UNIVERSITY OF CAPE COAST

SENIOR HIGH SCHOOL TEACHERS' PERCEIVED INFLUENCE OF LARGE-SCALE TESTING ACCOUNTABILITY ON TEACHING AND LEARNING AT BIRIM CENTRAL MUNICIPALITY

BY

ISAAC ENNU BAIDOO

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

DECEMBER 2021

DECLARATION

Candidate's Declaration

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

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

Name: ……..

Supervisor's Declaration

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

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

Name: ……..

Co- supervisor's Signature:……..... Date…….

Name: ……..

ABSTRACT

The study's objective was to examine the influence of large-scale testing accountability on teaching and learning. The study used a descriptive survey design. The study enrolled 200 senior high school teachers in Birim Central Municipal. The study selected teachers from the schools using simple random sampling, specifically the lottery method. Descriptive statistics (frequency and percentages, means and standard deviations) were used to analyse the data collected for the research questions, while inferential statistics, including exploratory factor analysis, independent t-test, and ANOVA, were used to test the hypotheses and answer research question two. The study's findings indicate that senior high school teachers in Birim Central Municipality have adequate classroom assessment skills to use classroom assessment effectively to improve teaching and learning. The Ghana Education Service should use large scale testing to results to advocate better resource allocation in education and teacher professional development in order to improve teaching.

KEYWORDS

Large scale testing

Accountability

Teachers

Birim Central Municipality

Assessment

ACKNOWLEDGEMENTS

First and foremost, I would like to express my gratitude to the Almighty God for His love, blessings, and numerous successes bestowed upon me throughout my life. My heartfelt and incalculable thanks go to Dr. Kenneth Asamoah Gyimah and Prof. Mark Owusu Amponsah for their contributions, recommendations, and support toward the completion of this thesis. Additionally, I am indebted to my wife, Florence Agyei, brother, David Baidoo-Anu, and biological mother, Mary Donkoh, for their counsel, prayers, and moral support.

DEDICATION

To my children, Benedict Fiifi Ennu Baidoo and Audrey Nana Kroma Ennu Baidoo.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

Background to the Study

The issue of accountability in school systems has gained a significant amount of attention around the world. According to Lin (2011), the primary objective of these accountability systems in education is to help improve and encourage academic attainment. According to existing research, school accountability is mostly based on large-scale measures of student achievement obtained through testing (Figlio & leob, 2011). In lieu of this, large-scale testing has become the vehicle of choice for accountability around the world, and testing has evolved from a instrument for making student-related decisions to a lever for holding schools accountable (Brill, Grayson, Kuhn, & O'Donnell, 2018).

Large-scale testing has always existed in schools, with the explicit and exclusive goal of making decisions regarding individual students' educational status (Lin, 2011). According to existing research, one of the main goals of large-scale testing is to increase student accomplishment by holding people in charge of student education accountable (Papay, 2012). Researchers such as Smyth, Banks, and Calvert (as cited in Aysel, 2012) refer to large-scale testing as standardised examinations or assessments, and the results might have significant repercussions for schools and/or students. The descriptions of Smyth, Banks, and Calvert (2011) were extremely similar to Lin (2011). According to Lin (2011), large-scale assessments are used to make major educational judgments regarding students, teachers, schools, or school districts.

According to Papay (2012), large-scale assessments are a fair way of identifying the best applicants for scarce resources, and they have been used to assign students to various programmes or into the workforce.

Large-scale testing has been identified as critical for all stakeholders in the educational system in Sub-Saharan Africa, including educators, students, parents, public examining bodies, and policymakers (Brown, 2016). According to Brown, large-scale testing not only measures the quality of teaching and learning in schools but also influences the future careers of students. As a result of political, social, and economic constraints, accountability issues have gained recognition and relevance in government administrations around the world. This accountability system varies, unfolding with different speeds and impacts (Gracin, 2011). Political and public demand for higher accountability has resulted in an increase in testing requirements, placing educators and students under increased scrutiny to ensure efficient and successful learning (Lin, 2011).

Testing on a large scale is one of the most divisive problems in education. While some policymakers feel testing is an essential component of school improvement and development, others believe it jeopardizes the quality of teaching and learning when it is used to make high-stakes decisions (Papay, 2012). Proponents of large-scale testing assert that the pressure created by these tests motivates students to work harder and teachers to be more productive in their instructional practices. According to them, high-stakes tests provide students and teachers with additional information on their own knowledge and abilities, as well as a clearer picture of each student's strengths and weaknesses (Carnoy & Loeb, 2002; Hamilton & Gonzales, 2003; Hess, 2002; Roderick, Jacob & Bryk, 2002). Other experts agree that by holding schools accountable

for student achievement, high-performing schools can be rewarded while low-performing schools can receive additional support and resources (Turner, 2014).

Additionally, proponents assert that accountability systems based on test scores will contribute to bridging the achievement gap between low-and high-income pupils. Accountability systems are premised on the assumption that holding schools accountable for standardized test scores will increase student success, particularly in schools with a large percentage of low-income learners. Additionally, it is claimed that incentives associated with student achievement inspire administrators, teachers, and students to adjust their behaviour, such as teaching methods and study habits, in order to achieve desired outcomes. The purpose of disaggregated test scores is to identify at-risk children and schools, allowing for more efficient resource allocation (Lin, 2011).

Opponents of high-stakes tests argue they increase learners' risk of educational failure, subject students and teachers with unequal resources to the same standards, restrict and distort the curriculum, and exacerbate class and racial imbalances. Additionally, they contend that high-stakes testing conveys the idea that the major goal of schools should be to achieve high test scores (Papay, 2012). Whatever one's opinion on the usefulness and utility of large-scale testing, there is no doubt that these programmes have had a tremendous impact on students, teachers, and school administrators. Countries, for example, use large-scale exams to keep heads of schools and students accountable. Students in these systems do not graduate unless they pass an exit exam, and principals do not receive increases until student success improves to a certain level. Heads of schools are under pressure to raise students' performance

because they fear losing their contracts if they do not (Kellaghan, Greaney & Murray, 2009).

In addition, opponents of high-stakes accountability do not believe that large-scale testing, when used to make significant decisions, widens the achievement gap between advantaged and underprivileged students (Breneman, 2010). Educators, they argue, are forced to focus only on the content and abilities of the assessments because of the pressure to meet AYP and avoid penalty. As a result, students' educational experiences are negatively impacted, especially in schools with large percentages of low-income students (Bennett & Wilezol, 2013). In addition, the pressure to improve test scores may influence student placement decisions, such as retention and special education identification. They argue that test-based accountability practices hinder these institutions' capacity to recruit and retain highly skilled instructors (Lin, 2011).

Various countries, including Ghana, have made large scale testing the centerpiece of their school reform (Anane, 2015). In lieu of this, large scale testing is prevalent in educational system in Ghana. These tests are required of students at Junior High School three (JHS 3) and Senior High School three (SHS 3). Nationally mandated examinations are exceedingly competitive, regardless of level, and the results are utilised for high-stakes decisions: they determine students' passage to the next grade level.

At present, national-mandated tests such as the West African Senior School Certificate Examination (WASSCE) for SHS 3 students and the Basic Education Certificate Examination (BECE) for JHS 3 have become high-stakes (Amoako, 2019), as a result, outcome of these national mandated tests are used for deciding the quality and effectiveness of school, and teachers. Due to the

prevalent and dominance of large-scale testing, teachers have the perception that assessment is mainly for accountability (summative) purposes (Oduro, 2015).

Statement of the Problem

Even though our education system in Ghana has been much improved through policy, there are still issues with regard to student accomplishment (Akyeampong et al., 2007). Large-scale tests in Ghana, like in many underdeveloped countries, are high-stakes in nature. Tests that are used to make decisions about learners, educators, schools, and districts are called "high stakes." For example, high school graduation and grade promotion tests are examples of high-stakes tests for students. Teachers should be aware of which tests have the greatest impact on their professions and salaries. Institutions and districts frequently use high-stakes tests to determine how much money they should allocate to each school and how to restructure those schools (Lin, 2011; Papay, 2012).

In Ghana, there are a number of large-scale tests offered to students at various educational levels (i.e., final year students at junior high school and senior high school). The West African Senior School Certificate Examination (WASSCE) and the Basic Education Certificate Examination (BECE) are now high-stake tests imposed by the government (Anane, 2010; Amoako, 2019). According to the country's former Minister of Education, Dr. Mathew Opoku Prempeh, a performance contract would be signed soon by heads of senior high schools with the government to ensure that students' academic performance improves. A performance contract would be utilised to assess whether or not a headmaster would be transferred or made to stay, Dr Prempeh said on Tuesday, March 27, 2018. According to Dr. Prempeh, it is up to the government to ensure

that teachers' needs and desires are addressed. According to him, the Ministry of Education will not tolerate any inappropriate behaviour from school heads that could jeopardise the academic progress of students (GhanaWeb, 2018).

Consequently, what the Minister said was put into operation. Since September 2018, heads of senior high schools have been signing performance contracts with the ministry. They are rewarded or cautioned accordingly following the release of the WASSCE results. In lieu of this, teachers have the perception that assessment is primarily used for accountability (summative) purposes, obscuring its developmental (formative) function with regard to student learning (Oduro, 2015). "This assessment view limits teachers' ability to explore ways to involve their students in the assessment process and make them take ownership of their own learning" (Oduro, 2015. p. 113).

Due to the high stakes associated with these national mandated assessments, particularly in Ghana's senior high schools, many teachers consider assessment as primarily a means of determining teacher effectiveness and school excellence (Anane, 2010; Oduro, 2015). As a senior high school teacher in Birim Central Municipality, I've seen that large-scale testing has had an effect on my teaching and I've found myself largely teaching to the test. According to Turner (2014), this phenomenon is referred to as "washback." The phrase "washback" refers to the effect of tests on education and learning. Taking a test can influence participants' beliefs and attitudes, such as teachers, students, and administrators, which has an effect on the work they do (processes), which in turn has an effect on the learning results they obtain (products).

Amoako (2019) examined the BECE's perception as a high-stakes test in Ghana and its impact on curriculum implementation. According to the study,

using the BECE as a high-stakes test limited the curriculum. Although several studies have examined the impact of large-scale testing accountability on curriculum and learning at the elementary level, there are few studies on how large-scale testing accountability affects the teaching and learning of senior high school teachers, particularly in Birim Central Municipality. In light of this, the purpose of this study is to examine how large-scale testing accountability affects teaching and learning and to make recommendations for methods to improve teaching and learning in this era of large-scale testing domination. My concern is this: what happens as the stakes for large-scale testing increase? How does accountability for large-scale testing (high-stakes) affect teaching and learning?

Purpose of the Study

The study's main purpose is to investigate the influence of large-scale testing accountability on teaching and learning. The study objective will specifically find out;

1. teachers' classroom assessment practices
2. the perceived influence of large-scale testing accountability on teaching and learning
3. how teaching and learning can be improved in this era of large-scale testing dominance
4. whether there is significant difference in teachers' perceived influence of large-scale testing accountability on teaching based on academic qualification
5. whether there is significant difference in teachers' perceived influence of large- scale testing accountability on teaching based on years of teaching experience

Research Questions

Based on the study's purpose, the following research questions were formulated to guide the investigation:

1. What are senior high school teachers' assessments practices at the Birim central Municipal?
2. What are the perceived influences of large-scale testing accountability on senior high school teachers teaching and learning in Birim central Municipal?
3. How can teaching and learning be improved in this era of large scale-testing accountability in senior high schools in Birim Central Municipal

Research Hypothesis

1. H_0: There is no statistically significant difference in teachers' perceived influence of large-scale testing accountability on teaching based on academic qualification

   H_1: There is a statistically significant difference in teachers' perceived influence of large-scale testing accountability on teaching based on academic qualification
2. H_0: There is no statistically significant difference in teachers' perceived influence of large-scale testing accountability on teaching based on years of teaching experience

   H_1: There is a statistically significant difference in teachers' perceived influence of large scale testing accountability on teaching based on years of teaching experience

Significance of the Study

This study is very important in a number of ways since it arm educational policymakers with evidence about the positive and negative effects of large-scale testing accountability on teaching and learning. Knowledge of this help formulate policies that reduce the negative effect of large-scale testing accountability and improve teaching and learning.

The study also help teachers become aware of or realize how large-scale testing accountability influences their teaching so that they can avoid negative washback and help students achieve the learning expectations. The study would also help students to become aware of how large-scale testing influences their learning and also help them to develop positive attitudes towards learning. Finally, the study serve as an important reference source for researchers to use as a supplement to studies already undertaken in this area.

Delimitation of the Study

The study was confined to the Birim Central Municipality in the Eastern region of Ghana. The study focused on professional teachers in public senior high schools in Birim Central Municipality on their classroom assessment practices, academic qualifications, and years of teaching experience.

Limitations of the Study

The small sample size in one municipality in Ghana, however, limits the generalizability of the study's findings, which is a limitation of the study. Again, the study relied largely on questionnaires as the main data collection instruments, which can introduce biases. In some cases, this may be due to social desirability bias or a desire to protect one's privacy. However, assuring participants of anonymity and confidentiality of their responses helped reduce

the probability of this problem occurring. Notwithstanding, this study has provided some unique and compelling evidence.

Organization of the Study

The research in five chapters, each focusing on a distinct aspect of the research. The study's background, problem statement, purpose of the study, research questions and hypothesis, delimitation, limitation, and significance of the study were all explored in the first chapter. The second chapter analyzed related literature, while the third chapter discussed data collection methods, sample size and sampling procedures, as well as descriptions of the research instruments used to collect data and the method of data analysis. The fourth chapter discusses the anticipated results. Chapter five summarizes the study's findings, conclusions, and suggestions.

CHAPTER TWO

LITERATURE REVIEW

Introduction

The study explores senior high school teachers' perceived influence of large-scale testing accountability on teaching and learning at Birim Central Municipal. The literature review is based on similar studies that have been done in various countries around the world and on the sub-themes drawn from the research questions.

Theoretical Framework

Based on the overarching objective of the study, the economic theory of the principal-agent problem was employed to provide a theoretical lens for the study. An economic theory of the principal-agent problem was proposed by Grossman and Hart in 1983. This theory has been extensively used to provide a justification for the proliferation of the use of data-motivated accountability in schools (Figlio & Loeb, 2011; Rothstein, Jacobsen, & Wilder, 2008). The principal-agent problem occurs when one person or group (such as schools or teachers) acts on behalf of another person or group (such as parents or governments). Suboptimal outcomes occur when there is misalignment in the aims and incentives of the principals (i.e., parents, governments) and the agents (i.e., schools, teachers).

In this study, I regard the government, the Ghana education service, parents, and educational policy actors as the core "principals" in the educational system, while secondary school teachers are the key agents. That is, secondary school teachers are the key agents or groups that act (i.e., teach secondary school students) on behalf of the Ghana education service, parents, educational policy

actors, and the government. Misalignment in the aims and incentives between teachers and governments or educational policy makers results in a "principal-agents problem" between teachers and educational policy makers (Jerrim & Sims, 2021). To solve this "principal-agent problem" in education, Figlio and Loeb (2011) stated that government and educational policy makers (i.e., principals) use data-driven accountability to force teachers (i.e., agents) to align their goals and objectives to the goals and objectives of the government and educational policy makers. Figlio and Loeb (2011) emphatically stated that "the information content in school accountability systems can provide a powerful mechanism for overcoming the principal-agent problem.

Assessing schools against the common metric of standardized student test scores provides policymakers and members of the general public with independent information regarding how well schools and school districts (and potentially teachers) are doing in comparison to their peers and outside performance standards. Measuring and reporting school performance and attaching positive and negative consequences to meeting or failing to meet performance objectives provides incentives that encourage educators to concentrate on the subjects and materials that are being measured and to potentially alter the methods through which they educate students. The measurement and reporting of a school's progress allows policymakers to assess how successful a school has been in meeting the state's achievement goals "(p. 386).

Gleaning from the theory, school-based accountability is premised on the idea that public pressure from openly reported data will lead to enhanced student success and achievement (Figlio & Loeb, 2011; Supovitz, 2009;

Rothstein, Jacobsen, & Wilder, 2008). However, extant literature explicates that data-driven accountability has unintended negative consequences for the agents (i.e., teachers) (Grupe & Nitschke, 2013; Smith & Amick, 1989). For example, Smith and Amick (1989) found that accountability put too much stress on teachers. Similarly, Grupe and Nitschke (2013) maintained that close monitoring and punishments attached to accountability lead to stress and anxiety. Further, data-driven accountability has also been found to cause a loss in job autonomy (Jerrim & Sims, 2021). Thus, teachers do not have the liberty to work due to data-driven accountability. Finally, data-driven accountability forces teachers to do more work, such as preparing students for tests (Perryman & Calvert, 2019). This adds extra work to teachers and puts them under pressure, consequently adversely affecting classroom teaching and learning.

Conceptual Review

Concept of Educational Assessment

According to Allen (2004), educational assessment is the systematic method of gathering and evaluating analytical data on experience, abilities, attitudes, and beliefs in order to improve student learning and refine programmes. Bardes and Denton (2001) thought of educational assessments as systematic ways of gathering data under the same conditions and drawing a conclusion about an employee's knowledge, qualifications, and potential.

Bardes and Denton again reported that educational assessment is information about educational programmes undertaken to improve training and development that has been systematically collected, examined, and used. According to the Yukon Education Department (2015), educational assessment is the systematic process of gathering data from a range of sources in order to

determine the most appropriate type of education.It identifies the strengths and needs of the student and helps to design and implement effective strategies for teaching and learning.

Educational assessment is thus inherently linked to student learning and performance in educational programmes. Evaluating the performance and learning of students involves a number of assessors and includes considering both hard and soft skills. The effect of assessments on student performance was considerably observed by Struyven et al. (2006). The approach to learning by Students determine their own ways of thinking and testing in the classroom. Pellegrino and Goldman (2017) and Shepard (2000) suggested an improvement in classroom assessment with a view to enhancing learning, such as assessment content, evaluation results use, etc. To promote learning, Pellegrino, Goldman, and Shepard outlined a number of strategies to improve educational assessment, including the content and characteristics of assessment and the utilisation of assessment data.

It is noted that because evaluation has an important impact on the student's approach to learning, evaluation patterns have shifted from student testing to student appraisal (Birenbaum & Feldman, 1998). The contemporary evaluation system aims to improve the connection between what students must study and what they are expected to know upon graduation (Gulikers, Bastiaens, & Kirschner, 2006). The urge, however, persists regardless of whether students are trained to thrive on a test or to construct meanings that sustain them over time. The aims of educational assessment vary according to the stakeholders (students, teachers, stakeholders, parents, schools, and policy-making stakeholders). Although instructors and administrators commonly select

evaluation forms and tasks, Romanoski, Cavanagh, Fisher, Waldrip, and Dorman (2005) assert that given that both student preferences and teacher rationale can impact how students proceed with learning and are assessed, I would consider it appropriate to incorporate student and teacher opinions when designing evaluation systems.

According to Goodrum and Hackling (2005), educational evaluation should ideally enhance learning, provide feedback on the student's progress, foster self-confidence and self-esteem, and foster assessment skills. Additionally, they claim that effective learning occurs when instruction, evaluation, and results are in sync. Educational evaluation is crucial to learning since it is directly related to instruction and learning outcomes. According to Goodrum, Hackling, and Rennie (2001), evaluation is a fundamental component of the educational and learning processes. According to Brown and Pendlebury (1992), the objective of appraisal is to provide advice and feedback to the learner. The nature, scope, and feedback provided by this advice are governed by the evaluation's goal. To prepare students for life, evaluation is based on the premise that learning does not occur exclusively during formal education but occurs throughout life (Black and Williams, 2010).

Given the effect of educational assessment on learning, Black and Williams (2010) argued that it should assist students in comprehending their own learning by providing self-reporting and discouraging them from relying on others for information about their performance. Boud and Falchikov (2006) felt the need to give equal attention to all educational assessment components in conjunction with well-established assessment purposes to certify and evaluate current learning aids. The teacher identifies the necessary learning, the

associated educational assessment and criteria, the student's performance and awarded graduation as traditional approaches to assessment. Such methods place the student in a passive rather than an active role in the assessment process, which contradicts the need for long-term assessment procedures that aid students in preparing for lifelong learning beyond the academy.

In 2006, Boud and Falchikov stated that educational evaluation activities should not just satisfy students' urgent certification needs or feedback on their current learning, but also contribute to their future education. Students will need to be prepared to appraise themselves and their performance and their learning as super-complexities that change the information required in a job when they leave the confines of an educational environment. In 2002, Duke mentioned that educational evaluation can be conducted in a variety of methods. The consequences of student learning must be evaluated using a variety of techniques. Direct and indirect methods of evaluation are available. Paper and pencil tests and presentations are examples of direct evaluation methods, while surveys and interviews are examples of indirect assessment methods.

Types of Educational Assessment

For a comprehensive assessment plan, both formative and summative assessments should be included in the overall assessment process. These two types of assessments are distinguished by the context in which they occur within a programme.

Formative Educational Assessment

According to McTighe and O'Connor (2009), "educative evaluation" is a term used to describe formative educational assessment, which is typically used throughout a course or project. In an educational setting, an educator or a

student may conduct a formative assessment to provide feedback on a student's work, and this feedback is not always used for graduation purposes. Research or diagnostic trials, standardised tests or tests, oral questions or draught papers are all examples of formative educational assessments. Instructions are used in conjunction with formative assessment to improve student learning. Prior to taking a summative exam, students take a formative evaluation to see whether or not they've grasped the material.

According to Black and Williams (2010), formative educational assessment is interactive and is primarily used to form or modify a continuous learning process or activity. Formative educational assessment focuses on student motivation and learning in order to produce higher-quality work or thinking. It is critical to recognise that there are two distinct observers for formative educational assessment (Edmund, 2006). According to Edmund, teachers can assess their students' understanding of a subject by asking questions or listening to them talk about it in small groups.

For the purpose of identifying students' learning needs and implementing effective teaching strategies, teachers collect data about students informally. Giving students immediate feedback on what they've learned is a key component of formative educational assessment. Effective feedback is often regarded as having a major impact on student achievement (Marzano, Pickering, & Pollock, 2001). Formative educational assessment, according to Black and Williams (2010), is commonly undertaken at the beginning or during the course of a programme, providing direct evidence of study learning on a single course or at a certain moment in a programme. Formative assessment is frequently used in the classroom to help teachers enhance the quality of their students' learning

(Black & Williams 2010). Depending on the outcomes of a course's formative educational assessment, which is a crucial part of teaching and learning, the curriculum may be changed.

According to Black and Williams (2010; 1998), there are several benefits to using formative assessment in a classroom setting. For example, it allows programmes to analyse if the learning goals and objectives are accomplished in each segment of the course. Faculty involvement in the planning and implementation of course objectives and programme effects can also increase instruction quality (Bardes & Denton, 2001). Almost all major stakeholders focused on formational assessment in the classroom in order to synthesise research studies on classroom assessment. A synthesis of more than 250 studies revealed that it has a greater impact on learning than the summarised formative educational assessment used by Black and Wiliam (1998). Crooks (2001) concluded in his study review that the effect sizes for summative educational assessment are consistently lower than the impact size for formative educational assessment.

It is worth noting that formative educational assessment information from the classroom can contribute to a comprehensive assessment plan by allowing special points in an appraisal programme to be identified and progress toward learning results to be monitored (Bardes & Denton, 2001). According to Sasser (2018), formative educational assessment should be used to qualitatively monitor students' learning rather than quantitatively examine it (final exam). As a result, formative evaluation is the best way to assess learning. As a result, formative assessment is preferable for assessing learning and summative assessment is preferable for assessing learning measurement.

Formative educational assessment, according to Reddy (2018), is important for changing student behaviour in the classroom. To deal with unexpected results and respond to emerging properties, formative evaluation is required. Reddy (2018) outlined some of the benefits of formative evaluation: The primary goal of formative educational assessment is to assist students in developing knowledge and skills. Instructors, leaders, or teachers can use this assessment category to identify individuals' needs and direct them to their educational goals. This method identifies the individual's obstacles and difficulties and applies appropriate remedies to overcome them. An assessment is provided by the instructor or teacher with a formative assessment to ensure that the individuals have mastered the concept that has been taught to them.

Formative educational assessment is beneficial because it helps to plan for the future by determining whether teaching or other career-related methods can be modified. Weakness is identified early and remediation is carried out. This keeps people on track and results in continuous feedback. In the event of a change in the teaching method or task, future planning with formative evaluation is recommended. Formative educational assessment encompasses a wide range of diagnoses that students or individuals require. Feedback is a critical factor in allowing students to reflect on what they have learnt and why. Formative educational assessment helps students improve their performance and achieve successful outcomes.

It is well known that when formative educational assessment is reduced to a mini summative educational assessment, or to a number of teaching techniques for improving levels and qualifications, problems are presented. In the case of larger accountability schemes such as National

Curriculum Assessment, a serious threat to the efficiency of formative evaluation occurs (Shuichi, 2016). For Sasser (2018) and Reddy (2018), some teachers may complain that the lesson is not even finished, and that during the course they will sacrifice time to appraise them.

In the classroom, formative educational evaluation is considered a time-consuming procedure when teachers regularly observe students. As a result, these evaluations are time-and money-consuming because they require constant data collection. As a result, teachers may be unable to practise due to time constraints as a result of regular data collection, analysis, and reporting, as well as the refinement of new implementations and their success.The planning and evaluation of the formative educational assessment can be an exhausting process for the teacher. Again, many teachers do not have the skills to formally evaluate their students (Reddy, 2018).

Summative Educational Assessment

According to Mctighe and O'Connor (2005), end-of-course assessments are common in educational settings. In the classroom, summative educational assessments are frequently used to classify students. Then, a review of what the students have learnt and how well they comprehend the subject is conducted. In projects, this form of assessment is often tested, analysed, or performed. Summative educational assessments are widely used when determining whether a student has completed, passed, or failed a grade level in school. Summative educational assessments are used to judge whether or not a student has met or exceeded the course's learning goals (Edmonds, 2006).Summative educational assessments are typically administered at the end of a unit or course to document what students have learned. Looking at a teacher's grades should provide insight

into what the main instructional objectives or results have been in a grade period. Summative educational assessments (tests, quizzes, projects, reports, and written tasks) will most likely inform the teacher of the student's mastery of the skills or content.

The level at which students need to "master" content and thinking, is a key aspect of summative assessment. Tests defining "mastering" content for events, titles and facts are less likely than the tests which ask students to write about large conflicts or topics that reoccur over a time period to build the thinking ability of students. The summative educational assessment in the classroom is comprehensive, gives responsibility and is used at the end of the programme to verify the level of learning according to Angelo and Crow (1993). For example, if students acquire the knowledge necessary to pass a certification examination following completion of the programme, taking the test would be a summative educational assessment, as it would be based on the cumulative learning experience. The summative classroom assessment programme's objectives frequently reflect the collaborative aspect of education (Palomba & Banta, 1999). Summative educational assessments are critical at the conclusion of any educational programme to confirm that students have met the program's objectives.

As a result, they put in less effort to study advance their academic achievements in the future. One major drawback to a summative evaluation is that it focuses the final analysis on output. In the case of impediments or problems, the process of learning may be difficult. Bardes and Denton (2001) articulated that a comprehensive plan should be followed by different methods and measures to gather detailed data and credible assessment results. The

foundation for an evaluation plan will eventually be the collection of summative educational assessment information, which can be stand-alone.

Reddy (2018) reports on the summative educational assessment at the end of the assessment cycle. It is a kind of evaluation that evaluates the value of the task until the programme is concluded. The summative educational assessment focuses mainly on the results. Summative educational assessment may also be referred to as evaluation techniques used to measure individuals' and students' results. Summative educational assessments, on the other hand, are used in education to evaluate students on what they have learned. The following merits from a summative assessment are proposed by Reddy.

Summative educational assessment uses assignments, testing, projects, and more to follow certain evaluative strategies. Thus, the teacher can determine whether the students have learned the topic and understand it. Summative refers to the utilisation of a task rather than the design of the test, assignment, or self-assessment. Thus, the instructor can ascertain the degree to which students comprehend the instructional materials. Typically, summative educational assessment occurs at the conclusion of any term of instruction. As a result, summative educational assessment is viewed as an evaluation rather than a diagnosis. Additionally, they are used to determine the effectiveness of instructional programmes. Another significant benefit is that they are used to track progress toward objectives and goals. Additionally, summative assessment is used to make course placement decisions.

Summative educational assessment outcomes are those that are documented as scores or grades in learners' academic records. They can be found in test scores, letters, or report cards that are used in the college

admissions process. Numerous schools, districts, and courses view summative educational assessment as a critical component of their evaluation system. The inclusion of a summative educational evaluation assists students in developing a learning environment and serves as a motivational factor. This is a formative assessment of learning based on the results. The summative educational assessment result is viewed as a good enhancement feature. This type of assessment increases self-esteem and lays the groundwork for particular alterations in behaviour at work or school (Reddy, 2018).

Summative data can be used by trainers and educators to identify weak areas where outcomes are consistently low. This can be used to enhance the outcomes obtained through the use of alternative procedures. New training can be implemented to ensure future success. A summative educational assessment can be used to evaluate the teacher or instructor. This assessment can be used to examine the efficacy of all teachers or instructors. A summative evaluation satisfies the school's responsibility requirements for the teacher. Although the summative evaluation does not include value added as a metric of the finished product, it is not defective. A critique levelled regarding summative educational evaluation is that it is reductive, with learners discovering their level of knowledge acquisition far too late for it to be useful (Mctighe & O'Connor, 2005).

The following limitations are proposed by Reddy (2018): A summative assessment demonstrates the motivation and impact of students to show that students that have performed poorly have a lower self-esteem at the end. As the results are at their end, there is no chance of recovery. When learning is considered, this is not an accurate reflection. In a summative educational

assessment, nothing is done to identify obstacles or challenges in good time. Instructional problems are not identified until the summative assessment is used in schools and becomes critical. Because summative evaluation is a single test at the end of the entire academic session, it causes nearly all people anxiety and trouble. They are nervous and afraid when they have to take a summative educational test.

In considering summative educational assessments, the focus is primarily on the teacher's performance during testing. Overall, summative educational assessment is not perfect because even outstanding students can confront issues that can make them fall. A student can get nervous or tense due to the pressure of examinations. The main reason would be Summative educational assessment is therefore not considered to be the best reflection of school learning. Repeated testing for students at low levels reduces self-confidence and self-esteem.

When summative educational assessment results are more prominent for students than for schools or authorities, they have a negative impact on low-level students. Secondary school students with low levels of education have the potential to perform worse over time. Additionally, it is viewed as a process that limits a capable individual. Anxiety is another factor, exacerbated in particular by girls, that contributes to the widening of the gap between those with higher and lower levels of attainment. Teachers modify their lessons to meet the needs of the test and deviate from the curriculum and content. Instruction techniques can be distorted. Additionally, summative educational assessment questions may not be organised similarly to formative educational assessment questions. Teachers and instructors can devote additional time to summative educational

assessments, which cannot truly improve an individual's knowledge. Additionally, teachers employ a didactic style that many students may not find ideal or convenient.

Testing as an Assessment Technique

Testing is therefore one of the strong instruments for measuring students' skills and enhancing their learning attitudes. Hughes (2003) supports this notion by saying that testing is an instrument for measuring student language skills. According to Brown (2004), a test is a method of assessing an learner's skills or performance in a specific subject. Additionally, in the same vein, Nitko (1983) stated that testing is the process of observing and documenting one or more characteristics of a person using either a number or category system. Using a numerical scale or category system, Cronbach in Azwar (2005) defined the term "test" as "a systematic technique for watching a person's behaviour and reporting it." It is important to note that a test is a systemic description, collection, and interpretation technique that is used to measure the test user's ability, knowledge, and performance in the learning process and to gain a judgement of value. It is the goal of a test to provide reliable information about a student's abilities and knowledge. The test results show that teaching and learning have been successful.

A test can be classified based on the information it provides. Tinambunan (1998) classifies testing into four categories: “placement tests, formative tests, diagnostic tests, and summative tests”. English Language is just one of many different types of language tests.Baily (1998) thinks there are eight different types of language evaluation. Language domination, talent testing, admissions, and placement tests are just a few of the tests that students take to

determine their fit for a certain programme. According to experts in language testing, the types of tests that will be reviewed and used in this study include accomplishment tests and summative tests.

Taking an achievement test is a way of gauging the accomplishment of individual students, groups of students, or entire classes in a language course. According to Richard's book, Dictionary of Language Teaching and Applied Linguistics (1992:3), "an achievement test (test for performance) is a test that examines how much a learner has effectively acquired with relation to a certain amount of time or course of instruction," he writes. Also, according to Brown (2004a), a performance test measures how far students can get through a curriculum of content in the time allotted for the test. At the end of the course of study, an accomplishment test is administered to determine the students' learning outcomes. The scope of the test should reflect the course or units that students have been taught.

When a course or training unit is finished, a summative test is administered to see how students performed. According to Brown (2004), there are summative tests, such as Kentucky's Commonwealth Accountability Tests Systems (CATS), that are mandated by the federal government. Summative tests are designed to indicate how pupils at the same grade level stack up against their peers. Thus, summative assessments are used to evaluate a student's overall performance after completing a set of courses. In most cases, the summative test is based on one semester's worth of material. The scope of a test that emphasises complete learning results encompasses all of the topics. The summative test results are used to establish whether or not students are capable of following the higher level of instruction in the subsequent instruction programme, as well as

their development over the semester. Report cards and printed sheets are used to express the degree to which an individual has achieved an accomplishment.

Large-Scale Assessment in Education

For example, large-scale assessments are standardised examinations administered at the national or international level that provide a picture of a group of students' learning performance over the course of an academic year. The use of these evaluations has increased and extended throughout the world. A broader range of subjects, such as digital skills, computer and information literacy, socio-emotional abilities, and comprehension of civics and citizenship concepts and challenges, are examined on a larger scale than reading and mathematics (UNESCO, 2019; Chudowsky & Pellegrino, 2003).According to Emler, Zhao, Deng, Yin, and Wang (2019), the use of large-scale testing assessment (LSAs) in education has increased greatly. Tobin, Lietz, Nugroho, Vivekanandan, and Nyamkhuu (2019) claim that the use of LSAs in education has expanded significantly (2015).

Large-scale assessments of student learning are standardised to allow for comparison between individuals, schools, and, in some circumstances, countries in previous decades. Many countries have developed more LSAs to determine teacher efficiency, school performance, and resource allocations. Both BECE and WASSCE are commonly used in Ghana to rank schools, encourage students to pursue further education, and hold teachers accountable. The importance and use of large-scale SAs in education has grown. The LSAs can be effective instruments for holding educators accountable, as well as efficient techniques for gathering and verifying policies. They are also considered objective and incorruptible tools for sorting and selecting competitors (Emler, Zhao, Deng,

Yin, & Wang, 2019). As the use and impact of large-scale assessments grows, so does the criticism leveled at them. Despite the fact that LSA has never been without controversy, its popularity has grown (Lewis & Lingard, 2015). There is no technical solution to the LSA issues.

These issues are unavoidable yet undesirable side consequences of effective therapies. Like any other educational programme or practice, LSAs in education are accompanied by inescapable secondary effects while generating beneficial results (Karsten, Visscher, & De Jong, 2001; Zhao, 2017, 2018d). That is, while LSAs can be effective in resolving some situations, they can also be inevitably harmful, because the mechanism that causes it is also useful and efficient. The goal of this research is to assess the of large-scale testing on education.

Large-scale assessments have been used for a variety of purposes. Three of them are in charge of the primary functions of selecting, comparing, and evaluating. Responsibility LSAs are assessments that hold teachers accountable for the learning of their students. Large-scale assessments are often used in many forms of selection, from educational opportunities like college and graduate programmes to employment like government officials. The Scholastic Assessment Test (SAT) and the American College Testing (ACT) are two examples of LSAs in the United States (US). A third reason LSAs have been utilised is to make comparisons of educational quality across regions, states, and nations. The Trends in International Mathematics and Science (TIMSS) is one of the most well-known international comparison studies (Emler, Zhao, Deng, Yin, & Wang, 2019).

Occasionally, the purposes overlap, and one LSA may be used for many reasons. While the national college entrance exam is primarily used for admissions purposes, it is well recognised that students' exam performance has an effect on schools and teachers. While the national college entrance exam is primarily used for admissions purposes, it is well recognised that students' exam performance has an effect on institutions and educators. It is also connected to educators' financial security, as student performance on the exam is directly related to teachers' income and other benefits.. Regardless of its goal, large-scale assessment is determined by its scale. A large-scale assessment involves the evaluation of a large number of individuals. A test can be deemed large-scale assessment only if it is given to a significant number of learners. In other words, irrespective of the assessment's objectives, the fact that it has a significant impact on a vast number of people makes it a large-scale assessment.. As a result of their large scale, large scale assessments have a number of distinguishing characteristics.

Uniformity

The validity and reliability of a wide-scale assessment must be maintained in a number of ways so that it can be presented to a large number of students. First, large-scale assessments must be in a uniform format or standardised so that accountability, selection, and comparison may be done objectively. Since diverse constructs can't be measured for various students, they should have a uniform structure or framework as well. It is also necessary to administer and score LSAs uniformly, as well as report the results consistently. This calls for a high degree of standardisation in the

administration, scoring, and reporting processes (Emler, Zhao, Deng, Yin, & Wang, 2019).

High Cost

Large-scale evaluations cost a lot of money .Regardless of who pays for it, designing, developing, managing, scoring, and reporting on LSA is expensive. It costs a lot of money for the Ghanaian government to conduct large-scale assessments. The Ministry of Education spent $68 million on only past questions for candidates taken West African Senior High School Certificate Examination (WASSCE).

Broad Impact

Large-scale assessments have a wide range of consequences for stakeholders. Designed LSAs have a huge impact on the lives of many students, test takers, teachers, schools, and, not least, their families, regardless of the function they serve. For the most part, large-scale assessments have a detrimental impact on test takers. An LSA can have a great impact on the lives of a lot of people who are related to the test takers. Assessments for postsecondary school admissions frequently have a direct impact on the test takers' families and, in certain cases, teachers, school administrators, and even local government officials who are held accountable for the outcomes.

High Stakes

Some stakeholders (such as students, teachers, and head teachers) view all LSAs as having high stakes. If you're a test user, some LSAs require a great deal of attention. It is simple to comprehend. Results from large-scale testing can be used in post-secondary education, for example. Students' scores

on large-scale tests are once again employed as a measure of teacher effectiveness and school quality (Emler, Zhao, Deng, Yin, & Wang, 2019).

Effects of Large-Scale Assessment in Education

The negative impact of LSAs on education is due to a combination of their power in dictatorial education and their limited range of skills, knowledge, and human qualities. There have been reports of negative or detrimental consequences in the literature. One of the most commonly used phrases to describe these impacts is "unintended consequences. 'Collateral damage" is another term for it (Nichols & Berliner, 2007). This section discusses some of the most alarming LSA side effects documented in the literature.

Distortion of Education

The distortion of education is one of the most harmful effects of LSAs. There are different levels of distortion. First, by persuading the general public that LSA performance correctly represents educational quality, LSAs alter the educational goal, albeit unintentionally. Few would dispute that education entails more than learning how to perform effective testing, let alone testing that covers such a small number of subjects (Sjberg, 2015; Zhao, 2016c). Large-scale assessments, on the other hand, have succeeded in instilling in the public the perception that training is primarily intended to prepare for testing and that test scores thus reflect the quality of education. The media portrayed education quality in TIMSS and PISA countries as excellent (Dillion, 2010; Sellar & Lingard, 2015), and the results were widely reported (Baird et al., 2016; Gruber, 2006).

Because it is dependent on the student's test scores, it is additional evidence that the National Assessment Programme-Literacy and Numeracy

(NAPLAN) "has become a de-facto assessment of schools" (Hout & Elliott, 2011). Second, if the curriculum's sole objective is to prepare students for tests, it is skewed. The distortion is primarily due to a reduction in the curriculum. Despite the fact that many educational systems provide for a considerably broader programme than the restricted number of subjects frequently reviewed with LSAs, the subjects evaluated are at the centre of focus and receive far more resources, pushing non-assessed individuals to the periphery. Exams are heavily influenced by what pupils should know and what they don't (Kirk-Patrick & Zang, 2011). Only disciplines that have been rigorously assessed via the National College Entrance Exam are permitted in China (Yu & Suen, 2015; Zhao, 2014).

Since the introduction of the NCLB in the United States, there has been a movement toward curricula reduction (Berliner, 2011; Klenowski, 2011; Klinger & Rogers, 2011; Tienken & Zhao, 2013). Between 75 and 150 minutes per week, the classroom was added to each of the two regions in U.S. students' classrooms, where students were evaluated on two major subjects (reading and math). The test disciplinary strands added up to 300 minutes of teaching when the two were combined (Berliner, 2011) Of course, this extra time must be offset by reductions in other subjects, abilities, and ways of thinking, such as social studies, science, physical education, recesses, art, and music (Berliner, 2011). The National Assessment Programme-Literacy and Numeracy has also created a more specialised programme for Australian students (Dulfer, 2012).

Finally, by their instruction, LSAs distort education, resulting in the teaching test phenomenon. Test teaching is instruction that is exclusively focused on and usually confined to the test's topic. It also implies that students

are assessed throughout the class for their abilities. Test teaching is a common occurrence in schools, in which students are taught not only how to take tests but also how to assess knowledge points in a variety of ways, such as by repeating model questions or passing tests from past years (Zhao, 2009, 2015b). The fact that the amount of time provided to pupils in the United States to prepare for state accountability testing has increased since NCLB has demonstrated the impact of LSA on distortion (Menken, 2006). There is also evidence that deep learning and other critical skills such as creativity, problem-solving, information organisation, and self-monitored abilities have considerably reduced the amount of time spent learning (Berliner, 2011; Klinger & Rogers, 2011).

When education is confined to test preparation, students lose access to skills that are vital to their future and the future of society (Zhao, 2012, 2014, 2015a). For example, the most important talents for today's learners to succeed include creativity, problem solving, knowledge organisation, self-monitoring, underachievement, and other skills and abilities (Duckworth & Yeager, 2015). In addition, the shortened programme allows students to explore and develop other talents as well as learn more about their passions (Berliner, 2011; Klinger & Rogers, 2011). According to Volante (2005), "excluded and reduced routes are no less useful in and of themselves; they are a vital element of a student's educational experience." For a large segment of the workforce in fields such as the visual arts, music, and physics, finding a new job is "critical" (p. 2). The fact that students cannot pursue various fields of thinking, creativity, self-esteem, commitment, and the desire to learn for the rest of their lives, on the other hand, reduces and affects different ways of thinking, creativity, self-esteem,

commitment, and the desire to learn for the rest of their lives (Berliner, 2011; Volante, 2005, 2006; Zhao, 2016c).

Discouragement and Psychological Challenges

Educators and students can become demoralised as a result of LSAs in the classroom. Most teachers begin their careers in education with a great desire and ambition to help students and foster lifelong learning. Teachers and leaders' motivation is negatively affected by LSAs' particular stressors and variables. Leaders of schools say they are torn between having more accountability for results and having less authority to determine the best practices and effectiveness of teachers (Goodwin, Cunningham, & Childress, 2003; Prytula, Noonan, & Hellsten, 2013). As a result, teachers report higher levels of stress and anxiety, as well as deterioration in their relationships with their colleagues and the administration, as well as emotions of guilt, shame, and failure (Berliner, 2011; Prytula et al., 2013). Because of the accountability mindset, teachers and administrators place a great importance on their learners' performance, creating a culture of fear and regression among the school faculty and staff, who believe they must do whatever it takes to earn the grades they deserve (Berliner, 2011; Klinger & Rogers, 2011).

Demoralization and psychological damage can occur to students as well. Large scale assessments in various countries, including China and South Korea, are frequently blamed for putting a tremendous amount of stress on students (Ho, 2012) .Large-scale assessments are employed as admission criteria in countries where only the greatest achievers are granted the option to enroll in elite colleges. Students in these countries have to put in years of intensive preparation. There is a lot of pressure on the students, not just from the exam

itself but also from their family members, teachers and school administrators, as well as the media and policymakers (Kirkpatrick & Zang, 2011; Kristof, 2011). National College Entrance Exam in China has long received criticism for inducing hazardous stress levels. A majority of Chinese students reported experiencing headaches and gastrointestinal pain due to their anxiety over upcoming exams (Hesketh et al., 2010). Students in Korea are the most dissatisfied in the 2015 Programme for International Student Assessment (PISA) compared to other nations in the Organization for Economic Co-operation and Development (OECD) (So & Kang, 2014).

While some stress can contribute to positive outcomes and advancement, excessive stress on students can be detrimental to their physical and mental health. Students avoid normal social interaction as a result of the psychological stress generated by negative exam-related experiences, which may have a long-term impact on their entire lives (Muthanna & Sang, 2015). To cope with the stresses of college, several students have turned to drinking and smoking (Unger et al., 2001). According to a 2013 study (Deng, 2014), more than 90% of elementary and middle school suicides happen after a student encounters school-related stress. Students commit exam-induced suicide in Japan, Singapore, China (Hong Kong and Taiwan), Korea, Cambodia, and Vietnam, among other places (Cui, Cheng, Xu, Chen, & Wang, 2011). Due to the test's high stakes and few chances, students devote the majority of their time to studying and even forego physical exercise. Students in their final year of high school in China are more inclined to stay in the classroom and study rather than participate in physical education classes, resulting in a high obesity rate among these students.

Corruption and Malpractices

Another unintended consequence of LSAs on education is moral corruption. The more quantitative social indicators are utilised in social decision-making, the more vulnerable they are to corruption pressures and the more likely they are to distort and corrupt the social processes they are supposed to monitor. This is the law of Donald Campbell (Campbell, 1976, p. 49). LSAs are not free from the rule of law because they are a quantitative indicator utilised in social decision-making. They corrupt and distort the truth. When it comes to LSA-caused corruption in education, cheating on tests is one of the most obvious examples. Certain countries have been accused of cheating on the PISA test (Loveless, 2014; Sands, 2017).

According to the Chinese media, malpractices in LSAs is a widespread problem in China (Kirkpatrick & Zang, 2011; Moore, 2013; Zhao, 2014). However, examination malpractices are not limited to China. Rather, it happens in countries where LSAs are used to make critical and high-stake decisions (Hassan, 2019).

There have been instances of teachers and school leaders engaging in unethical behaviour. Exact test preparation materials have been found to be used by teachers (Towles-Reeves et al., 2006), as well as copies of the test or look-alike items for preparation (Albrecht & Joles, 2003; Volante, 2005). Teachers have also given test answer keys ahead of the exam (Popham, 2000; Volante, 2006), and relaxed testing protocols such as timing and seating arrangements (Popham, 2000). Teachers and administrators send students on field trips or encourage them to stay at home in the hopes that they'll transfer or drop out, all for the sake of a few extra points (Albrecht & Joles, 2003; Volante, 2006). In

the United States, Atlanta public schools were the location of the most widely publicised cheating incident. In 2008, the Georgia Department of Education designated 58 Atlanta schools as potential hotspots for cheating and 44 as proven hotspots for cheating. At least 178 educators were identified as participants, with at least 80 admitting to taking part. It was one of the longest trials in Georgia's history, and a number of the defendants were imprisoned (McCray, 2018). Equal access, best practises, and the most effective education have all been compromised by LSAs, which were intended to secure these things but instead succeeded in eroding the empathy and abilities of those within educational systems, compromising their ethics, beliefs, and potential.

Exacerbating Inequity and Injustice

Additionally, LSAs exacerbate educational disparity. There is already much educational disparity and injustice in place as a result of cultural, social, racial, geographical, and historical factors. As a result of this, students, through no fault of their own, have significantly differing educational possibilities in terms of quality. It has long been known that a child's parental history is the most important predictor of his or her educational attainment and future success (Berliner, 2006; Hanushek, 2016; Lee & Burkam, 2002; Zhang & Zhao, 2014). In a variety of ways, LSAs contribute to inequity and injustice.

To begin with, using LSA scores to grant socioeconomic opportunities such as college admissions is biassed toward children from low-income families, as these students frequently perform worse than their advantaged counterparts due to unequal opportunities for education. Historically, large-scale evaluations favoured affluent students. SAT scores, for example, have a substantial positive relationship with family income (Dixon-Roman, Everson,

& McArdle, 2017), indicating that students from affluent families do better. Consequently, children from low-income families have fewer options to attend elite colleges, restricting their opportunities for upward socioeconomic progress and compounding the current imbalance and injustice faced by underprivileged children.

However, such essential decisions should not necessarily be made through LSAs. The case for using LSAs to make critical and high-stake decisions such as college admissions is based on the mistaken belief in meritocracy and the premise that LSAs can properly assess individuals' merit in order to award social resources and positions (Zhao, 2016b, 2018a, 2018b). Nevertheless, the seeming meritocracy does not exist. (Lemann, 2000), and LSAs cannot adequately capture the qualities necessary for future success. For example, the two most widely used college admissions tests, the SAT and ACT, have been shown to be poor predictors of college and life success (Hiss & Franks, 2014; Noble & Sawyer, 2002).

Second, LSAs contribute to disparity and injustice in education by informing policies and practises that are intended to benefit but really hurt students. And the impact is exacerbated for underprivileged children. For example, the findings of LSAs were the primary driver of the "achievement gap mania" in the United States (Hess, 2011). Large-scale assessments have consistently shown significant disparities in achievement amongst different groups of students. Although the gaps are mainly the result of ethnic and socioeconomic inequality, as well as other factors beyond the control of schools and teachers (Berliner, 2006; Ladson-Billings, 2006), the US government has chosen to use policies to hold educators accountable for bridging the gaps

(Berliner, 2006; Ladson-Billings, 2006). Additionally, LSAs are used to track development and hold teachers and schools accountable. As previously noted in this study, the consequences have been terrible (Nichols & Berliner, 2007; Zhao, 2018).

The policies recommended to close the gap did not succeed in terms of test scores; instead, they resulted in a limited curriculum, distorted education, and disillusioned educators. However, the harm to poor children has been greater: curriculum restriction, instructional distortion, demoralisation, and corruption scandals have occurred more frequently in schools serving underprivileged students. Reduced time spent in non-tested areas is statistically greater in poorer schools than in wealthy schools, worsening social segregation and disparities rather than closing achievement gaps (Berliner, 2011). In other words, efforts to reduce the accomplishment divide have exacerbated the opportunity divide, resulting in increased inequity and injustice (Tienken & Zhao, 2013).

Third, LSAs contribute to educational imbalance and injustice by weakening the self-efficacy of underprivileged students. Because LSAs are so narrow in their scope, they rarely catch the abilities of students from underprivileged backgrounds, but frequently communicate to these students that they are not good, as indicated by their low results. In other words, impoverished students are continuously reminded they are inadequate at everything and require remediation. These students are deprived of opportunities to demonstrate their abilities since they are compelled to complete remedial courses in order to raise their LSA scores.

Oppressive Innovation in Education

Another unintended consequence of LSAs is the stifling of educational innovation. It is widely acknowledged that today's education is out of date. Education must evolve in order to meet the challenges of technological transformation. Thus, innovation is required to improve education for the future (Barber, Donnelly, & Rizvi, 2012; Zhao, 2012). In at least two ways, LSAs hinder innovation. First, LSAs serve as a rear-view mirror for policymakers and educators, directing individuals to look backward (Zhao, 2018c). LSAs are widely used to recognise successful educational policies and practises and to urge others to follow in their footsteps. The evidence of LSA outcomes has led to the identification of several so-called evidence-based effective educational practices. Policies and practises in countries like Shanghai, Singapore, Finland, Korea, and other high-achieving educational structures have been acknowledged as worthy of copying by others based on TIMSS and PISA, see (Jensen, 2012; Schleicher, 2018; Tucker, 2011). While LSAs do show that they are effective, they aren't always effective in all aspects of education.

Large scale assessments really stifle innovation. Forcing individuals to believe that the answers to the future are already out there inhibits innovation. Furthermore, uniformity stifles innovation. Many educators and parents are convinced that great educational policies and practises already exist, and all they need to do is find and implement them where they are most needed. As a result, education policies and practises around the world are becoming more uniform (Schleicher, 2018). As a result of homogenization, there are fewer opportunities for originality and creativity (Zhao & Gearin, 2016).

Empirical Review

Teachers' Practice of formative assessment in Ghana

Bortey (2019) examined the use of formative assessment by SHS Mathematics tutors in the Cape Coast Metropolis. A quantitative method was employed, involving a descriptive survey of the entire population (census). A questionnaire on formative assessment techniques was used to collect data from all 148 mathematics tutors in the Cape Coast Metropolis. The analysis was conducted using descriptive statistics (frequency and percentages) and inferential statistics (one-sample t-test, Pearson Product Moment Correlation Coefficient, and one-way analysis of variance). The statistical tests were conducted at a significance level of 0.05. The study's findings suggested that SHS mathematics teachers in the Cape Coast Metropolis possessed limited knowledge of but a favourable attitude toward formative assessment practices. Formative assessment practises such as classwork and oral questioning were discovered to be often used.

Bekoe, Eshun, and Bordoh (2013) also conducted a similar study, examining the formative assessment strategies used by instructors to assess teacher-trainees' social studies learning in Ghanaian colleges of education. The study collected data through a case study method in three colleges of education in Ghana's Central Region, where both instructors and students were purposefully and conveniently picked for the study. Nine social studies tutors were given an interview guide and classroom observation checklists. Teachers were found to be insufficiently knowledgeable about formative assessment. Again, the study discovered that, as a result of teachers' hurry in constructing formative

assessment and scoring, tutors prioritised the cognitive domain above the affective and psychomotor domains, which were also critical.

Additionally, Awoniyi (2016) conducted research to ascertain how instructors implement school-based assessment (SBA) in the classroom. The study enrolled 110 teachers, 100 male and 10 female, in mathematics. The research instruments were a questionnaire and an interview schedule. Frequencies and percentages were used to analyse the data. The study discovered that senior high school mathematics teachers in Ghana's Cape Coast Metropolis did not have a better understanding of SBA principles and so did not use them. Appiah (2018) examined primary school teachers' practice of the School-Based Assessment (SBA) system using a descriptive survey design with both quantitative and qualitative research methodology. He sampled 240 primary school teachers using cluster and simple random sampling techniques in the Asikuma-Odoben-Brakwa District. Using a self-developed observational check-list and a self-developed questionnaire for data collection, he found that the primary school teachers practiced school-based assessment (SBA) aligned with the policy guidelines.

Amoako (2018) also conducted a meta-analysis of formative assessment practises in Ghana. The study's goal was to examine the methodology and findings of earlier studies (1990–2016) on teachers' formative assessment practice at all educational levels. Amoako observed that, on average, teachers at all levels of education adopt and practice formative assessment. However, caution must be taken in generalising this to mean that teachers have adequate knowledge, skills, and a positive attitude towards the practice of formative assessment since no empirical evidence has shown such a relationship.

Eshun, Bordoh, Bassaw, and Mensah (2014) used formative assessment to examine social studies students' learning at the selected colleges of education.

The study used a case study approach and was conducted in three colleges of education in Ghana's Central Region. The data was combined to produce a single case. The course tutors and colleges were both purposefully and conveniently chosen for the study. Nine social studies tutors were given interviewees and classroom observation checklists. The study's research topic was "to what extent does a tutor use formative assessment to evaluate student learning?" According to the findings of the study, most teachers do not employ assessment strategies to improve the teaching and learning process.

Teachers in Ghana have the perception that assessments are primarily intended to hold them accountable, which obscures the developmental role they play in students' educational progress (Oduro, 2015). As a result, testing has superseded all other forms of assessment, and teachers regard assessment as testing. "This view of assessment limited teachers' ability to explore ways to engage their students in the assessment process and to empower them to take ownership of their learning (Oduro, 2015, p. 113)." Teachers will have this perception as a result of their insufficient knowledge of formative assessment practices.

Teachers Assessment Practices from other Jurisdictions

In recent years, the measuring community has paid more attention to classroom assessment. Teaching and student learning are evaluated by teachers, so there is a lot of concern about how teachers are assessing their students. In classroom assessment, a wide range of activities are included, such as preparing paper-and-pencil examinations and performance measures, grading, analysing and interpreting the results of standardised tests, and making decisions about

test results. It is important for teachers to be aware of the advantages and disadvantages associated with various assessment techniques while employing paper-and-pencil assessments and performance measures (Stiggins, 1992). Content validity can be ensured by ensuring that test items are aligned with course objectives and instruction (Airasian, 1994), as well as by ensuring that a representative sample of instructional materials is used in the development of test items.

Validity and reliability in performance assessments can be improved by using observable and clearly defined performance tasks (Airasian, 1994; Baron, 1991; Shavelson, Baxter, & Pine, 1991), detailed rating procedures, several samples of behaviour evaluated by many judges, and recording scoring results during the assessment process (Dunbar, Koretz & Hoover,1991) (1985, Stiggins & Bridgeford). Teachers should be able to construct their own assessments using test data and item analysis (Carey, 1994; Gregory, 1996).

Students' achievements are positively related to their teachers' abilities to construct or pick high-quality classroom tests and use them effectively to aid learning, according to study (Stiggins, 2010). Teachers' opinions on teaching, learning, and epistemological principles are reflected in how they think about evaluation (Brown, 2008). There might be a lot of diversity in assessment due to various instructor beliefs (Leung, 2004). Teachers who lean toward constructivism are more likely to put the student at the centre of their assessment methods; on the other hand, teachers who lean toward behaviourism are more likely to emphasise content transmission while ignoring student needs (Carless, 2011). Similarly, instructors' perceptions of assessment play a key role in their analysis of assessment understanding and application of actual assessment

practice (Brown, 2008). Teachers are more inclined to embrace new knowledge and practises that are compatible with their assessment philosophy, while rejecting those that are not. Teachers' attitudes about assessment, whether favourable or negative, might influence the success or failure of new assessment policies, as well as the effectiveness or ineffectiveness of assessment knowledge learning. At the institutional level, the current situation in which teachers work has an impact on their assessment procedures. Although some teachers support assessment for learning principles, they are constrained by institutional values (Yu, 2015).

Due to time restrictions, teachers in a school culture that sets a high emphasis on curriculum or topic covering often feel rushed to complete the prescribed syllabus and, as a result, are less likely to conduct formative assessment activities (Carless, 2011). In practice, teachers must also meet stakeholder needs, such as learners' interests (Xu & Liu, 2009). To engage students in classroom discussion, teachers of English as a Foreign/Second Language (EFL/ESL) students with weak language skills may adjust their questioning pattern by asking less hard memory questions (Shomoossi, 2004).

In the classroom, teacher questioning has long been used to arouse students' interest, aid teacher instruction, and assess learning results (Sanders, 1966). According to research, convergent and divergent inquiries, or low and high cognitive queries, have different instructional properties. Students' factual knowledge may be assessed, their interest aroused, and they could be alerted to the facts they'll need to answer higher-order thinking problems (Musumeci, 1996). They may better engage language novices in classroom engagement since they are less frightening (Shomoossi, 2004). Random data evoked by

recall questions, on the other hand, are quickly forgotten (Sanders, 1966), and teachers' focus on these types of questions risks engaging students in rote learning and inhibits critical thinking skills development (Tan, 2007). Divergent questions, on the other hand, are more likely to imitate the development of students' thinking and encourage abilities such as synthesis and evaluation since they require students to not only memorise facts but also to reorganise information in their heads to construct an answer (Gall, 1970). Higher-order thinking is said to occur when pupils are offered increasingly challenging questions as long as they are actively engaged in thinking and reflection (Black et al., 2003).

Because classroom replies represent the externalisation of individual thought contained in language, it is assumed that they are a good predictor of pupils' current level of learning (Leung & Mohan, 2004). There is some evidence that the most prevalent responses (i.e., individual student responses, choral responses, student no responses, and instructor self-answers) do not always effectively reflect students' thinking. Teacher feedback is essential for student development, but creating good feedback that moves learning ahead is challenging. In a model that divides teacher feedback at four levels, Hattie and Timperley (2007) claim that teacher feedback at the process and self-regulation levels promotes learning more effectively (self, task, process, and self-regulation).

According to research of teachers' real practises, the less effective self- and task-levels of classroom feedback are often used. Other issues, according to Black and Wiliam (1998), include teachers just stating whether or not students' answers are correct or incorrect without any follow-up interventions

to improve learning (Leung & Mohan, 2004). Through classroom testing and assessment, many educators are able to effectively assess students' working knowledge in a variety of areas of the curriculum (Kubiszyn & Borich, 2013). Many educators disagree about the importance of classroom testing and measuring in assessing students' academic success.

Furthermore, some educators argue that mandatory high-stakes testing may lead to classroom instruction that contradicts a teacher's ideals and ideas about teaching, learning, and classroom assessment processes (Abrams, Pedulla, & Madaus, 2003). Recent study in the field of classroom testing and measurement shows that for some kids, frequent classroom testing can be highly difficult and stressful (Ramirez & Beilock, 2011). Furthermore, pupils who are uninterested in testing and external examinations perform lower on tests and external examinations than their classmates who enjoy testing and examinations in the classroom, according to the research (Ramirez & Beilock, 2011).

Hanson and Carpenter (2011) studied the impact of cooperative learning practises on classroom testing and the implications for nursing education and practice in a similar study. In nursing schools, cooperative learning practises boosted students' critical thinking and learning capacities, according to the study (Hanson & Carpenter, 2011). Furthermore, the study found that group assessments, portfolios, and other current classroom assessment procedures have a larger positive impact on students' classroom progress than regular testing and external tests. In the realms of curriculum and education, grading is a widely used and recognised concept in many countries throughout the world.

While grading is an important tool for evaluating students' progress in the classroom, it is useless unless and until it is used for its intended purpose. As a result, grading systems are useless unless used in tandem with the educational system (Nagel, 2015). As a result, teachers must be aware of the grading system's goal, the manner in which it will be used, and the mode in which grades will be used to assess each learner's degree of progress in the classroom. Furthermore, the grading system should be designed to fulfil the requirements of both teachers and students. Furthermore, the grading system should be implemented in such a way that it accurately reflects each student's level of success and growth in the classroom (Rowtree, 1987).

McMillan, Myran, and Workman (2002) assert that assessing students' work can be difficult due to the time commitment and consideration of a variety of external variables. External factors include the level of thinking of the assessment, the teachers' overall educational values, and their attitudes toward classroom assessment techniques. Additionally, it is widely recognised that the educational ideas and beliefs of instructors have a significant impact on their classroom practice (Pyle & Deluca, 2013; Zacharos, Koliopoulos, Dokimaki, & Kassoumi, 2007). According to Dobrow, Smith, and Poner (2011), teachers' values, attitudes, and beliefs can also have a direct impact on the quality of classroom feedback.

Additionally, the researchers asserted that grading can have a mixed effect on a learner's overall performance. As a result, giving students with feedback on their work's quality and progress toward mastery is crucial to the grading process. Constructive feedback on students' learning is crucial since it enables students to improve the quality of their work. Additionally, providing

feedback assists students in achieving desired learning outcomes and provides crucial direction for attaining required standards (Kubiszyn & Borich, 2013). As a result, students benefit from constructive criticism since it enables them to improve the quality and productivity of their learning (Pyle & Deluca, 2013). Unfortunately, some instructors lack the knowledge and abilities essential to provide learners with the high-quality feedback necessary to achieve the desired learning objectives.

Educators will have a difficult time giving learners with the quality feedback and direction they require to enhance the quality and outcome of their learning under these circumstances (McMillan, Myran, & Workman, 2002). Effective feedback is crucial for students' success, according to Hanson and Carpenter (2011). As a result, effective feedback should be timely, succinct, and specific about what went well and what the learner should do to further improve the quality of their work (Kubiszyn & Borich, 2013). Brookhart (1993) argues that some educators are more conventional in their mode of instruction and classroom assessment techniques, and thus are more averse to school systems' grading policies and procedures.

As a result, these educators frequently depart from established assessment and grading practises. As a result, these teachers are more receptive to using a more traditional grading system than they are to implementing more contemporary teaching and classroom evaluation practises. Additionally, these diverse perspectives on classroom assessment will have an effect on the quality of instruction and the extent to which students learn and advance in the classroom. Many classroom teachers continue to consider classroom assessment as a matter of issuing grades and administering tests. Regardless of their high

ambitions (Zacharos, Koliopoulos, Dokimaki, & Kassoumi, 2007), many teachers believe that students are blank slates ready to be filled with the knowledge and skills required for success in life. Students face significant pressure to perform well on state-mandated standardized examinations, which contributes to this perception (Dixon, 2011). Several experts, notably Cheng, Rogers, and Hu (2004), suggest that educators should prioritise students' ability to think logically and imaginatively over their ability to score well on state-mandated standardised tests.

This may occur, however, only when professors are willing to engage students in healthy debate guided by appropriate classroom evaluation methods. Teachers, on the other hand, must be able to assess learners' educational needs successfully in order to challenge them and fully involve them in relevant classroom assessment activities. Numerous studies have established that many teachers refer to classroom assessment as testing and quizzing (Campbell & Evans, 2000). Numerous scholars, including Calculator and Black (2009), claim that our sociocultural and economic conditions have an effect on our values and ideas. Jane (2013) conducted a popular study in which she assessed teachers' perspectives on classroom assessment techniques in South Africa.

The study showed that tests and quizzes were considerably more frequently used as classroom assessment tools than any other sort of assessment technique. Additionally, the study discovered that teachers' expertise, values, and opinions influenced the sorts of items chosen for tests and quizzes. Also, Jane discovered that the majority of teachers view the use of assessment for learning strategies as a time-consuming process that adds another layer of responsibility to their teaching assignment.

This could have a significant impact on how some teachers view classroom assessment and how it is implemented. Because of this, it is thought that instructors' perceptions of classroom assessment practises are influenced by their socioeconomic, cultural, religious, and political views on education (Lambert & Lines, 2000). Most teachers in the classroom use this feature in their classroom assessment practises. Igbalajobi (1983) reviewed elementary school teachers' educational and training needs and discovered a lack of training for instructors in the field of classroom assessment techniques. It is through such education that teachers would be able to assess the competencies required to aid students in achieving their goals. Emberger (2007) contends that preservice teacher education programmes do not place a high premium on instructors' classroom assessment practices, allowing some teachers to assess students in the very same manner they were assessed in college.

Continuous classroom assessment must be questioned through the introduction of novel attitudes and techniques. There is a strong case to be made that instructors' assessment practises are strongly influenced by their personal beliefs as well as their social, cultural, and environmental context (Campbell & Evans, 2010). According to an article by Bond (2011), teachers' opinions on classroom assessment are influenced by the level of training they get. Additionally, educators who have received outstanding preservice training have high aspirations for their students' academic achievement. These teachers are more likely to employ classroom assessment strategies and to take responsibility for their students' achievement than other teachers. Bond continued to say that there are a number of ways these teachers will monitor the development taking place in their classrooms.

Teachers Perception about High-Stakes Testing

Education Week released an article in which the author presented the results of a recent survey conducted by Scholastic and the Bill and Melinda Gates Foundation on teacher attitudes toward high-stakes testing (Rebora, 2012). The study drew its conclusions from a survey of more than 10,000 educators in public schools. According to Rebora, only 28% of teachers agree that standardised tests are necessary and only 26% believe that assessments effectively reflect their students' knowledge. Teachers expressed worry that assessments were not well-aligned with the current learning objectives of their students.

Davis (2011) researched the impact of standards-based reform and high-stakes testing on high school teachers' practices and opinions. High-stakes testing, according to the research, limits instructors' creativity, time, and ability to investigate topics in depth. Furthermore, the researcher revealed that, while the majority of instructors felt pressured by the implementation of high-stakes examinations, there was no significant difference in the practices of teachers who delivered the tests versus those who did not. Taylor, Sheppard, Kinner, and Rosenthal (2003) conducted a survey of about 1,000 teachers to find out how they felt about high-stakes testing and state standards. According to the study, standards have a greater impact on education than testing.

While teachers believed that state standards helped them focus their instruction on reading and math and improve their writing, their feelings about high-stakes testing were mixed. While teachers thought the tests improved their writing instruction, they also altered the curriculum (80% agreed), transferred instructional time away from non-tested topics, increased test preparation time,

and reduced faculty morale (81% agreed). Yamashita (2011) investigated how teachers in the District of Columbia Public School system altered their teaching in response to high-stakes testing. To ensure compliance with the NCLB act, teachers who administered the newly implemented assessment were interviewed. While teachers' teaching methods did not change as a result of the assessments, topic instruction did. She came to the conclusion that, while educators were under pressure to perform well on tests and changed their instruction as a result, tests alone were insufficient to impact how teachers instruct their students. Over 900 Texas educators were polled by Reese, Gordon, and Price (2004) to find out how they felt about the Texas Assessment of Academic Skills (TAAS). According to participants in a study, the examined curriculum received special attention but could not stimulate children to learn. Teachers also stated that high-stakes assessments were useless in evaluating student learning or school effectiveness.

The data showed that instructor' perceptions of test effects, proponents' arguments for testing, and studies on real pedagogy and student growth all differed, according to the data. In one component of the study, teachers' stress levels as a result of the pressure to perform well on high-stakes assessments were investigated in one component. Primary schools were the most stressed out (90 percent agreed), followed by middle schools (74% agreed), and high schools (74% agreed).67 percent. Jones and Egley (2004) conducted a study similar to this one to determine whether teachers believed that high-stakes testing programmes were moving public schools in the proper direction. The study analysed data from a survey of 708 public school instructors in Florida. Teachers from grades 3-5 were surveyed.

The participants completed an online test consisting of yes-or-no and open-ended questions. Over 79.9% of instructors stated that the high-stakes testing scheme was leading schools astray. They believed that the examinations were administered in an inefficient manner and that a single assessment of students' learning and progress was insufficient (2004). Positive and negative findings were included in the test results. According to Jones and Egley, teachers reported adverse effects on their curriculum, teaching and learning, and teacher and student motivation as a result of the assessment. Although the researchers discovered fewer positive effects, they did note that assessments held students, educators, and parents accountable for their actions.

Effects of High-Stake Testing

Using a common "yardstick" in the form of a common assessment, students across schools are treated equally and fairly (Phelps, 2008; O'Conner, 2009). It has been influenced by the results of large-scale assessments in the perception of the need for special attention for students with disabilities (Roderick & Engel, 2001). When students are given a summary of their performance, it has a significant impact on their academic success (Phelps, 2012). Teachers and administrators returning to their schools and classrooms after participating in item development, item review panels, and scoring students' responses to open-ended questions return with enhanced training and experience in item development and scoring (Cizek, 2001). Professional development conferences and training can also be evaluated using the results of large-scale assessments, Cizek added.

The findings of large-scale assessments can be used by education departments and school district staff to verify that the curriculum has been

properly addressed (Lissitz & Schafer, 2002). Having large-scale assessment results that have been made available to the public can serve as a jumping-off point for discussions about how an accountability system should be constructed and implemented in order to improve students' educational experiences and achievements (Cizek, 2001; Ferrera, 2005; Mirazchiyski, 2013; Paton, 2013). Large-scale assessments have been criticised for narrowing instructional content, reducing instructional time for test preparation activities, and including students who are no longer teachers' students in test results (Burrows, Groce & Webeck, 2005; Darling-Hammond, Earl & Katz, 2006; National Council on Measurement in Education, 2012; Shepard, 2010).

Gordon and Reese (1997) interviewed twenty Texas teachers about the effects of high-stakes testing. According to the study's teachers, the test had a negative effect on students, teachers, and schools. Some research has been conducted on teachers' perspectives on teaching exclusively to required standards, as well as their feelings of pressure and stress in this regard, as well as the effect of the standards on their motivation. Teachers face pressure, stress, and a lack of motivation to educate in meaningful ways, according to scholars who have studied their perspectives on high-stakes testing (Al-Fadhli & Singh, 2010; Donnelly & Sadler, 2009; Jones & Egley, 2004). Teachers have expressed this sentiment as a result of professional incentives for high student achievement and possible harsh consequences for poor student performance (Assaf, 2006, 2008; Brumback, 2013; Gabriel, 2010).

Educators have stated that they are under pressure to improve their students' exam scores (Bhattacharyya et al., 2013). In 2013, after it was discovered that teachers were manipulating test scores to indicate student

development, 38 states and the District of Columbia were reprimanded. Gutierrez et al., 2014; Schaeffer, 2013). At the height of a massive cheating scandal in Georgia, the district superintendent observed, "When performance pay is introduced and your evaluation may be based on how well your children perform on standardised tests, it's an incredible amount of pressure" (Gabriel, 2010, p. 4). This discovery implies that teachers' preparation for state-mandated assessments requires more than simply supporting students in earning high grades. Administrators analyse and compensate or penalise teachers based on their pupils' test scores in salary and teaching evaluations.

Increased accountability causes increased stress for teachers, which may be tied to an increase in the number of teachers quitting the field (Berryhill, Linney, & Fromewick, 2009). Teachers may follow regulations in only a rudimentary or superficial manner as a result of the stress of results-driven teaching, as well as a sense of despair and loss of control, since they believe they have lost their autonomy. British teachers' opinions about annual reporting of student performance were described by Perryman, Ball, Maguire, and Braun (2011). This belief that instructors must jump through hoops in order to meet targets may cause psychological dissonance as they lose their sense of professional autonomy (Perryman et al., 2011, p. 186). This observation shows that British educators preparing learners yearly reports felt constrained in their teaching and were unable to apply their own curriculum in favour of relying entirely on obligatory guidelines and resources.

High stakes testing accountability has been linked to a lack of drive to educate creatively (Ciani, Summers, & Easter, 2008). According to one piece of research, one of the main concerns of instructors about high stakes testing

was that it "narrowed the curriculum by requiring teachers to teach only the areas that was tested" (Jones & Egley, 2004). Teachers were expected to build their lessons around demonstrative objects that were either identical to or similar to real test items. Rubin (2011) argued that "NCLB is affecting teachers now, both in ELA [English Language Arts] and across the curriculum," and that "NCLB is weakening teachers, their practice, and their lifetime commitment to the teaching profession" (p. 407). NCLB rules, according to educators at the study school, have had a negative influence on their teaching and have forced many to consider leaving the profession.

Many teachers expressed dissatisfaction with NCLB's restrictions and an unwillingness to educate their students (Grierson, personal communication, March 25, 2014). In a 5-year study of 58 instructors in three Mississippi Delta schools, Al-Fadhli and Singh (2010) identified discrepancies in teachers' satisfaction with NCLB and motivation. The schools were located in the same area as the current research location. School A satisfied the NCLB standards, school B could not, and school C did not change, according to the Al-Fadhli and Singh study. According to the teachers at the three schools, the NCLB regulations helped to strengthen accountability in each of them. Teachers, on the other hand, had different viewpoints. Teachers at School B credited accountability for the school's positive changes. The systems, according to the teachers at School C, were extremely complex and harmful to their teaching and students. Overall, some teachers voiced concern about not having enough class time to cover curricular topics and being unable to deliver challenging topics to high achievers due to a lack of class time. Unlike earlier studies, the NCLB criteria seemed to boost teachers' motivation to teach, with School C teachers

reporting the least motivation (Al-Fadhli & Singh, 2010). The conclusions of the study indicate that teachers have doubts regarding NCLB. However, not all of the teachers believed the NCLB standards had a negative impact on their teaching; rather, they all said the requirements helped them be more accountable and increased their motivation to educate.

CHAPTER THREE

RESEARCH METHODS

Introduction

The purpose of the study was to explore senior high school teachers' perceived influence of large-scale testing accountability on teaching and learning at Birim Central Municipal. This chapter explains the research methods used in this study.

Research Design

The term "research design" refers to the fundamental framework of a study, the nature of the hypothesis, and the variables studied (Gay, 1992). Research design, according to Mouton (2001), is "a plan or blueprint for how the research would be conducted." It establishes a framework for any study to follow. As a result, it represents the data collection and analysis technique for a certain concept. The study design was a descriptive survey. Babbie (1990) endorsed the descriptive survey for the purpose of generalising from a sample to a population in order to make judgments about the population's features, attributes, or behaviour. A descriptive survey is a data collection technique used to test hypotheses or address research questions concerning the study's status (Gay 1992). A descriptive survey's primary purpose is to observe, describe, and document the characteristics of a situation as it naturally occurs.

This strategy incorporates a variety of data collection techniques, including pre-testing, questionnaires, observation, interviews, and document review (Amedahe & Asamoah-Gyimah, 2003). According to Polit and Hunger (1995), descriptive surveys are primarily concerned with describing, observing,

and documenting characteristics of a situation as it unfolds naturally, rather than with explaining it. According to them, the design offers the advantage of eliciting a large number of responses from a diverse range of individuals. A descriptive survey entails the distribution of a set of questions to a large group of people through the mail, telephone, or in-person. Simultaneously, it presents a more realistic picture of events and attempts to explain human perception and behaviour using facts obtained at a certain point in time. It is appropriate when a researcher wishes to describe a population's characteristics through the use of unbiased samples of individuals who complete questionnaires, interviews, or exams (Fraenkel & Wallen, 1993).

On the other hand, if stringent safeguards are not implemented, descriptive survey data may be distorted as a result of the introduction of bias into the research design (Amedahe & Asamoah-Gyimah, 2003). Another problem with descriptive surveys is that, while they rely on direct observation for data collection, they require data organisation and presentation in a systematic manner before reliable conclusions can be formed (Jacobs, 2011). The design is appropriate for the study because it is intended to elicit views about the impact of large-scale testing accountability on teaching and learning and because it allows for the collection of massive amounts of data on the subject in a short period of time. The key problem with the design is demand characteristics, as respondents attempt to provide answers that mirror their perceptions of the responses desired by the researcher. Despite this disadvantage, the optimum design for this study was chosen to be this one.

This is a descriptive study, as it was conducted to ascertain senior high school teachers' perceptions of the influence of large-scale testing accountability on teaching and learning at Birim Central Municipal High. Birim Central Municipal Assembly is in Ghana's Eastern Region, in the southwestern portion. To the north, the municipality borders Akyemansa and Kwaebibirem; to the west, Birim South District; to the south, Asikuma-Odoben-Brakwa and Agona East Districts; and to the west, West Akim (to the east). It covers a total of 1,090 km^2 of land. According to the 2010 Population and Housing Census, the Birim Central Municipality has a population of 144,869 people, accounting for nearly 6% of the total population of the Eastern region. Males make up 47.1% of the population, while females make up 52.9%, resulting in a sex ratio of 92 (number of males to 100 females) .The municipalities is located in the wet semi-equatorial climatic zone, which means it receives a lot of rain. The climatic conditions in the municipality are similar to those seen in Ghana's central belt. The average monthly temperature in the municipality is approximately 26 °C, with daily temperatures ranging from 21 °C to 35 °C.

Every year, there is a double maximum rainfall. The major rainy season lasts from April to July, whereas the minor season lasts from September to early November. The wettest month of the year is June. Humidity is highest during the rainy season and lowest during the dry season. The average relative humidity in the municipality is around 80%. (Ghana Statistical Service, GIS, 2010).

Figure 1: Map of Birim Central Municipal

Source: Ghana Statistical Service, GIS

Population of the Study

In research, "population" is defined by Gay (1992) as the group in which the researcher is interested and to which the researcher wishes to generalise the study's results. He went on to add that the defined population has at least one characteristic that distinguishes it from other groups. A study population is defined by Polit and Hungler (1996) as "the totality of cases that match a prescribed set of criteria." The researcher wants to generalise his findings to all of the participants who took part in the study. According to Johnson (1994), participants in a population must have the data needed for the study.

This study's target population is all senior high school teachers in Birim Central Municipality. The study's accessible population was Birim Central

Municipality's public senior high school teachers. The municipality operates three (3) senior high schools with a total teacher population of 347.

Table 1: Distribution of Population

SHS	Number of SHS Teachers
Akim Oda SHS	126
Attafuah SHTS	117
St Francis SHTS	104
Total	347

Source: GES, Birim Central Municipality, 2019

Sample Size and Sampling Procedure

According to Sarantakos (1998), a sample is a carefully chosen portion of the population.. Sarantakos added that, it is through the use of samples that the researcher hoped to generalise the findings of a study. The techniques and procedures used to select samples from the target population are known as sampling techniques and procedures. The method of choosing a subset of a population to represent the entire population is known as sampling (Polit & Hungler, 1999). According to Amedahe (2002), a sample size of 5% to 20% of the population is acceptable for generalisation, hence a sample size of 200, representing 57.6% of the population, is highly adequate for generalisation.

It was also necessary to use Babbie's (2001) formula to determine the sample size for each school in order to obtain a comparable representation of teachers from each of the participating schools.

The formula is

$$s = \frac{(n)}{N} \times k$$

Where (s) stands for the sample to be selected from a school;

(n) stands for the entire population of the school;

(N) stands for the size of the target population (347); and

(K) stands for the sample size (200).

Hence, for example, Akim Oda Senior High School whose teachers' population was given as 126, applying the Babbie's (2001) formula, gave the outcome as shown below:

$\frac{126}{347} \times 200$

=73 approximately

Table 2: Sample and Sampling Procedure

SHS	Number of SHS Teachers
Akim Oda SHS	73
Attafuah SHTS	67
St Francis SHTS	60
Total	200

Source: Field Survey

All three public senior high school teachers were considered for the study, excluding non-professionals. Proportionate sampling was used to select 73 teachers out of 126 teachers in Oda SHS, 67 teachers out of 117 teachers in Attafuah SHTS and 60 teachers out of 104 in St. Francis SHTS, giving a total

of 200 selected teachers out of 347 for the study. I decided to employ the lottery method. On sheets of paper, the names of all 126 teachers at Oda SHS were recorded.

After putting the papers in a container and shaking them vigorously, I chose 73 teachers. The same procedure was used to choose 67 and 60 teachers from Attafuah SHTS and St. Francis SHTS, respectively I used simple random to ensure that each teacher had an equal probability of being chosen.

Data Collection Instruments

Instrumentation is the process of developing tools for collecting data in the field. These include questionnaires and an interview schedule, to name a few. Although a variety of instruments could have been used to collect data, a questionnaire was determined to be the most appropriate for the study. Researchers and respondents alike save time because questionnaires are simple to administer, simple to complete, and quick to score (Knowles, 1980). Despite the many advantages of questionnaires, dishonesty can be a concern. This means that respondents may not tell the truth in their answers. This may be due to social desirability bias or a desire to keep information private. By assuring respondents that their privacy is protected and that the process does not permit personal identification, I was able to reduce dishonesty. Additionally, it has the weakness of respondents skipping complicated questions, which can affect the study's outcome. This notwithstanding, I simplified my questions to avoid question skipping and thus increase completion rates. I began developing the questionnaire by reading extensively on the impact of large-scale testing accountability on teaching and learning. I adapted the questionnaire used in this study.

This study adapted a teacher survey on high-stakes testing. The National Board on Educational Testing and Public Policy developed, validated, and initially used the instrument (Pedulla et al., 2003). Four sections comprised the questionnaire: A, B, C, and D. Section A was designed to elicit demographic information about the teachers. Section B, comprised of 10 items, was used to elicit information about teachers' assessment practises. Section C contained 20 items that gathered data from teachers' surveys regarding the influence of large-scale testing accountability on teaching and learning. Finally, a self-developed questionnaire that formed section D and was comprised of 8 items elicited information on ways to improve teaching and learning in this era of large-scale testing dominance.

Validity

As a result of selecting a representative sample from the domain of indicators for the concepts, my data was subjected to content validity. The term for this is sampling validity. Validity is defined by Mugenda and Mugenda (1999) as the accuracy and significance of inferences drawn from research findings. Content-related evidence aids in determining whether the instrument's content contains an adequate or appropriate sample of the domain it is meant to represent or reflect. To ensure the instrument's validity, content-related evidence and face-validity were used. This was accomplished by submitting the questionnaire to an educational research expert. The instrument's validity and reliability were determined in order to ensure that it is internally consistent.

Pilot-testing and Reliability

The instruments that were used were pilot-tested by me. According to Donald (1990), pilot testing helps the researcher determine if the study is

feasible and worthwhile to undertake, as well as the suitability and practicality of the data collection instrument. The instrument was pilot tested to check for understanding and ambiguity, as well as to correct any misunderstandings that may have arisen as a result of the items' framing and construction. Pilot testing of the adapted instruments as well as self-developed instruments was done at Kade Senior High/Technical School in Kwaebibirem Municipality, which borders Birim Central Municipality. I chose Kade Senior High/Technical School because the participants (teachers) share many of the same characteristics as the study participants. Cronbach's alpha was 0.83 for the questionnaire's reliability statistic.

Data Collection Procedure

The data were gathered from senior high school instructors in Birim Central Municipality. An introductory letter from the Head of the University of Cape Coast's Department of Education and Psychology, requesting the cooperation of the selected senior high school headmasters and headmistresses in data collection. One teacher was trained and engaged as a research assistant in each school to assist me with administration and data collection. I briefed the respondents on the study's objectives and the importance of providing honest responses to the items during the administration of the instruments, as the study is for academic purposes and will also be useful to the school, teachers, and students. The data from the respondents took 7 weeks to collect. The reason for this is that all three schools run a double system. When the "Gold Track" students vacated, some of the respondents took the questionnaire home with them. Upon arrangement, I travelled to the towns in which some of the

respondents were staying to collect the questionnaires. Others too. I had to wait until the schools reopened. I had a 94% return rate.

Ethical Consideration

The correct rules of conduct required when conducting research are referred to as research ethics. It emphasizes the importance of participants being informed about the purposes, objectives, and possible negative consequences of their participation (Seidman, 2006). Additionally, it states that they have the right to revoke their consent at any time after it has been given. Informed consent, according to Cohen (2000) and Mertens (2010), is inferred from the participant's right to freedom. Researchers have an ethical obligation to ensure the safety of their participants. The researcher is fully responsible for conducting ethical research. Researchers must make every effort to ensure that the study does not have an adverse effect on the physical, social, or psychological well-being of research participants. When possible, research relationships should be characterised by mutual respect and trust. Prior to participating in this study, each participant was informed of the purpose of the study.

Punch (2008) believes that researchers should be ethically conscientious, particularly in social research that involves the collection of personal data. Moral considerations and respect for participants are critical in social research. As a result, this study considered several ethical issues. The study addressed all ethical concerns, including informed consent, anonymity, and confidentiality.

Participants were informed that involvement in the study was completely voluntary and that they could withdraw at any time. The anonymity

of study respondents was given a lot of thought in this study. According to Oliver (2010), anonymity is a critical ethical issue in research because it enables participants to maintain their anonymity. For the objectives of this study, fictional names that could not be linked back to the participants were used for identification. To prevent breaching participants' privacy needlessly, I visited schools prior to the start of data collection to acquire their consent. To adhere to the ethical principle of anonymity, no respondents' names or personally identifiable information were collected. This safeguards respondents from victimisation if certain responses are deemed unpalatable by other stakeholders.

Concerning confidentiality, every attempt was made to keep the participants' responses secret. Participants were informed that their responses would remain confidential, that no one they knew would have access to the information they supplied, and that their names would not be included in the study.Most crucially, in terms of ethical concerns, information cited from previous studies on the influence of large-scale testing accountability on teaching and learning was properly acknowledged through citation and referencing to avoid academic dishonesty, often known as plagiarism.

Data Processing and Analysis

The data analysis process provided me with facts and figures that enabled me to understand the study's findings and make statements about them. The data was compiled and updated in order to resolve unanswered questions, either partially or completely. The questions were serially numbered to simplify their identification, presentation, and analysis statistically. This precaution must be taken to guarantee that minor errors in data tabulation are detected promptly. Additionally, responses to the various questionnaire items were collected,

collated, and analysed statistically. Each item on the questionnaire was coded. Items were scored on a four-point Likert scale (4–1), with 4 being the highest and 1 being the lowest.

The background information of the participants was analysed using percentages and frequencies. Data on research question 1and 3 were analysed and answered using means and standard deviations. To answer research question 2, I used exploratory factor analysis (EFA) with principal component analysis. Hypothesis one was tested using an independent t-test, while hypothesis two was tested using an ANOVA. The responses to research questions one and three were analysed using means and standard deviations. Because the scale has a maximum value of 4 and a minimum value of 1, the average was 2.5. I used this value as my test value, 2.5, to determine how teachers use formative assessment and how teaching and learning can be improved.

Exploratory factor analysis was used to better understand the structural patterns arising from teachers' responses to the influence of large-scale testing accountability on teaching and learning. The data was inspected before starting the factor analysis to ensure that it was suitable for factor analysis. Because it is appropriate for short-scale response options and small sample sizes, principal axis factor analysis was used (Bandalos & Finney, 2018). The component correlation matrix was examined using oblimin (oblique rotation). The 20 items that examine teachers' perceptions of large-scale accountability in teaching and learning were subjected to exploratory factor analysis. The Kaiser–Meyer–Olkin measure, KMO =.634 (or "middling," according to Hutcheson & Sofroniou, 1999), validated the sampling adequacy for the analysis. This

indicates that the sample size was sufficient for factor analysis. The result of Bartlett's test was significant (.000), indicating that the sphericity assumption was met.

Table 3: Kaiser-Meyer-Olkin Measure of Sampling Adequacy.

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.634
Bartlett's Test of Sphericity	Approx. Chi-Square	1001.581
	df	190
	Sig.	.000

Furthermore, eigenvalues for each factor in the data were obtained through a preliminary analysis. Seven factors had eigenvalues over Kaiser's criterion of 1, explaining 64.04% of the variance. In order to determine the best factor structure pattern, Monte Carlo PCA for Parallel Analysis (Watkins, 2000) and scree plot were used. The results of this analysis supported three factors, explaining 38.54% of the variance and with eigenvalues greater than the corresponding criterion values in a random data set, generated with a matrix of the same size (20 variables 200 participants).

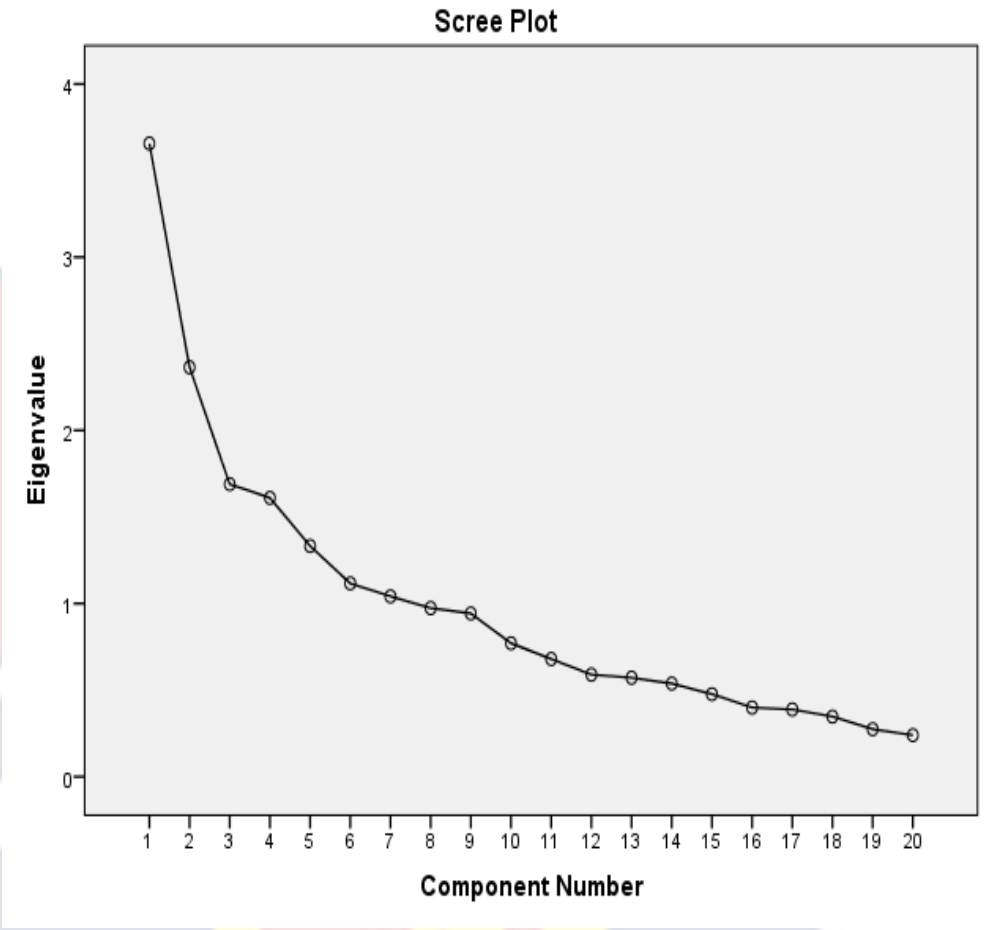


Figure 2: Scree Plot

Because factor loadings below 0.4 did not represent significant values, they were suppressed (Field, 2013). Items (12, 13, 7, 5, and 3) did not load onto any factor and were thus deleted. The factors were named based on the items that clustered around the three factors. Internal consistency (i.e., Cronbach's alpha) was more than 0.70 for all three components, signifying high reliability (Field, 2013).The three-factor structure was used to understand teachers' perceived influence on large scale testing accountability on teaching and learning.

Hypothesis :One-way analysis of covariance was used to test whether significant differences existed in demographics (i.e., teaching experience). ANOVA was deemed appropriate because the demographics are categorical

with three or more independent (unrelated) groups. Moreover, an independent t-test was used to test if a significant difference existed in the teachers' educational qualifications. Because teachers in my study indicated only bachelors and master's degrees, an independent t-test was deemed fit for testing this hypothesis.

Chapter Summary

This chapter focused on the research design, population, sampling procedure, data collection devices and procedures, and data processing analyses**.** The descriptive research design was deemed appropriate for this study, and the population for this study was public senior high schools in Birim Central Municipality. To sample participants for the study, simple random sampling, specifically the lottery method, was used. For background information, percentages and frequencies were used, while means and standard deviations were used to analyse all research questions and ANOVA for hypotheses.

CHAPTER FOUR

RESULTS AND DISCUSSIONS

Introduction

The study's aim was to explore senior high school teachers' perceptions of the influence of large-scale testing accountability on teaching and learning at Birim Central Municipal. This chapter is about the presentation and analysis of the data that I gathered from the senior high school teachers who participated in the study. According to the research questions, the data was analyzed and discussed. Teachers from the three public senior high schools in the Birim Central Municipality participated in the study. Frequency distributions, percentages, standard deviations, and means, as well as ANOVA, were used to analyse the data. The demographic characteristics of respondents are described in the first section of this chapter. The research findings are presented in five sections in the second part, according to the research questions and the hypothesis formulated

Demographic Characteristics of the Respondents

Gender, educational level, and teaching experience were all considered as demographic characteristics of teachers. This demographic information aided in the understanding of the types of respondents included in the study. Table 4 summarises the background information on the respondents who were considered for the study.

Table 4: Demographics of the Participants

Demographic Variables	Frequency	Percentage
Gender		
Female	51	25.5
Male	149	74.5
Educational Background		
Diploma	0	0
Bachelor's Degree	154	77
Master's Degree	46	23
PhD Degree	0	0
Years of Experience		
0-5 years	20	10.0
6-10years	166	83.0
Above 10years	14	7.0

Source: Field Survey, 2019

As shown in Table 4, 25.5% (51) of the 200 senior high school teachers who participated were female, while 74.5% (149) were male. The majority of teachers (154, representing 77%) held a bachelor's degree. Additionally, when asked about their teaching experience, the majority of teachers indicated that they had taught for between 6 and 10 years. This demonstrates that the teachers in this study possess the necessary classroom experience to respond appropriately to the question and provide credible information about large-scale testing accountability.

Results

This section discusses the field data in connection to the study's research questions. The four-point Likert scale questionnaire was analyzed using means and standard deviations.

Research Question 1: What are Birim central Municipal senior high school teachers' assessments practices.

To address this research question, I asked teachers from selected senior high schools to estimate their proficiency with each assessment practise and circle the appropriate number. For instance, if they believe their abilities in "Explaining assessment results clearly to parents" are "very high," you would circle number 4. Alternatively, if they believe their assessment skills are "very low," they would circle number 1.A mean of 2.50 and above indicates teachers' agreement of having adequate classroom assessment skills needed to move teaching and learning forward, while a mean below 2.50 shows that teachers do not have adequate classroom assessment skills needed to move teaching and learning forward. Table 5 shows the responses of the teachers who took part in the study.

Table 5: Senior high school teachers' assessments practices

Statements	Mean	Std. Deviation	Kurtosis	
				Std. Error
"Selecting multiple methods of assessment (e.g., tests, observations, projects)"	3.57	.638	.586	.342
"Creating assessment tasks that accommodate the needs of a variety of students"	3.72	.803	-.279	.342
"Developing assessment tasks with different formats (e.g., multiple-choice, fill-in-blank, short answer)"	3.94	.812	-.973	.342
"Sampling from the domain defined by learning goals to write assessment items"	3.43	.712	-.175	.342
"Determining if an assessment is aligned with required standards (national curriculum goals)"	3.37	.905	-.079	.342
"Interpreting summary scores reported with standardized test results (e.g., mean, percentile rank)"	3.00	1.017	-.832	.342
"Recognizing when assessment results are being used inappropriately by others"	3.20	.837	.006	.342
"Revising test items based on item analysis"	3.24	1.125	-.711	.342
"Using the results of formative assessment to adjust future lesson plans"	3.98	.824	.178	.342
"Explaining assessment results clearly to parents for decision making"	3.24	1.014	.028	.342

Source: Field Survey, 2019

The results from Table 5 clearly indicate teachers' agreement of having adequate classroom assessment skills needed to appropriately use classroom assessment to improve teaching and learning. Although teachers generally agree that they have adequate assessment skills, the results show that teachers are

highly skilled in adjusting future lesson plans based on formative assessment results (M = 3.98, SD =.824).Furthermore, Birim Central Municipal senior high school teachers agreed that they are extremely skilled at developing assessment tasks in a variety of formats (e.g., multiple-choice, fill-in-the-blank, short answer) (M = 3.94, SD =.812).Notwithstanding, the results show that teachers are not highly skilled in interpreting summary scores associated with standardised test results (e.g., mean, percentile rank). This assessment skill had the lowest mean among all the other skills (M = 3.00, SD = 1.017).

Research Question 2: What are the perceived influences of large-scale testing accountability (high-stake) on teaching and learning?

Table 6 present the result of the exploratory factor analysis

Table 6: Exploratory factor analysis of perceived influences of large-scale testing accountability (high-stake) on teaching and learning

Items	Factors		
	Motivate Teachers and students	Pressure on Teachers	Pressures on Students
"14. Large -scale testing has increased teachers' morale in my school"	.745		
"1. Scores on the large-scale testing accurately reflect the quality of education students have received"	.715		
"2. The large-scale testing motivates previously unmotivated students to learn"	.702		
"15. High stake test is an appropriate way to assess what student know"	.544		

"16. The national curriculum is aligned with the large - scale testing"	.488		
"17. Ensure teachers are evaluated based on their students' scores on high-stakes tests".	.446		
"9. Teachers in my school want to transfer out of form three where the large -scale testing is administered"		.716	
"11. Administrators in my school believe students' large-scale testing scores reflect the quality of teachers' instruction"		.676	
"10. Teachers feel pressure from parents to raise scores on the large- scale testing"		.661	
"8. There is so much pressure for high scores on the large- scale testing that teacher only teach to the test."		.554	
"6. Teachers feel pressure from the headmaster to raise scores on the large - scale testing"		.508	
"18. The use of large-scale tests has resulted in an increase in students' test anxiety"			.697
"20. On average, teachers spend at least half of their instructional time (or more) preparing their students for large-scale tests"			.647
"19. Large-scale testing programme has led teachers to teach in ways that contradict their own ideas of good educational practice"			.401
Variance explained %	*18.28*	*11.82*	*8.44*
Reliability	*0.743*	*0.702*	*0.714*

Note. Extraction method: Principal Axis Factoring. Rotation method: Oblimin with Kaiser Normalization. Factor loadings lower than .4 were suppressed.

Table 6 presents the factor structure and loadings, reliability coefficients (i.e., Cronbach's alpha), and variance explained. A three-factor structure was identified to understand teachers' perceived influence of large-scale testing accountability on teaching and learning. Factor 1, which was termed as *motivation to teachers*, was centered on 6 items, which implies that teachers perceive large-scale testing accountability as a way of motivating teachers and students. Teachers perceive large-scale accountability as their morale in school. Also, large-scale testing motivates previously unmotivated students to learn. Factor 2 was termed "*pressure on teachers."* This factor was made up of 5 items, which implies that large-scale accountability puts pressure on teachers. Teachers perceive that large-scale testing accountability forces them to teach to the test. Teachers further indicated that they feel pressured by head teachers to raise students' scores. Moreover, school administrators and parents also put pressure on them to increase students' scores. Finally, factor 3 was termed "*pressure on students."* The third factor was made up of 3 items. Teachers indicated that large-scale accountability put pressure on students, thereby increasing their test anxiety. Teachers also spend extra time preparing students for the test, which consequently increases the number of instructional hours.

Research Question 3: How can teaching and learning be improved in this era of large scale testing accountability in Senior High Schools in Birim Central Municipality

Table 7: Improving teaching and learning in this era of large scale-testing accountability (high-stake)

Statements	Mean	Std. Deviation	Skewness	
				Std. Error
"Results of large-scale testing should not be seen as the sole basis for teacher effectiveness"	3.35	.787	-1.321	.172
"Items on the large-scale testing should accurately reflect the quality of education students have received"	3.18	.538	.123	.172
"Large scale testing results data should be used to improve teaching and learning".	3.34	.588	-.252	.172
"Large scale testing results should be used to examine school policies related to curriculum and instruction".	3.18	.556	.048	.172
"Large scale testing results should be used to encourage teachers' professional development to improve instruction".	3.15	.577	-.014	.172
"Large scale testing results should be used in attracting additional resources for students with greatest needs"	3.23	.613	-.700	.172
"Large scale testing results should be reported in a manner that encourages	3.02	.511	-.195	.172

professional collegiality"				
"Large scale testing results should be used to promote better allocation of state educational resources"	3.16	.786	-.856	.172

Source: Field data, 2019.

The results from Table 7 clearly shows teachers agreement to the statements about how large-scale testing can be leveraged to move learning forward. Teachers agreed that large-scale testing results should not be used as a sole basis to determine teacher effectiveness (M=3.35, SD=.787). Rather, they suggested that large-scale testing results be used to promote better allocation of national educational resources (M=3.16, SD=.786) and to urge teachers to improve their professional development in order to improve instruction (M=3.15, SD=.577).

Moreover, teachers indicated that large-scale testing results should be used to evaluate curriculum and instruction policies in schools (M = 3.18, SD =.556).Teachers further suggested that large scale testing results should be reported in a manner that encourages professional collegiality (M=3.02, SD=.511) and foster unity among schools, teachers and administrators.

Results from Research Hypotheses

This section presents the analysis of the hypothesis.

Research Hypothesis One: H_0: I sought to find out whether statistically significant differences exist in teachers' assessment practices based on academic qualification. To achieve this, independent sample t-test was deemed appropriate. This is because the respondents had either bachelor's degree or master's degree. The results are presented in Table 8 below.

Table 8: Result of t-test Analysis of teachers' assessment practices on the basis of academic qualification.

Gender	n	Mean	Std. D	df	t-value	Sig
Bachelor's degree	154	34.0325	5.54885	198	-3.381*	.001
Master's degree	46	37.1333	4.90640			

Source: Field data, 2019. * Significant, p< .05 (2-tailed)

Table 8 presents result of t-test analysis of teachers' assessment practices on the basis of academic qualification. From the Table 8, the t- test is significant at 0.05 level of significance. I therefore fail to reject the null hypothesis which stated, that "there is no statistically significant difference in teachers' assessment practices on the basis of academic qualification" *t* (198) = -3.381*, sig= 0.001, p < .05. Therefore, the alternate hypothesis that "there is statistically significant difference in teachers' assessment practices on the basis of academic qualification" was upheld. This therefore implies that there is statistically significant difference in teacher assessment practices on the basis of academic qualification. From the descriptive statistics, teachers with master's degree use assessment results appropriately (M=37.133, SD=4.906) than teachers with bachelor's degree (M=34.03, SD=5.548).

Research Hypothesis Two: H_0:

There is no statistically significant difference in teachers' assessment practices on the basis of years of teaching experience. I sought to determine whether there are statistically significant differences in assessment practices among teachers based on their years of experience in teaching. This was accomplished using a one-way analysis of variance (ANOVA). The following

assumptions were verified before running the one-way ANOVA: Normality, linearity, and homogeneity of variance were among the assumptions. The normality test is shown in Table 9, and the homogeneity of variance is shown in Table 10.

Table 9: Tests of Normality

	Years of Experience	Shapiro Wilk Test		
		Statistic	df	Sig.
STAP	0-5years	.166	20	.149
	6-10years	.076	166	.051
	Above 10years	.221	14	.063

a. Lilliefors Significance Correction

From Table 9, the Significance value (Sig) for Shapiro Wilk Test is more than the alpha or critical value of p= 0.05 shows that the data is normally distributed at .05 alpha level.

Table 10: Test of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
STAP	Based on Mean	.279	2	197	.757

From Table 10, the Significance value (Sig) for Levene' test is 0.757 which is more than the alpha or critical value of p= 0.05 shows that the assumption of homogeneity has not been violated for this sample that is [F (2, 197) = .279, p= .757] at the .05 alpha level.

Table 11: Summary of ANOVA results Comparing in teachers' assessment practices on the basis of years of teaching experience

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	69.425	2	34.712	1.119	.329
Within Groups	6111.355	197	31.022		
Total	6180.780	199			

* Significant, p< .05 (2-tailed)

From the one-way Analysis of Variance, *F (2, 197) =1.119, Sig. = .329, p. < 0.05*. From the analysis of the one-way variance in Table 11, the Sig. value of .329 is more than the p. value of 0.05 *(p< 0.05)* this means that there are no significant differences between means of the teachers' years of experience concerning their assessment practices.

Discussions

The study investigated senior high school teachers' perception of the influence of large-scale testing accountability on teaching and learning at Birim Central Municipality. Three research questions and two hypotheses were tested. These were; 1) what are senior high school teachers' assessment practises at the Birim Central Municipal? 2) What are the perceived influences of large-scale testing accountability on senior high school teachers' teaching and learning in Birim Central Municipal? 3) How can teaching and learning be improved in this era of large-scale testing accountability in senior high schools in Birim Central Municipal? There is no statistically significant difference in teachers' perceived

influence of large-scale testing accountability on teaching based on academic qualification. Based on years of teaching experience, there is no statistically significant difference in teachers' perceptions of the influence of large-scale testing accountability on teaching.

The study clearly indicates that teachers have adequate classroom assessment literacy and the skills needed to appropriately use assessment to improve teaching and learning. In lieu of this, teachers are highly skilled in developing assessment tasks with different formats and using the outcomes of formative assessment to modify future lesson plans. The findings support a meta-analysis conducted by Amoako (2018) on formative assessment practises in Ghana. Amoako (2018) concluded that, on average, formative assessment is embraced and practised among teachers at all levels of education. Notwithstanding, findings from this study contrast with the work of Bortey (2019), who found that secondary school mathematics teachers in the Cape Coast Metropolis had low knowledge levels but a positive perception of formative assessment practices. Similarly, Bekoe, Eshun, and Bordoh (2013) discovered that teachers lack adequate formative assessment knowledge.

The study discovered that, as a result of teachers' haste in developing formative assessment and rating, teachers overemphasised the cognitive domain at the expense of the equally essential affective and psychomotor domains. Awoniyi (2016) discovered that mathematics teachers at senior high schools in Ghana's Cape Coast Metropolis lacked a deeper knowledge of SBA principles and hence failed to implement them. Additionally, Eshun, Bordoh, Bassaw, and Mensah (2014) used formative assessment to examine social studies students' learning in selected colleges of education in Ghana's Central region. The data

indicates that the majority of teachers do not employ assessment tools to enhance the teaching and learning process.

Regarding teachers' perceptions of large-scale testing accountability, the study found that although teachers perceive large-scale testing accountability as a motivation and an increase in their morale in schools, they indicate that they feel pressured by such accountability. Teachers aren't the only ones feeling the pressure; students are as well. A theory known as the "principal-agent problem" can be supported by these results. Publicly reported data is used to hold schools accountable for the success and achievement of their pupils (Figlio & Loeb, 2011; Supovitz, 2009; Rothstein, Jacobsen, & Wilder, 2008).However, data-driven accountability has unintended negative consequences for the agents (i.e., teachers) (Grupe & Nitschke, 2013; Smith & Amick, 1989). For example, Smith and Amick (1989) found that accountability put too much stress on teachers. Similarly, Grupe and Nitschke (2013) maintained that close monitoring and punishments attached to accountability lead to stress and anxiety. Further, data-driven accountability has also been found to cause a loss in job autonomy (Jerrim & Sims, 2021). Thus, teachers do not have the liberty to work due to data-driven accountability. Finally, data-driven accountability forces teachers to do more work, such as preparing students for tests (Perryman & Calvert, 2019). This adds extra work to teachers and puts them under pressure, consequently adversely affecting classroom teaching and learning.

This result confirms the work of (McMillan, 2005), who found that when LSA results are used for making high-stakes decisions, it encourages teachers to teach to the test, with the primary purpose of increasing students' scores on external tests. The results further support the work of Kellaghan, Greaney, and

Murray (2009), which pointed out that students' performance on LSAs is used to evaluate the quality of teachers, schools, and the overall worth of students.

According to Kellaghan, Greaney, and Murray, teachers face pressure to raise academic success and the threat of losing their contracts if they fail to make academic gains. This finding also corroborates prior study (Browne, 2016; Kellaghan, Greaney, & Murray, 2009), which discovered that teachers are also under pressure to demonstrate to parents that they are preparing their students to pass critical exams that lead to the next level of schooling.In lieu of the competitive nature often associated with LSA and the newer phenomenon of teacher evaluation based on students' LSA test results creates further pressure for teachers to 'teach to the test' which is generally referred as washback effect (Carless, 2011).

The findings support previous research that found that high-stakes tests elevate learners' risk of academic failure, hold learners and educators with unequal resources to the same standards, limit and distort the curriculum, and intensify class and racial disparities. Additionally, they argue that high-stakes testing conveys the idea that the prime objective of education should be to score well on examinations (2005Bracey, 2000; Burger & Krueger, 2003; McMillan, 2005). Regardless of one's feelings about the value and utility of large-scale testing, it is indisputable that these programmes have had a substantial impact on learners, educators, and school administrators.

For example, some states have developed large-scale assessments that hold headteachers and students accountable. In these systems, students do not graduate unless they pass an exit exam, and headteachers do not get raises unless student achievement improves to a predetermined level. Headteachers feel the

pressure to improve student achievement and fear losing their contracts if they fail to make academic gains (Kellaghan, Greaney & Murray, 2009).

These findings back up the findings of researchers who discovered that large-scale testing generates the necessary pressure for students to work more and teachers to adopt more effective approaches (Carnoy & Loeb, 2002; Hamilton & Gonzales, 2003; Hess, 2002; Roderick, Jacob & Bryk, 2002). They argue that high-stakes assessments give students and teachers more information about their own knowledge and skills, as well as a better understanding of individual students' strengths and weaknesses (Carnoy & Loeb, Hamilton & Gonzales, Hess, Roderick, Jacob & Bryk). Other experts agree that by holding schools accountable for student accomplishment, high-performing schools can be praised while low-performing schools can be targeted for further assistance and resources (Barnes, 2005; American Psychological Association, 2001; Jones & Egley, 2007; Stecher, 2002).

While teachers suggested that large-scale testing should be improved, they emphasised that large-scale testing should not be viewed as the main determinant of teacher effectiveness and school quality. Large-scale testing results should instead be used to advocate for a more equitable distribution of state educational resources and to support teachers' personal and professional growth so that instruction may be improved. Additionally, the results of large-scale testing should be used to analyse school policies on curriculum and instruction and to advocate for a more efficient allocation of state educational resources. Additionally, the study discovered that teachers' educational background has a significant impact on how they assess literacy or knowledge. Teachers with a master's degree have a greater understanding of assessment than

those with a bachelor's degree. The study, however, discovered no statistically significant differences in teachers' assessment practises based on their teaching experience. This finding corroborates the findings of Asare (2021), who discovered that teachers' experience had no effect on the formative assessment practises in the classroom by basic school teachers.

CHAPTER FIVE

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

Introduction

This chapter deals with the summary, conclusion, and recommendations based on the findings. It summarizes the research methods employed in collecting data needed to address the research questions and hypotheses formulated. Based on the findings, conclusions were reached, which enabled recommendations to be made to appropriate authorities. Also, suggestions are made for further research.

Summary

The purpose of the study was to assess the influence of large-scale testing accountability on teaching and learning. However, the study specifically sought to find out Birim central municipal senior high school teachers' classroom assessment practices, teachers' perceived influences of large-scale testing accountability (high-stake) on teaching and learning, and how teaching and learning can be improved in this era of large-scale testing accountability (high-stake). Again, the study sought to find out whether statistically significant differences exist in teachers' assessment practises on the basis of academic qualification and teaching experience.

The study was conducted using a descriptive survey design. Senior high school teachers in Birim central municipal constitute the study's target population. To select the teachers from the schools, the researchers used simple random sampling, namely the lottery method. Furthermore, Babbie's (2001) formula was utilised to determine the sample size for each of the three schools. The teachers' responses were elicited using a questionnaire. Descriptive

statistics (frequency and percentages, means and standard deviations) were used to analyse the data collected on the research questions, and inferential statistics (Exploratory factor analysis, independent t-test, and ANOVA) were used to test the hypotheses and answer research question two.

Key Findings

The following are the study's key findings:

1. The study's findings indicate that senior high school teachers in Birim Central Municipality have adequate classroom assessment skills needed to appropriately use classroom assessment to improve teaching and learning.

2. The study identifies a three-factor structure to understand teachers' perceived influence of large-scale testing accountability on teaching and learning. Factor 1, which was termed as *motivation for teachers*, was centred on six items, which implies that teachers perceive large-scale testing accountability as a way of motivating teachers and students. Factor two was termed "*pressure on teachers."* This factor was made up of five items, which implies that large-scale accountability puts pressure on teachers. Finally, factor three was termed "*pressure on students."* Teachers indicated that large-scale accountability put pressure on students, thereby increasing their test anxiety.

3. Moreover, teachers indicated that large-scale testing can be leveraged to move learning forward. Teachers indicated that the results of large-scale testing should not be seen as the sole basis for teacher effectiveness. Instead,, they claimed, large-scale testing results should be used to advocate for a more equitable allocation of state educational resources

and to boost teachers' professional development in order to strengthen instructional delivery.

4. The study again revealed that there is a statistically significant difference in teacher assessment practices on the basis of academic qualification. Teachers with a master's degree use assessment results more appropriately than teachers with a bachelor's degree.
5. Finally, the study found no statistically significant differences in teachers' years of experience concerning their assessment practices.

Conclusion

Outcomes from this study show that senior high school teachers in Birim Central Municipality have adequate classroom assessment skills needed to appropriately use classroom assessment to improve teaching and learning. However, these teachers have the skills set to adequately practice classroom assessment in a way that moves learning forward. Although these findings contrast with most similar work that has been done in other areas (central region) in Ghana (Awoniyi, 2016; Bortey, 2019; Bekoe, Eshun & Bordoh, 2013; Eshun, Bordoh, Bassaw & Mensah, 2014), they provide a preliminary understanding of the teachers' assessment literacy in the Birim central municipality in the Eastern region of Ghana.

On the other hand, teachers have also highlighted concerns regarding the use of large-scale testing results to make high-stakes decisions. For example, teachers express anxiety about how the results from large-scale testing are used to evaluate or determine schools and teachers' effectiveness. Teachers indicated that using large-scale testing results to make high-stake decisions creates pressure to force them to teach to test.

Investigating the effect of high stakes testing in Ghana's basic education, Amoako (2019) reported that teachers engaged learners in a more condensed curriculum as a result of the accountability pressures associated with the Basic Education Certificate Examination (BECE). Moreover, Oduro (2015) argued that the majority of teachers conceive assessment as mainly for accountability purposes, which tends to obscure the improvement (formative) function. This perspective on assessment has prevented teachers from exploring strategies to engage their students in the assessment process and encourage them to take responsibility for their own learning (Oduro, 2015).

Provision of teaching and learning materials (Adane, 2013; Baidoo-Anu, 2018; Baidoo-Anu & Mensah, 2018), and educational infrastructure (Baidoo-Anu, 2018), could have little impact on students' overall academic performance if teachers do not engage in formative assessment practises that move learning forward (Black & Williams, 2010). Again, irrespective of the assessment literacy of teachers, if large-scale testing results are not used appropriately, it will negatively affect teachers' assessment practices. According to Volente (2006), the advantage or disadvantage of large-scale assessment is determined by its use. Large-scale testing should be used to promote a more efficient allocation of educational resources and professional development for teachers in order to improve instruction. The ultimate value of large-scale assessment programmes is determined by their capacity to facilitate rather than obstruct student and teacher learning. To do this, large-scale assessments must be viewed as necessary, but not as indicators for making high-stakes decisions, Volente added. Again, to avoid unhealthy competition between schools and teachers,

LSA results should be communicated in a way that fosters professional collegiality.

Additionally, the majority of teachers are not highly skilled in interpreting summary scores associated with standardised test results (e.g., mean, percentile rank). This will make it difficult for most teachers to make meaning of the test results that will be handed to them. Subsequently, they will not be able to use it to improve teaching and learning. Teaching teachers how to interpret large-scale testing results will allow them to make better use of the report they have received.

Recommendation

The following recommendations for policy and practice have been made based on the study's findings and conclusions.

1. Ghana Education Service should use large-scale testing results to advocate for better resource allocation in education and teacher professional development in order to improve teaching within the Birim Central Municipality.
2. Again, evaluation of Birim Central Municipality schools and teachers' performance on the basis of students' performance on large-scale testing creates tension and puts pressure on teachers to teach to the test, which eventually narrows the curriculum. Education stakeholders should eschew using LSA results for evaluating schools and teachers' performance within the Birim Central Municipality.
3. Educational stakeholders responsible for using large-scale testing results should make sure that results are used to examine school policies

associated with curriculum and instruction within the Birim Central Municipality.

4. The results show that teachers are not highly skilled in interpreting summary scores associated with standardized test outcomes (e.g., mean, percentile rank). I recommend that the Ghana Education Service in Birim Central Municipality organizes workshops for senior high school teachers to orient them on how to interpret summary scores reported with standardised test results.

Suggestion for further Studies

The present study focused on only senior high school teachers in Birim Central Municipal, in the Eastern region. This will prevent generational mobility beyond the region. Therefore, I suggest that a further study be conducted to look at all senior high schools in other regions as well. This will aid generalization across the entire country.

REFERENCES

Abrams, L. M., Pedulla, J. J., & Madaus, G. F. (2003). Views from the Classroom: Teachers' Opinions of Statewide Testing Programs. *Theory into Practice*, *42*(1), 18–29.

Act, E. S. S. (2015). Every student succeeds act (ESSA). *Pub. L*, 114-95.

Adane, L. O. (2013). Factors affecting low academic achievement of pupils in Kemp Methodist Junior High School in Aburi. Eastern Region. Unpublished M.A. Thesis, University of Ghana, Legon.

Airasian, P. W. (1994). Classroom assessment. New York: McGraw-Hill.

Dunbar, S. B., Koretz, D. M., & Hoover, H. D. (1991). Quality control in the development and use of performance assessments. Applied Measurement in Education, 4(4), 289–303

Akyeampong K, Djangmah J., Oduro A., Seidu A., & Hunt, F. (2007) .Access to basic education in Ghana: The Evidence and the Issues – Country Analytic Report: CREATE, University of Sussex

Albrecht, S. F., & Joles, C. (2003). Accountability and access to opportunity: Mutually exclusive tenets under a high-stakes testing mandate. *Preventing School Failure: Alternative Education for Children and Youth*, *47*(2), 86-91.

Al-Fadhli, H. M., & Singh, M. (2010). Unequal moving to being equal: Impact of No Child Left Behind in the Mississippi delta. Journal of Negro Education, 79(1),18-32. Retrieved from http://search.proquest.com/docview/366345288? accountid=14872.

Amedahe, F. K., & Asamoah-Gyimah, A. (2003). Introduction to educational research for diploma in basic education.

Anamuah-Mensah Committee Report (2002). Meeting the challenges of Education in the Twenty First Century, Review report of the President's Committee on Education Reforms in Ghana: Author.

Anane, E. (2010). Effects of high-stakes testing on instruction in senior high school. International. Journal of Research in Education, 2(1) 58-66.

Assaf, L. (2006). One reading specialist's response to high-stakes testing pressures.International Reading Association, 3, 156-167. doi:10.1598/RT/60.26

Assaf, L. C. (2008). Professional identity of a reading teacher: Responding to high-stakes testing pressures. Teachers and Teaching: Theory and Practice, 14(3), 239-252. doi:10.1080/13540600802006137

Awoniyi, F. C. (2016). The understanding of senior high school mathematics teachers of School-Based Assessment and its challenges in the Cape Coast Metropolis. *British Journal of Education, 4*(10), 22-38.

Aysela T. (2012). An Exploration of the Effects of High-Stakes Examinations on the Teaching and Learning of Mathematics in Post-Primary Education in Ireland and Turkey. Unpublished doctoral Thesis, The National University of Ireland Maynooth.

Baidoo-Anu, D. (2018). *Students and Teachers Attitudes Responsible for Poor Academic Performance of Junior High School Students*. *2*(2), 1–8. https://doi.org/10.9734/AJESS/2018/40174

Baidoo-Anu, D. (2018). Perceived School Environmental, Home Conditions and Academic Performance of Junior High School Students in Asikuma-Odoben-Brakwa District, Ghana. *Journal of Education, Society and Behavioural Science*, *24*(2), 1–7.

https://doi.org/10.9734/jesbs/2018/39617

Baidoo-anu, D., & Mensah, G. E. (2018). The Perceptions of Junior High School Students and Teachers towards Teaching and Learning of Integrated Science at Komenda-Edina-Eguafo- Abrim District. Asian *Journal of Social Studies 2*(2), 1–8. https://doi.org/10.9734/AJESS/2018/40173

Bailey, K. M. (1996). Working for washback: A review of the washback concept in language testing.' *Language Testing, 13*, 257-279.

Baird, J. A., Hopfenbeck, T. N., Elwood, J., Caro, D., & Ahmed, A. (2014). Predictability in the Irish leaving certificate. *OUCEA Report*, *14*(1).

Baird, J. A., Johnson, S., Hopfenbeck, T. N., Isaacs, T., Sprague, T., Stobart, G., & Yu, G. (2016). On the supranational spell of PISA in policy. *Educational Research*, *58*(2), 121-138.

Baker, J. H. (2019). *An introduction to English legal history*. Oxford University Press.

Baker, E., Baron, P., Darling-Hammond, L., Haertel, E., Ladd, H., Linn, R., & Shepard, L. (2010). Problems with the use of student test scores to evaluate teachers (pp. 1–22). Washington.

Barber, M., Donnelly, K., & Rizvi, S. (2012). Oceans of innovation. *The Atlantic, the Pacific, Global Leadership and the Future of Education. London: Institute for Public Policy Research*, 70.

Barber, M., & Mourshed, M. (2007). How the world's best performing systems come out on top.

Bardes, B., & Denton, J. (2001). Using the grading process for departmental and program assessment. In *American Association for Higher Education Conference*.

Barksdale-Ladd, M. A., & Thomas, K., F. (2000). What's at stakes in high-stakes testing: Teachers and

Parents speak out. Journal of Teacher Education, 51(5), 384-397.

Baron, J. B. (1991). Strategies for the development of effective performance exercises. Applied Measurement in Education, 4(4), 305–318.

Bennett, W. J., & Wilezol, D. (2013). *Is college worth it? A former United States Secretary of Education and a liberal arts graduate expose the broken promise of higher education*. Thomas Nelson.

Berliner, D. C. (2006). Educational psychology: Searching for essence throughout a century of influence. *Handbook of educational psychology, 2*, 3-42.

Berliner, D. C. (2011). Rational responses to high stakes testing: The case of curriculum narrowing and the harm that follows. *Cambridge Journal of Education, 41*(3),287-302.

Berliner, D. C., Popham, W. J., & Shepard, L. A. (2000, April). *Three blueprints for a revolution: How to halt the harm caused by high-stakes tests.* General Session presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Berry (Ed.), *Assessment reform in education: Policy and practice* (pp. 63-74). London: Springer.

Berryhill, J., Linney, J. A., & Fromewick, J. (2009). The effects of education accountability on teachers: Are policies too stress-provoking for their

own good? International Journal of Education Policy & Leadership, 4(5), 1-14. Retrieved from http://journals.sfu.ca/ijepl/index.php/ijepl/article/viewFile/99/68

Bhattacharyya, S., Junot, M., & Clark, H. (2013). Can you hear us? Voices raised against standardized testing by novice teachers. Creative Education, 4, 633-639. doi: 10.4236/ce.2013.410091

Birenbaum, M., & Feldman, R. A. (1998). Relationships between learning patterns and attitudes towards two assessment formats. Educational Research, 40(1), 90-98

Black, P., Harrison, C., Lee, C., Marshall, B., and Wiliam, D. (2003). *Assessment for Learning: Putting it Into Practice*. Buckingham: Open University Press.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in education, 5*(1), 7-74.

Bond, N. (2011). Preparing preservice teachers to become teacher leaders. *The Educational Forum, 75*(4), 280-297.

Bortey, J. (2019). *Formative Assessment Practices among Senior High School Mathematics Teachers in the Cape Coast Metropolis*. Unpublished Thesis, University of Cape Coast: Cape Coast

Breneman, D. W. (2010). *Liberal arts colleges: Thriving, surviving, or endangered?* Brookings Institution Press.

Brill, F., Grayson, H., Kuhn, L. & O'Donnell, S. (2018). What impact does accountability have on curriculum, standards and engagement in education? A *literature review*. NFER. Accessed 09/01/2020 from

https:// www. nfer. ac. uk/ media/ 3032/ nfer_ accou ntabi lity_ liter ature_ review_ 2018. Pdf

Bracey, G. (2000). *High Stakes Testing.* Center for Education Research, Analysis, and Innovation, School of Education, University of Wisconsin-Milwaukee, Milwaukee, WI. Retrieved from http://nepc.colorado.edu/files/cerai-00-32.htm.

Brandt, R. (1995). Punished by rewards: A conversation with Alfie Kohn. *Educational Leadership, 53*(1). Retrieved from Alfie Kohn website: http://www.alfiekohn.org/ teaching/pbracwak.htm

Brookhart, S. M... (1993). Teachers' Grading Practices: Meaning and Values. *Journal of Educational Measurement, 30*(2), 123–142.

Brookhart, S. M. (2011). Educational Assessment Knowledge and Skills for Teachers. *Educational Measurement: Issues and Practice, 30*(1), 3-12. doi: 10.1111/j.1745-3992.2010.00195.x

Brown, G. T. L. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Policy, Principles and Practice, 11*(3), 305-322.

Brown, G. T. L. (2008). *Conceptions of Assessment: Understanding What Assessment Means to Teachers and Students.* New York, NY: Nova Science Publishers.

Burger, J.M., & Krueger, M. (2003). A Balanced Approach to High-Stakes Achievement Testing: An Analysis of the Literature With Policy Implications. *International Electronic Journal for Leadership in Learning, 7*(4).Retrieved from http://www.ucalgary.ca/iejll/burger_krueger/.

Burrows, S., Groce, E., & Webeck, M. L. (2005). Social studies education in the age of testing and accountability. *Educational Measurement: Issues and Practice, 3,* 13–20.

Calculator, S. N., & Black, T. (2009). Validation of an inventory of best practices in the provision of augmentative and alternative communication services to students with severe disabilities in general education classrooms. *American Journal of Speech - Language Pathology*, *18*(4), 329 - 42. Retrieved from EBSCOhost.

Campbell, D. T. (1976). Assessing the impact of planned social change. Occasional Paper Series, #8.

Campbell, C., & Evans, J. (2000). Investigation of preservice teachers' classroom assessment practices during student teaching. *The Journal of Educational Research*, *93*(6), 350. Retrieved from EBSCOhost.

Carless, D. (2011). *From Testing to Productive Student Learning: Implementing Formative Assessment in Confucian-Heritage Settings*. New York, NY: Routledge.

Carnoy, M., & Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. *Educational Evaluation and Policy Analysis*, *24*(4), 305–331.

Chester, M. D. (2005a). Making valid and consistent inferences about school effectiveness from multiple measures. *Educational Measurement: Issues and Practice, 4,* 40–52.

Chester, M. D. (2005b). Measuring the impact of state accountability programs. *Educational Measurement: Issues and Practice, 4,* 3–4.

Childs, R. A., & Fung, L. (2009). “The first year they cried”: How teachers address test stress. *Canadian Journal of Educational Administration and Policy, 64.* Retrieved from University of Manitoba website: http://www.umanitoba.ca/publications/ cjeap

Chingos, M. M. (2012). Strength in numbers. *State spending on K-12 assessment systems.*

Chudowsky, N. & Pellegrino, J. W. (2003). "Large-Scale Assessments That Support Learning: What Will It Take?”. *Theory into Practice. **42** (1), 75–83.*

Chung, P. J., & Chea, H. Y. E. O. N. W. O. O. (2016). South Korea’s accountability policy system and national achievement test. *The global testing culture: Shaping education policy, perceptions, and practice,* 249-260.

Ciani, K. D., Summers, J. J., & Easter, M. A. (2008). A top-down analysis of high school teacher motivation. Contemporary Educational Psychology, 33, 533-560. doi: 10.1080/01443411003660232

Cizek, G. J. (2001).More unintended consequences of high-stakes testing. *Educational Measurement: Issues and Practice, 20*(4), 19–27.

Crooks, T. (2011). Assessment for learning in the accountability era: New Zealand. *Studies in Educational Evaluation, 37,* 71-77. doi: 10.1016/j.stueduc.2011.03.002

Davis, M. F. (2011). The influence of high-stakes testing on science teacher perceptions and practices. (Ed.D., Walden University). ProQuest Dissertations and Theses. (861938019).

Darling-Hammond, L., Ancess, J., & Falk, B. (1994). *Authentic assessment in action: Studies of schools and students at work.* New York, NY: Teachers College Press.

Decker, D. M., & Bolt, S. E. (2008). Challenges and opportunities for promoting student achievement through large-scale assessment results: Research, reflections, and future directions. *Assessment for Effective Intervention, 34*(1), 43-51.

Dhindsa, H. S., Omar, K., & Waldrip, B. (2007). Upper secondary Bruneian science students' perceptions of assessment. *International Journal of Science Education, 29*(10), 1261-1280.

Dixon, H., (2011) Infusing peer assessment into classroom programmes: descriptions of practice. *Set: Research Information for Teachers* [Wellington], (2), 3 Retrieved from EBSCOhost.

Dixon-Román, E., Everson, H., & McArdle, J. (2013). Race, poverty and SAT scores: Modeling the influences of family income on black and white high school students' SAT performance. *Teachers College Record, 115*(4), 1-33.

Dobrow, S. R., Smith, W.K., & Poner, M.A. (2011). Managing the Grading Paradox: Leveraging the Power of Choice in the Classroom. *Academy Of Management Learning & Education, 10*(2), 261-276.

Donnelly, L. A., & Sadler, T. D. (2009). High school science teachers' views of standards and accountability. Science Education Policy, 93, 1050-1075. doi:10.1002/sce.20347.

Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher*, *44*(4), 237-251.

Earl, L. M. (1999). Assessment and Accountability in Education: Improvement or Surveillance?. *Education Canada*, *39*(3), 4.

Earl, L., & Katz, S. (2006). *Leading in a data rich world: Harnessing data for school improvement.* Thousand Oaks, CA: Corwin Press

Elbousty, Y. (2009). *High Stakes Testing Literature Review and Critique.* Paper presented at the Annual Meeting of the Northeastern Educational Research Association, Rocky Hill, CT, October 2009. Retrieved from http://digitalcommons.uconn.edu/cgi/viewcontent.cgi?article=1013&context=nera_2009

Emberger, M. (2007). Helping teachers improve classroom assessments. *Principal Leadership, 7(9), 24-29.*

Emler, T. E., Zhao, Y., Deng, J., Yin, D., & Wang, Y. (2019). Side effects of large-scale assessments in education. *ECNU Review of Education*, *2*(3), 279-296.

Eshun, I., Bordoh, A., Bassaw, K.T., & Mensah, F. M. (2014). Evaluation of social studies students' learning using formative assessment in selected Colleges of Education in Ghana. *British Journal of Education in Ghana*, *2*(2), 39-55.

FairTest. (2007). *The Dangerous Consequences of High-Stakes Standardized Testing.* Retrieved from http://www.fairtest.org/dangerous-consequences-highstakes-standardized-tes.

Florida Department of Education. (2010). *Frequently Asked Questions: Assessment and School Performance*/

Ferrara, D. (Ed.). (2005). [Special issue]. *Educational Measurement: Issues and Practice*, *24*(4).,18-22

Figazzolo, L. (2009). *Impact of PISA 2006 on the education policy debate*. Education Internat..

Figlio, D. & Loeb, S. (2011). School Accountability. pp 383-421 in Eric A. Hanushek, Stephen Machin, and Ludger Woessmann (eds*) Handbooks in Economics* Volume 3, North-Holland

Gabriel, T. (2010). Under pressure, teachers tamper with tests. New YorkTimes. Retrieved from http://bedfordtownmeeting.com/ uploads/ NYT_6_10_10_Test_Tampering.pdf

Gall, M. D. (1970). The use of questions in teaching. *Rev. Educ. Res.* 44, 707–721. doi: 10.3102/00346543040005707

Karsten, S., Visscher, A., & De Jong, T. (2001). Another side to the coin: the unintended effects of the publication of school performance data in England and France. *Comparative education*, *37*(2), 231-242.

Goldin, C., & Katz, L. F. (2008). Transitions: Career and family life cycles of the educational elite. *American Economic Review*, *98*(2), 363-69.

Goldstein, H. (1997). Methods in school effectiveness research. *School effectiveness and school improvement, 8,* 369–395.

Goodrum, D., Hackling, M., & Rennie, L (2001). The status and quality of teaching and learning of science in Australian schools. Canberra Department of Education: Training and Youth Affairs.

Gracin, D. G. (2011). *Requirements in mathematics textbooks and PISA assessment*. na.

Gregory, R. J. (1996). Psychological testing: History, principles, and applications (2nd ed.). Boston: Allyn and Bacon.

Grossman, S., & Hart, O. (1983). An analysis of the principal-agent problem. *Econometrica, 51*(1), 7–46.

Gruber, K. H. (2006). The German 'PISA-Shock': some aspects of the extraordinary impact of the OECD's PISA study on the German education system. *Cross-national attraction in education: Accounts from England and Germany*, 195-208.

Grupe, D., & Nitschke, J. (2013). Uncertainty and anticipation in anxiety. An integrated neurobiological and psychological perspective. *Nature Reviews Neuroscience, 14*(7), 488–501

Gutierrez, R. (2014). Improving education and the mistaken focus on "raising test scores"and "closing the achievement gap."In P. C. Gorski & K. Zenkov (Eds.), The big lies of school reform: Finding better solutions for the future of public education (pp. 17-28). New York, NY: Routledge. Retrieved from https://books.google.com/books?hl=en&lr=&id=7rc TAwAAQBAJ&oi=fnd&pg=PA17&dqd

Hamilton, L. & Koretz, D. (2002). Tests and their use in test-based accountability systems in L. Hamilton, B. Stecher and S. Klein (eds) *Making sense of test-based accountability in education.* RAND.

Hamilton, L. S., Stecher, B. M., & Klein, S. P. (Eds.). (2002). *Making sense out of test based accountability in education*. Washington, DC: RAND.

Hanson, M. J. S., & Carpenter, D. R. (2011). Integrating cooperative learning into classroom testing: implications for nursing education and practice. *Nursing Education Perspectives*, *32*(4),270.

Hanushek, E. A. (2016). What matters for student achievement. *Education Next*, *16*(2), 18-26.

Hanushek, E., & Raymond, M. (2005). Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management*, *24*(2), 297–327.

Hassan, D. K. (2019). Potentiating the innovation capacity of brainstorming testing approaches in architecture education medium. *International Journal of Technology and Design Education*, *29*(3), 583-602.

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

Hess, F. M. (2002). Reform, resistance ... retreat? The predictable politics of accountability in Virginia. In D. Ravitch (Ed.), *Brooking paper on education policy* (pp. 69-122). Washington: DC: Brookings Institute.

Heubert, J. P. (2000). High-stakes testing opportunities and risks for students of color, English Language learners, and students with disabilities. *National Center on Accessing the General Curriculum*, New York.

Hiss, W. C., & Franks, V. W. (2014). Defining promise: Optional standardized testing policies in American college and university admissions. Report of the National Association for College Admission Counseling (NACAC).

Hoffman, J., Assaf, L., & Paris, S. (2001). High-stakes testing in reading: Today in Texas, tomorrow? *The Reading Teacher*, *54*(5), 482-494.

Hout, M., & Elliott, E. Stuart (2011). Incentives and test-based accountability in education.

Hughes, A. (1993). Backwash and TOEFL 2000. Unpublished manuscript, University of Reading, England.

Igbalajobi, O. (1983). A needs assessment for an in-service training program for elementary classroom teachers of mildly handicapped children in Nigeria. *International Education*, *12*(2), 18.

Jacob, B.A. (2005). Accountability, Incentives and Behavior: The Impact of High-Stakes Testing in the Chicago

Jane, S. M. (2013). A vision of improvement of learning: South African teachers' conceptions of classroom assessment. Perspectives in Education, *31*(2), 14-21. Retrieved from EBSCOhost

Jerrim, J., Sims, S. School accountability and teacher stress: international evidence from the OECD TALIS study. *Educ Asse Eval Acc* (2021). https://doi.org/10.1007/s11092-021-09360-0

Jensen, A. (2012). *Educational Differences (RLE Edu L)*. Routledge.

Jones, B. D., & Egley, R. J. (2004). Voices from the frontlines: Teachers' perceptions of high-stakes testing. Education Policy Analysis Archives,12(39).http://dx.doi.org/10.14507/epaa.v12n39.2004

Jones, B.D., & Egley, R.J. (2007). Learning to Take Tests or Learning for Understanding? Teachers' Beliefs about Test-Based Accountability. *Educational Forum*, *71*(3), 232-248

Kane, T. J., & Staiger, D. O. (2002). Volatility in school test scores: Implications for test-based accountability systems. In D. Ravitch (Ed.),

Brookings papers on education policy 2002 (pp. 235–283). Washington, DC: Brookings Institute.

Kellaghan, T., Greaney, V., & Murray, S. (2009). *Using the results of a national assessment of educational achievement* (Vol. 5). World Bank Publications.

Kirkpatrick, R., & Zang, Y. (2011). The negative influences of exam-oriented education on Chinese high school students: Backwash from classroom to child. *Language testing in Asia*, *1*(3), 1-10.

Klenowski, V. (2011a). Assessment for learning in the accoutability era: Queensland, Australia. *Studies in Educational Evaluation, 37*(1), 78-83.

Klenowski, V. (2011b). Assessment reform and educational change in Australia. In R.

Klinger, D. A., DeLuca, C., & Miller, T. (2008). The evolving culture of largescale assessments in Canadian education. *Canadian Journal of Educational Administration and Policy, 76,* 1–34.

Klinger, D. A., & Rogers, W. T. (2011). Teachers' perceptions of large-scale assessment programs within low-stakes accountability frameworks. *International Journal of Testing, 11,* 1–22.

Klinger, D. A., Shula, L. A., & Wade-Wooley, L. (2009). Towards an understanding of gender difference in literacy achievement. Toronto, ON: Education Quality and Accountability Office. Retrieved from EQAO website: http://www.eqao.com/ Research/pdf/E/FINAL_ENGLISH_Gender_Gap_Report_As_of_May _11_2010. pdf

Knowles, M. S. (1980). From pedagogy to andragogy. *Religious Education*.

Kobrin, J. L., Patterson, B. F., Shaw, E. J., Mattern, K. D., & Barbuti, S. M. (2008). Validity of the SAT® for Predicting First-Year College Grade Point Average. Research Report No. 2008-5. *College Board*.

Kohn, A. (2000). *The case against standardized testing: Raising the scores, ruining the schools*. Portsmouth, NH: Heinemann.

Kubiszyn, T., & Borich, G., D. (2013). *Educational testing & measurement: Classroom applications and practice* (10th ed.). Hoboken, NJ: Wiley.

Kumekpor, T. K. (2002). *Research methods and techniques of social research*. SonLife Press & Services.

Ladson-Billings, G. (2006). From the achievement gap to the education debt: Understanding achievement in US schools. *Educational researcher*, *35*(7), 3-12.

Lambert, D., & Lines, D. (2000). *Understanding assessment: Purposes, perceptions, practice*. London: Routledge/Falmer

Lee, V. E., & Burkam, D. T. (2002). *Inequality at the starting gate: Social background differences in achievement as children begin school*. Economic Policy Institute, 1660 L Street, NW, Suite 1200, Washington, DC 20036.

Lewis, A. (2000). High-Stakes Testing: Trends and Issues. *Mid-Continent Research for Education and Learning Policy Brief*, April 2000. Retrieved from http://www.mcrel.org/pdf/policybriefs/5993pi_pbhighstakestesting.pdf

Leung, C. (2004). Developing formative teacher assessment: knowledge, practice, and changes. *Lang. Assess. Q.* 1, 19–41. doi: 10.1207/s15434311laq0101_3

Leung, C., and Mohan, B. (2004). Teacher formative assessment and talk in classroom context: assessment as discourse and assessment of discourse. *Lang. Test.* 21, 335–359. doi: 10.1191/0265532204lt287oa

Linn, R. (2011). Test-based accountability. *The Gordon Commission on the Future of Assessment in Education. Consultado en http://www. gordoncommission. org/rsc/pdf/linn_test_based_accountability. pdf.*

Linn, R. L. (2003). Accountability: Responsibility and reasonable expectations. *Educational Researcher, 32*(3), 3–13.

Linn, R. L., Baker, E. L., & Betebenner, D. W. (2002). Accountability systems: Implications of requirements of the no child left behind act of 2001. *Educational researcher, 31*(6), 3-16.

Lissitz, R. W., & Schafer, W. D. (2002). *Assessment in Educational Reform: Both Means and Ends*. College Division, Allyn & Bacon, Inc., 75 Arlington Street, Suite 300, Boston. MA 02116.

Marchant, G.J., & Paulson, S.E. (2005). The Relationship of High School Graduation Exams to Graduation Rates and SAT Scores. *Education Policy Analysis Archives, 13*(6). ERIC Document Reproduction Service No. EJ846516.

Marzano, R. J., Pickering, D., & Pollock, J. E. (2001). *Classroom instruction that works: Research-based strategies for increasing student achievement*. Ascd.

McMillan, J. H., Myran, S., & Workman, D. (2002). Elementary teachers' classroom assessment and grading practices. *The Journal of Educational Research*, *95*(4), 203.

McMillan, J.H. (2005). *The Impact of High-Stakes Test Results on Teachers' Instructional and Classroom Assessment Practices.* Metropolitan Educational Research Consortium, Virginia Commonwealth University.

Menken, K. (2006). Teaching to the test: How No Child Left Behind impacts language policy, curriculum, and instruction for English language learners. *Bilingual Research Journal*, *30*(2), 521-546.

Mika, S. (2005). *Testing Testing: Do Schools Do Too Much?* Gallup, Inc. Retrieved from http://www.gallup.com/ poll/19552/testing-testing-schools-too-much.aspx.

Mirazchiyski, P. (2013). *Providing School- Level Reports from International Large- Scale Assessments: Methodological Considerations, Limitations, and Possible Solutions.* Amsterdam: International Association for the Evaluation of Educational Achievement

Mitchell, R. (2006). *Research Review: Effects of High-Stakes Testing on Instruction.* The Center for Public Education, Alexandria, VA. Retrieved from http://www.centerforpubliceducation.org.

Moses, M.S., & Nanna, M.J. (2007). The Testing Culture and the Persistence of High Stakes Testing Reforms. *Education and Culture*, *23*(1), 55-72.

Mugenda, O. M., & Mugenda, A. G. (1999). *Research methods: Quantitative and qualitative approaches*. Acts press.

Mulvenon, S.W., Connors, J.V., & Lenares, D. (2001). *Impact of Accountability and School Testing on Students: Is There Evidence of Anxiety?* Paper presented at the Annual Meeting of the Mid-South Educational Research Association, Little Rock, AR, November 2001. ERIC Document Reproduction Service No. 460155.

Musumeci, D. (1996). Teacher-learner negotiation in content-based instruction: communication at cross-purposes? *Appl. Linguist.* 17, 286–325. doi: 10.1093/applin/17.3.286.

Muthanna, A., & Sang, G. (2015). Undergraduate Chinese students' perspectives on Gaokao examination: Strengths, weaknesses, and implications. *International Journal of Research Studies in Education*, *5*(2), 3-12.

Nagel, D. (2015). *Effective grading practices for secondary teachers: Practical strategies to prevent failure, recover credits, and increase standards-based/referenced grading*. Thousand Oaks, California: Corwin.

National Council on Measurement in Education. (2012). *Testing data integrity in the administration of statewide student assessment programs.* Madison, WI: Author.

Nichols, S. L. (2007). High-stakes testing: Does it increase achievement? *Journal of Applied School Psychology*, *23*(2), 47–64.

Nichols, S. L., & Berliner, D. C. (2007). *Collateral damage: How high-stakes testing corrupts America's schools*. Cambridge, Massachusetts: Harvard Education Press.

Nichols, S.L., & Berliner, D.C. (2008). Testing the Joy Out of Learning. *Educational Leadership, 65*(6), 14-18.

Noble, J., & Sawyer, R. (2002). Accuracy of High School Grades and College Admissions Test Scores for Predicting Different Levels of Academic Achievement in College. AIR 2002 Forum Paper.

Oduro, O.E (2015). *Assessment in mathematics classrooms in ghana: a study of teachers' practices.* Sussex: University of Sussex.

Palomba, C. A., & Banta, T. W. (1999). *Assessment Essentials: Planning, Implementing, and Improving Assessment in Higher Education. Higher and Adult Education Series*. Jossey-Bass, Inc., Publishers, 350 Sansome Street, San Francisco, CA 94104.

Paris, S. G., & Urdan, T. (2000). Policies and practices of high-stakes testing that influence teachers and schools. *Issues in Education, 6*(1/2), 83-108.

Papay, J. (2012). Refocusing the debate: Assessing the purposes and tools of teacher evaluation. *Harvard Educational Review*, *82*(1), 123-141.

Paton, P. (2013, March 31). Girls "marked up in lessons to reward good behavior." *The Telegraph (London).* Retrieved from The Telegraph website: http://www. telegraph.co.uk/education/educationnews/9963834/Girls-marked-up-in-lessons-to-reward-good-behaviour.html

Pedulla, J.J., Abrams, L.M., Madaus, G.F., Russell, M.K., Ramos, M.A., & Miao, J. (2003). *Perceived Effects of State-Mandated Testing Programs on Teaching and Learning: Findings from a National Survey of Teachers.* National Board on Educational Testing and Public

Policy, Lynch School of Education, Boston College, Chestnut Hill, MA. Retrieved from http://www.bc.edu/research/nbetpp/statements/nbr2summary.pdf

Pellegrino, J. W., & Goldman, S. R. (2017). Beyond rhetoric: Realities and complexities of integrating assessment into classroom teaching and learning. In *The future of assessment* (pp. 7-52). Routledge.

Perryman, J., Ball, S., Maguire, M., & Braun, A. (2011). Life in the pressure cooker: School league tables and English and mathematics teachers' responses to accountability in a results-driven era. British Journal of Educational Studies, 59(2), 179-195. Retrieved from http://mmiweb.org.uk/hull/site/pt/downloads/ pbl_pas_PressureCooker.pdf

Perryman, J. & Calvert, G. (2019). What motivates people to teach, and why do they leave? Accountability,performativity and teacher retention. *British Journal of Educational Studies*. https:// doi. org/ 10. 1080/ 00071 005. 2019. 15894 17.

Phelps, R. (2008). The role and importance of standardized testing in the world of teaching and training. *Nonpartisan Education Review, 4*(3), 1–9.

Phelps, R. (2012). The effect of testing on achievement: Meta-analysis and research summary, 1910–2010. *International Journal of Testing, 12,* 21–43.

Popham, W. J. (2000). *Modern educational measurement: Practical guidelines for educational leaders* (6th ed.). Boston: Allyn & Bacon.

Popham, W. J. (2002). *High-stakes tests: Harmful, permanent, fixable.* Paper presented at the Annual Conference of the American Research Council, New Orleans, LA.

Popham, W. J. (2003). Trouble with Testing: Why Standards-based Assessment Doesn't Measure Up. *American School Board Journal, 190*(2), 14-17.

Prytula, M., Noonan, B., & Hellsten, L. (2013). Toward Instructional Leadership: Principals'Perceptions of Large-Scale Assessment in Schools. *Canadian Journal of Educational Administration and Policy*, (140).

Pyle, A., & Deluca, C. (2013). Assessment in the kindergarten classroom: An empirical study of teachers' assessment approaches. *Early Childhood Education Journal, 41*(5), 373- 380. Retrieved from EBSCOhost.

Qi, L. (2005). Stakeholders' conflicting aims undermine the washback function of a high-stakes test. *Language Testing 22*(2): 142–173.

Ramirez, G., & Beilock, S. (2011). Writing about testing worries boosts exam performance in the classroom. *Science, 331*(6014), 211-213.

Ravitch, D. (2010). *The death and life of the great American school system: How testing and choice are undermining education.* New York, NY: Basic Books.

Rebora, A. (2012). Teachers place little value on standardized testing. Education Week, 31(26), 14-14.

Reese, M., Gordon, S. P., & Price, L. R. (2004). Teachers' perceptions of high-stakes testing. Journal of School Leadership, 14(5), 464-496.

Roderick, M., & Engel, M. (2001). The grasshopper and the ant: Motivational responses of low achieving students to high-stakes testing. *Educational Analysis and Policy Analysis, 23*(3), 197–227.

Roderick, M., Jacob, B., & Bryk, A. (2002). The impact of high-stakes testing in Chicago on student achievement in promotional gate grades. *Educational Evaluation and Policy Analysis, 24*(4), 333-357.

Romanoski, J., Cavanagh, R., Fisher, D., Waldrip, B., & Dorman, J. (2005). Measuring student perceptions of classroom assessment. In *AARE 2005 International Education Research Conference-Creative Dissent: Constructive Solutions* (pp. 2-12). AARE Inc..

Rotberg, I. C., Berstein, K. J., & Ritter, S. B. (2001). *No Child Left Behind: Views on the Potential impact of the Bush administration's education proposals* (Occasional Paper Series). Washington, DC: Center for Curriculum, Standards and Technology, Institute for Education Policy Studies, Graduate School of Education and Human Development, The George Washington University.

Rothstein, R., Jacobsen, R., & Wilder, T. (2008). *Grading education: Getting accountability right.* NY: Teachers College Press and the Economic Policy Institute..

Rowtree, D. (1987). New Jersey. *Assessing students: How shall we know them*? Library of Congress Cataloging in- Publication Data.

Rubin, D. I. (2011). The disheartened teacher: Living in the age of standardization, high-stakes assessments, and No Child Left Behind (NCLB).Changing English, 18(4), 407-416. doi:10.1080/1358684x.2011.630197

Sanders, N. M. (1966). *Classroom Questions: What Kinds?* New York, NY: Harper and Row.

Sarantakos, S. (1998). Sampling procedures. In *Social Research* (pp. 139-164). Palgrave, London.

Schaeffer, B. (2013). Standardized exam cheating confirmed in 37 states and D.C.; New report shows widespread test score corruption. FairTest, National Center for Fair and Open Testing. Retrieved from http://fairtest.org/2013-cheating-report

Schleicher, A. (2018). *World class*. OECD Publishing, Paris.

Sellar, S., & Lingard, B. (2014). The OECD and the expansion of PISA: New global modes of governance in education. *British Educational Research Journal*, *40*(6), 917-936.

Shavelson, R. J., Baxter, G. P., & Pine, J. (1991). Performance assessment in science. Applied Measurement in Education, 4(4), 347–362.

Shepard, L. A. (1991). Psychometricians' beliefs about learning. *Educational Researcher*, *20*(6), 2–16.

Shepard, L. (2000). The role of assessment in a learning culture. Educational Researcher, 29(7), 4–14.

Shepard, L. A. (2010).What the marketplace has brought us: Item-by-item teaching with little instructional insight. *Peabody Journal of Education, 85,* 246–257.

Shomoossi, N. (2004). The effect of teachers' questioning behavior on EFL classroom interaction: a classroom research study. *Read. Matrix* 4, 96–104.

Sjøberg, S. (2015). OECD, PISA, and globalization: The influence of the international assessment regime. In *Education policy perils* (pp. 114-145). Routledge.

Smith, M. & Amick, B. (1989). Electronic monitoring at the workplace: Implications for employee control and job stress pp 275–289 in S. L. Sauter. J. J. Hurrell Jr and C. L. Cooper (eds) *Job Control and Worker Health*. Wiley.

Smyth, E., Banks, J., & Calvert, E. (2011).From Leaving Certificate to Leaving School (A Longitudinal Study of Sixth Year Students).*The Economic and Social Research Institute.*

Stiggins, R. J., & Bridgeford, N. J. (1985). The ecology of classroom assessment. Journal of Educational Measurement, 22(4), 271–286.

Struyven, K., Dochy, F., Janssens, S., Schelfhou, W., & Gielen, S. (2006). The overall effects of end-of-course assessment on student performance: A comparison between multiple choice testing, peer assessment, case-based assessment and portfolio assessment. *Studies in Educational Evaluation*, *32*(3), 202-222.

Suen, H. K., & Yu, L. (2006). Chronic consequences of high-stakes testing? Lessons from the Chinese civil service exam. *Comparative Education Review*, *50*(1), 46-65.

Supovitz, J. (2009). Can high-stakes testing leverage educational improvement? Prospects from the last decade of testing and accountability reform. *Journal of Educational Change, 10*, 211–227.

Tan, Z. (2007). Questioning in Chinese university EL classrooms: what lies beyond it? *RELC J.* 38, 87–103. doi: 10.1177/0033688206076161

Taylor, G., Shepard, L., Kinner, F., & Rosenthal, J. (2003). A survey of teachers' perspectives on high-stakes testing in Colorado: What gets taught, what gets lost. Center for Research on Evaluation, Diversity and Excellence, University of California, Santa Cruz

Thompson, S. (2001). The authentic standards movement and its evil twin. *Phi Delta Kappan, 82*(5), 358-363.

Thurlow, M. L., & Ysseldyke, J. E. (2001). Standard setting challenges for special populations. In G. J. Cizek (Ed.), *Setting performance standards: Concepts*,

Tienken, C. H., & Zhao, Y. (2013). How common standards and standardized testing widen the opportunity gap. In P. L. Carter & K. G. Welner (Eds.), Closing the opportunity gap: What America must do to give every child an even chance (pp. 111-122). New York, NY: Oxford University Press.

Tobin, M., Lietz, P., Nugroho, D., Vivekanandan, R., & Nyamkhuu, T. (2015). Using large-scale assessments of students' learning to inform education policy: Insights from the Asia-Pacific region.

Turner, R. (2014). *Assessing mathematical literacy*. K. Stacey (Ed.). Springer International Publishing AG.

UNESCO (2019). The promise of large-scale learning assessments: acknowledging limits to unlock opportunities. UNESCO. ISBN 978-92-3-100333-2.

United States Department of Education. (2003). No child left behind, accountability and adequate yearly progress (AYP). Washington, DC: US Department of Education. Retrieved from United States Government website: http://www2.ed.gov/nclb/ landing.jhtml

United States Department of Education. (2009). *Race to the Top program: Executive summary.* Washington, DC: US Department of Education. Retrieved from United States Government website: https://www2.ed.gov/programs/racetothetop/ executive-summary.pdf

Volante, L. (2005). Accountability, Student Assessment, and the Need for a Comprehensive Approach. *International Electronic Journal for Leadership in Learning*, *9*(6), n6.

Volante, L. (2006). An alternative vision for large-scale assessment in Canada. *Journal of Teaching and Learning*, *4*(1).

Von Zastrow, C., & Janc, H. (2004). *Academic atrophy: The condition of the liberal arts in America's public schools.* Washington, DC: Council for Basic Education

von der Embse, N.P. (2008). *High Stakes Testing, Socio-economic Status and Test Anxiety: Opportunities for Impact.* Thesis submitted to the Department of Educational Psychology, Miami University, Oxford, OH. Retrieved from http://etd.ohiolink.edu/send-pdf.cgi/von der Embse Nathaniel.pdf?miami1214510534

Wagner, T. (2008). Rigor redefined. *Educational leadership, 66*(2), 20-24.

Westchester Institute for Human Services Research. (2003). High-Stakes Testing. *The Balanced View, 7*(1).Retrieved from http://www.sharingsuccess.org/code/bv/testing.pdf.

Wagner, T. (2012). *Creating innovators: The making of young people who will change the world*. Simon and Schuster.

Wilde, M. (2008). *Are We Stressing Out Our Kids?* Retrieved from http://www.greatschools.org/parenting/

. Wiggins, G. P., & McTighe, J. (2005). *Understanding by design.* (Expanded 2nd ed.). Upper Saddle River, NJ: Merrill/ASCD.

Xu, Y., and Liu, Y. (2009). Teacher assessment knowledge and practice: a narrative inquiry of a Chinese college EFL teacher's experiences. *TESOL Q.* 43, 493–513. doi: 10.1002/j.1545-7249.2009.tb00246.x

Yamashita, M. Y. (2011). How does high stakes testing influence teachers' classroom instruction?: Institutional pressures and classroom instruction. University of Pittsburgh. Retrieved from ProQuest Dissertations and Theses. (Order No. 893188370).

Yu, W. M. (2015). Teacher leaders' perceptions and practice of student assessment reform in Hong Kong: a case study. *Plann. Chang.* 46, 175–192.

Zacharos, K., Koliopoulos, D., Dokimaki, M., & Kassoumi, H. (2007). Views of prospective early childhood education teachers, towards mathematics and its instruction. *European Journal of Teacher Education, 30*(3), 305-318.

Zhang, G., & Zhao, Y. (2014). Achievement gap in China. In *Closing the achievement gap from an international perspective* (pp. 217-228). Springer, Dordrecht.

Zhao, Y. (2012). Doublethink: The creativity-testing conflict. *Education Week, 31*(36), 26-32.

Zhao, Y. (2014). *Who's afraid of the big bad dragon? Why China has the best (and worst) education system in the world*. New York: Jossey-Bass.

Zhao, Y., & Gearin, B. (2016). Paradigm shift: Fostering an entrepreneurial mindset in schools. *RICERCAZIONE*, 39.

APPENDICES

APPENDIX A

UNIVERSITY OF CAPE COAST

DEPARTMENT OF EDUCATION AND PSYCHOLOGY

QUESTIONNAIRE FOR SHS TEACHERS

Dear Respondent,

I am Master of Philosophy student of University of Cape Coast conducting a research on the topic: **"**Exploring Senior High School Teachers' Perceived Influence of Large-Scale Testing Accountability on Teaching and Learning at Birim Central Municipal".

The goal of this study is to obtain evidence of viewpoints, opinions, and attitudes that teachers have regarding large- scale tests (high-stakes tests).

I, therefore, solicit your cooperation and consent to participate in this study. Your responses will be used for academic studies only and treated with utmost confidentiality.

Note: Large-scale tests (high -stake tests) are standardized examinations such as BECE, WASSCE , GCE 'A' Level and Teachers Promotion Examination of which the results are used to make decisions or have significant effects on the students (repetition, graduation, job opportunities , etc.), teachers (incentives, rewards, promotion, punishment, ,etc.) or schools/districts (resource allocation ,grading ,etc.).

SECTION A

DEMOGRAPHIC INFORMATION

Please, fill in the following demographic information, or thick the appropriate answer when choices are provided.

1. Gender: Male [ ] Female [ ]

2. Age in years: 15-20 [ ] 21-30 [ ] 31-40[ ] 41-50[ ] 51-60[ ] Above 60[ ]

3. Highest level of education: Diploma [ ] Bachelor's degree [ ] Master's degree [ ] Doctorate degree [ ]

4. Number of years of teaching: Less than 5[ ] 5-8 [ ] 9-12[ ] 13-16[ ] 17-20[ ] Above 20[ ]

5. Form**(s)** currently teaching: One [ ] Two [ ] Three [ ]

6. Subject(s) currently teaching:

……

7. Have you ever taken a course in Assessment in Schools or Educational Assessment?

Yes [ ] No [ ] (If yes, please continue)

SECTION B

Teacher's Assessment Practices

Scale of Teacher Assessment Practices (STAP)

This scale addresses issues in applying assessment practices in the classroom. There are 12 items relating to assessment practices that may be applied in the classroom. Each item is followed by a scale ranging from (very low) to (very high). Please, estimate the level of your skills with each practice and thick the appropriate scale. For example, if you feel that your skills are "Very High" with regard to "Explaining assessment results clearly to parents" then you would thick "Very High". In contrast, if you feel that your skills relating to this assessment practice are "Very Low" then you would thick "Very Low".

Statement	Very Low	Low	Acceptable	High	Very High
1. "Selecting multiple methods of assessment (e.g., tests, observations, projects)"					
2. "Creating assessment tasks that accommodate the needs of a variety of students"					
3." Developing assessment tasks with different formats (e.g., multiple-choice, fill-in-blank, short answer)"					
4. "Sampling from the domain defined by learning goals to write assessment items"					
5. "Determining if an assessment is aligned with required standards (national curriculum goals)"					
6."Interpreting summary scores reported with standardized test results (e.g., mean, percentile rank)"					

7. Recognizing when assessment results are being used inappropriately by others					
8. Revising test items based on item analysis					
9. Using the results of formative assessment to adjust future lesson plans					
10. Explaining assessment results clearly to parents for decision making					

SECTION C
TEACHERS' SURVEY ON THE INFLUENCE OF LARGE-SCALE TESTING ACCOUNTABILITY ON TEACHING AND LEARNING

Please, thick (√) the appropriate column to indicate your level of agreement with the following statements: Strongly Agree (SA), Agree (A), Disagree (D), Strongly Disagree (SD)

Statement	SA	A	D	SD
1. Scores on the large-scale testing accurately reflect the quality of education students have received				
2. The large scale testing motivates previously unmotivated students to learn				
3. Differences among schools on the large scale testing are more a reflection of students' background characteristics than of school effectiveness				
4. large-scale testing results have led to many students not furthering their education in my municipality				
5. Media coverage of large-scale testing issues has been unfair to teachers				

6. Teachers feel pressure from the headmaster to raise scores on the large - scale testing				
7. "Students' low performance on high-stakes tests is an indication of poor teaching."				
8. "There is so much pressure for high scores on the large- scale testing that teachers only teach to the test."				
9. "Teachers in my school want to transfer out of form 3 where the large -scale testing is administered"				
10. Teachers feel pressure from parents to raise scores on the large- scale testing				
11. Administrators in my school believe students' large-scale testing scores reflect the quality of teachers' instruction				
12. Teachers in my school have found ways to raise large- scale testing scores without really improving student learning				
13. My school has an atmosphere conducive for learning				
14. Large -scale testing has increased teachers' morale in my school				
15. High stake test is an appropriate way to assess what student know				
16. The national curriculum is aligned with the large - scale testing				
17. Teachers should be evaluated based on their students' scores on high-stakes tests?				
18.The use of large-scale tests has resulted in an increase in students' test anxiety				
19.Large-scale testing programme has led teachers to teach in ways that				

contradict their own ideas of good educational practice				
20.On average, teachers spend at least half of their instructional time (or more) preparing their students for large-scale tests				

SECTION D

HOW TEACHING AND LEARNING CAN BE IMPROVED IN THIS ERA OF LARGE SCALE TESTING DOMINANCE

Please, thick (√) the appropriate column to indicate your level of agreement with the following statements: Strongly Agree (SA), Agree (A), Disagree (D), Strongly Disagree (SD)

Statement	SA	A	D	SD
1.Results of large-scale testing should not be seen as the sole basis for teacher effectiveness				
2.Items on the large scale testing should accurately reflect the quality of education students have received				
3. Large scale testing results data should be used to improve teaching and learning.				
4. Large scale testing results should be used to examine school policies related to curriculum and instruction.				
5. Large scale testing results should be used to encourage teachers' professional development to improve instruction.				
6.Large scale testing results should be used in attracting additional resources for students with greatest needs				
7.Large scale testing results should be reported in a manner that encourages professional collegiality				
8.Large scale testing results should be used to promote better allocation of state educational resources				

Kindly provide any additional information you have here.

……

……

……

……

……

……

……

……

……

THANK YOU

APPENDIX B

RELIABILITY COEFFICIENT OF QUESTIONNAIRE

Reliability Statistics	
Cronbach's Alpha	N of Items
.830	38

\

APPENDIX C

INTRODUCTORY LETTER

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

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

UNIVERSITY POST OFFICE
CAPE COAST, GHANA

Our Ref:

Your Ref:

16th January, 2020

TO WHOM IT MAY CONCERN

Dear Sir/Madam,

THESIS WORK
LETTER OF INTRODUCTION: MR. ISAAC ENNU BAIDOO

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

Mr. Ennu Baidoo is researching on the topic: **''EXPLORING SENIOR HIGH SCHOOL TEACHERS' PERCEIVED INFLUENCE OF LARGE-SCALE TESTING ACCOUNTABILITY ON TEACHING AND LEARNING AT BIRIM CENTRAL MUNICIPAL.''**

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 with 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.

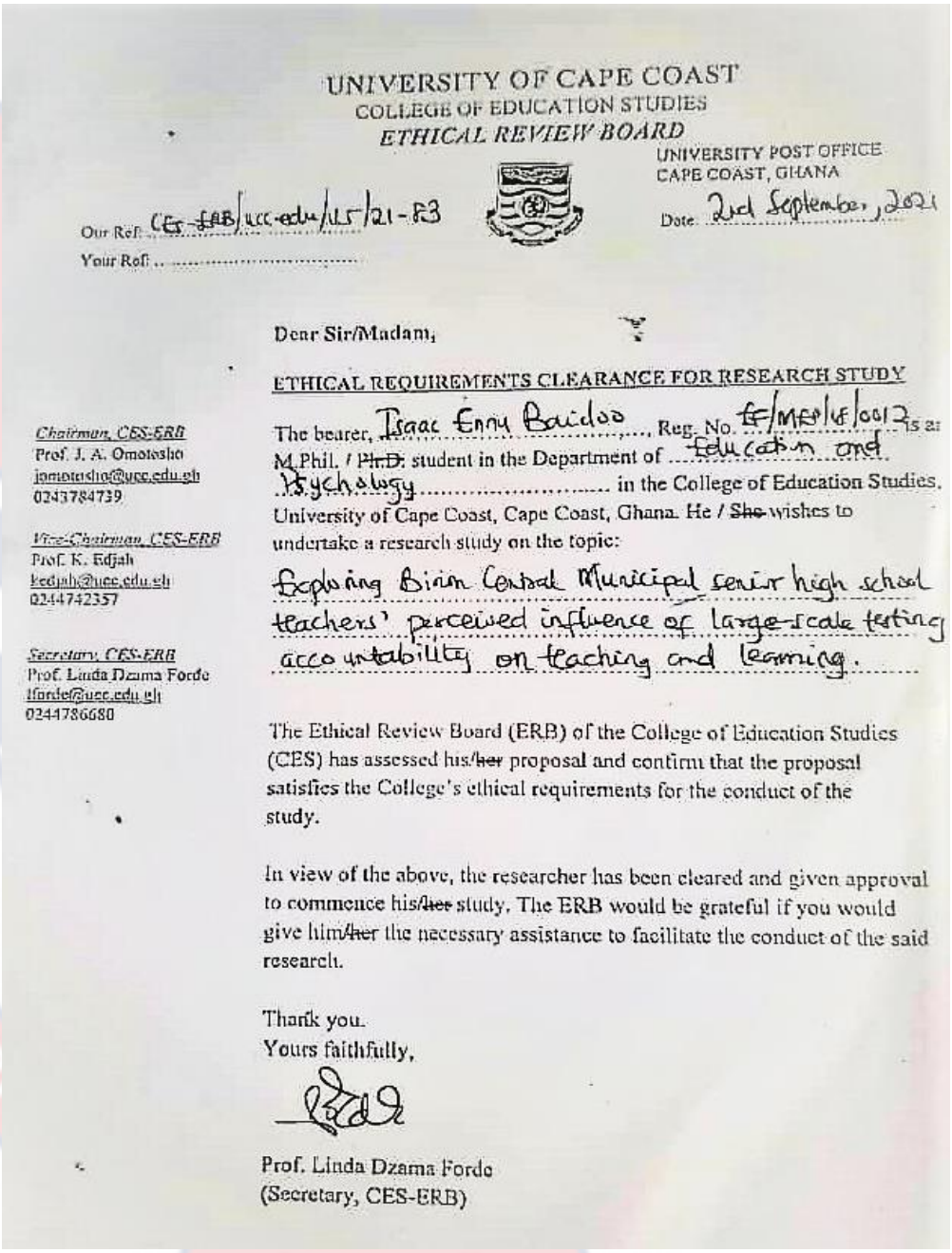
Thank you.

Yours faithfully,

Ama A. Ocran (Ms.)
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/ILS/21-P3
Date: 2nd September, 2021
Your Ref: .....

Dear Sir/Madam,

ETHICAL REQUIREMENTS CLEARANCE FOR RESEARCH STUDY

The bearer, Isaac Enny Baidoo, Reg. No. EF/MED/18/0012 is a 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:

Exploring Birim Central Municipal senior high school teachers' perceived influence of large-scale testing accountability on teaching and learning.

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

DECLARATION

Candidate's Declaration

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

Candidate's Signature:…….. Date:……

Name: Edward Fayah

Supervisor's Declaration

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

Supervisor's Signature:……. Date:……

Name: Mr. Eugene Adjei Johnson.

ABSTRACT

This study was to investigate the existence of the intuitive rule: 'Everything can be divided into two' in primary and junior high school students. It also investigated the students' perception of matter in the context of subdivisions.

Two schools in the Ga West district of the Greater Accra Region were selected based on convenience. The selected schools were a public school and a private school. All students in primary five to junior high school form three in the two selected schools were used. Where a class had more than one stream, one class was selected by simple random sampling using the shuffling method. In all 408 students participated in the study.

The main instrument used for data collection was a self report survey questionnaire with close-ended and open-ended questions. It covered 14 items with 5 items on subdivisions. Four of the items on subdivisions were on repeated halving tasks and one was on the serial dilution of sugar solution. The responses of the students were analyzed to determine the intuitive rule the students have.

The results of the study showed that in all the classes majority of the students have the intuitive rule, 'All divisions will end' meaning that subdivisions must eventually cease.

As a result of these findings, it was suggested that in teacher education, it is important to raise teachers' awareness of the role of intuitive rules in students' thinking, so as to enable them foresee possible cognitive obstacles.

ACKNOWLEDGEMENTS

I am grateful to my supervisor, Mr. Eugene Adjei Johnson of the Department of Science and Mathematics education, University of Cape Coast, for his support, encouragement and guidance. Mr. Godwin Aboagye of the Department of Science and Mathematics education, University of Cape Coast, was very helpful and I appreciate his effort.

I am also grateful to Professor Joseph Ghartey Ampiah who gave me the first journal that led me on this path. Many thanks to my friend Mr. Emmanuel Kissi who assisted me in obtaining most of the journals that I used for the work.

Special thanks to my wife Jemima who also ‘doubled’ as my typist.

DEDICATION

To Jemima, Eden-Victoria, and Edward Junior.

TABLE OF CONTENTS

LISTS OF FIGURES

CHAPTER ONE

INTRODUCTION

Background to the Study

A major thrust in mathematics and science education recently has been the study of students' conceptions and reasoning in science and mathematics. Many have pointed out the persistence of misconceptions, naive conceptions, intuitive conceptions and preconceptions (Stavy & Tirosh, 2000). According to Pine, Messer, and St. John (2001) children have pre-existing knowledge in the form of naive theories that are a pervasive feature of children's early understanding of science.

Many alternative conceptions held by children have been discovered in various fields of science, such as physics, astronomy, chemistry, biology and mathematics (Yair & Yair, 2004). According to Nussbaum (1985) the alternate views held by school children are embedded in their cognition, being based on intuition and daily experience. Therefore students come to the classroom with their own ideas or conceptions.For example, when young children compare two cups containing equal amounts of water, with one cup being narrower and taller than the other, they often incorrectly claim that ' The taller – the more' (Piaget, 1965).

According to Pine et al (2001), children do not come to primary science lessons as a 'tabula rasa' but come with rich knowledge about their physical worldbased on their everyday experience; although this rich

knowledge is laden with over generalisations, heuristics and misconceptions. For example,in the concept of gravity, large objects are believed to fall faster than small objects.

Research in the last decades has pointed out the persistence of alternative conceptions.These studies have highlighted the persistence of students' difficulties (or alternative conceptions). These conceptions are in variance with currently accepted scientific notions.It has been found that the students hold alternative ideas that are not in line with accepted scientific notions. That is, they are not always compatible with those accepted in science. Such difficulties lead students to give erroneous responses (Babai, 2009; Stavy & Tirosh, 1996a, 1996b, 2000; Tirosh & Stavy, 1996, 1999; Yair & Yair, 2004).

Mainstream research on students' incorrect responses to mathematics and scientific tasks, have been to identify the forms of reasoning required to provide correct responses to given tasks, arguing that incorrect responses result from lack of required cognitive schemes. Piaget is credited with it. He is the one who started work on children's errors and related it to their cognitive schemes (Piaget, 1965, 1968, 1969; Piaget & Inhelder, 1974). The problem with mainstream research, however, is that, it does not account for the variability in students' responses to different tasks that, according to these theories demand the same logical schemes (Stavy & Tirosh, 2000).

Another research on students' inappropriate responses on mathematics and scientific tasks has been the alternative conception paradigm (Driver, 1994; Vosniadou & Brewer, 1987). This approach aimed at detailed descriptions of certain particular alternative conceptions, where the child takes

an active constructive role in the knowledge - acquisition process. The student then brings to the learning situations alternative, internally coherent, robust and consistent perceptions.However, there is evidence that students tend to respond inconsistently to tasks related to the very same mathematical or scientific concept (diSessa, 1983; Tirosh, 1990). This evidence challenges the alternative conception paradigm.

While some researchers have observed some shared features in students' responses to different physics tasks, leading them to suggest underlying common cores of these responses (Andersson, 1986; Gutierrez & Ogborn, 1992); other researchers have looked for common routes and have tried to build an extensive theoretical framework that would account for these difficulties.

One framework is the intuitive rules theory developed by Stavy and Tirosh (1996a, 1996b,2000). They have proposed a theory to explain and predict students responses to scientific and mathematics tasks; the theory of intuitive rules. Many of the responses described in literature as alternate conceptions could be interpreted as evolving from several common intuitive rules. Some of the alternative conceptions in mathematics and science are based on the same intuitive rules (Tirosh & Stavy, 1999; Stavy & Tirosh, 2000).

According to the intuitive rules theory, students are affected by a small number of intuitive rules when solving a wide variety of conceptually non related science and mathematics tasks that share some common external features (Stavy & Tirosh, 2000).

Many common and incorrect responses to mathematics and scientific tasks can be interpreted as evolving from a small number of intuitive rules which are activated by specific external task features (Stavy, Babai, Tsamir, Tirosh, Lin, & McRobbi, 2006).

The essential claim of the intuitive rules theory is that human responses are determined mainly by irrelevant external features of the tasks and not by related concepts and ideas. Students responses to given mathematical and scientific tasks are often affected by common external features of these tasks which trigger the use of these intuitive rules (Stavy & Tirosh, 2000). This agrees with a previous study by Driver, Guesne and Tiberghien (1985) which concluded that there is thetendency for people to initially base their reasoning on observable features in a problem situation.

According to Stavy and Tirosh (2000), so far three types of responses have been identified, and accordingly three intuitive rules have been defined. These are 'More A – more B'; 'Same A – same B'; and 'Everything can be divided'.Two of the rules relate to comparison tasks ('More A – more B' and 'Same A- same B') and one rule relates to subdivision tasks ('Everything can be divided').

'More A – more B'

Stavy and Stachel (as cited in Stavy & Tirosh, 1996b, 2000) presented children aged 5 to 15 with two identical candles one of which was melted in their presence. They were asked whether the solid candle and the melted one weighed the same. Many children (6 to 10 year olds) argued that the solid candle is harder or stronger than the liquid candle and therefore weighed more. According to Stavy and Tirosh (2000), the above response is an application of

the intuitive rule ‘More A (the harder the candle) - More B (the more it weighs)’.

Strauss, Stavy and Orpaz (as cited in Stavy & Tirosh, 1996b, 2000) presented children with three cups containing equal amounts of hot water at the same temperature. The water from two of the cups was poured into an empty cup and the children were asked to compare the temperature of the water in the combined cup with that of the third cup’s contents. Majority of the children aged 6 to 8 claimed that the water in the combined cup was warmer. These incorrect judgements were justified with reference to the amount of water namely, ‘the more water, the warmer’.

A similar response was observed with older students who were presented with the same problem and also given numerical temperature values (e.g., 40 ^{0}C in each cup). Most children aged 7 to 11 argued that the temperature in the combined water was high than that in the original cup. These responses where often accompanied by arithmetic calculations (e.g., $40^0C + 40^0C = 80^0C$). According to Stavy and Tirosh, (1996b, 2000), this is also an application of the intuitive rule ‘More A (the more water) – More B (the warmer)’.

Piaget’s studies on the development of the concept of numbers were concerned with children’s dependence on length and density information when asked to compare the number of objects in two parallel rows containing the same number of objects. Young children (until about 5 or 6 years old) only paid attention to the relative length of the two rows. Rows of the same length were said to have the same number of objects; otherwise, the longer row was said to be more numerous than the shorter row. Some older children based

their judgements on the relative density of the two rows, stating that the denser row was more numerous (Piaget, 1965). According to Stavy and Tirosh (1996b, 2000) these two responses are in line with the intuitive rule 'More A (length of row or density of row) – More B (number of objects)'.

Tirosh (as cited in Stavy & Tirosh, 1996b), asked students aged 13 to 25 to compare the number of points in two line segments. According to Cantorian set theory, any two line segments contain the same number of points. Yet about half of the students across the entire age group claimed that 'The longer line segment contains more points'. This response seems to evolve from the intuitive rule 'More A (longer line segment) – more B (more points).

'Same A – same B'

Strauss and Stavy (as cited in Stavy & Tirosh, 2000; Tirosh & Stavy, 1999), presented children aged 4 to 15 with two cups of water and asked about the relative sweetness of the water after sugar was added to the cups. One cup was full of water and one teaspoon of sugar was mixed into it. This was done with the other same size but with half full cup. The children were asked whether they thought the sweetness of water in the two cups were the same or not, and if not in which cup the water was sweeter. Many of the young participants (4 to 8 year olds) argued that each cup contains one teaspoon of sugar and as a consequence, they must be equally sweet.

Very similar results were obtained in regard to the development of children's conceptions of temperature. Children were asked to compare the temperature of different amounts of water heated by the same number of candles for the same duration of time. They were also asked to compare the temperature of different amounts of water cooled by the same number of ice

cubes. Young children (4 to 9 year olds) incorrectly claimed that 'the temperature of the water in the two cups is equal because both were heated by the same number of candles' and that 'the temperature of the water in the two cups is equal because both are cooled by the same number of ice cubes' Strauss and Stavy (as cited in Stavy & Tirosh, 2000; Tirosh & Stavy, 1999).

The behaviours of young children in these studies were often interpreted as the application of incorrect alternative conceptions related to non differentiation between mass and concentration or non differentiation between heat and temperature (Erickson, 1979). Another explanation related to children's difficulty in coping with inverse ratio in the context of intensive quantities. According to Stavy and Tirosh (2000), such incorrect responses could be viewed as applications of the general rule 'Same A (same amount of sugar) –same B (same sweetness)'; 'Same A (same number of candles or ice cubes) – same B (same temperature)'.

Piaget, Inhelder and Szeminska (1960), asked young children to compare the length of a straight line with that of a wavy line. The lines were of different lengths but they begun and finished at parallel points on the page. Piaget et al (1960) reported that 84% of the children (4 to 5 year olds) incorrectly replied that the lines were equal in length. A typical response was 'They are both the same length (indicating the end points)'. Piaget et al (1960) interpreted this response by referring to children's development of the concept of length. They argued that 'at this stage, the length of a line is estimated solely in terms of its end points without reference to its rectilinearity' (p. 92). However, according to Stavy and Tirosh(2000), this response may also be

regarded as a case in which the intuitive rule ‘Same A (distance between end points) – Same B (length of the lines)’ is activated.

‘Everything can be divided’

The intuitive rules already discussed use comparison tasks. Another type of task used to investigate students’ conception in science and mathematics is the process of successive division. In mathematics, it is used to study students’ conception on infinity and in science it is used to study students’ conception on matter.

In a series of studies, students of different ages were presented with tasks related to the successive division (or repeated halving) of mathematical and physical objects (Stavy & Tirosh, 1993, 2000; Tirosh & Stavy, 1996). Students were asked whether the successive division of a line segment and a copper wire will end. Majority of the students at all grade levels responded in the same way to these tasks, although they are in two different domains. Some students responded that the repeated divisions of these two will end, that is, the division is finite. Some of the students also responded that the repeated division of the two will not end, that is, the division is infinite.

Only about a third of the students in the upper grades came up with a correct response to the two questions. When the question was repeated and a biological object (a carrot) was added, the students gave the same responses.It was also observed that at all grade levels the frequencies of a response pattern assuming infinity were higher than the frequencies of both the correct (different) responses and the response pattern assuming finiteness. According to Stavy and Tirosh (1996a, 2000), this is because these tasks elicit the use of the intuitive rule, ‘Everything can be divided’.

The correct response for the line segment which is mathematical (or geometrical) according to Euclidean geometry it can be infinitely divided. Therefore the scientifically correct answer would be infinite as this is the principal definition of such objects. The correct response for the copper wire which is physical (or material) is that the process of repeated halving comes to a halt when it reaches the atomic or molecular level. Beyond this, the material ceases to exist as such.

Statement of the Problem

Intuition is often described as a form of cognition related to specific content domain (concentration, area, perimeter, probability, volume etc.) and students' responses are then explained in the context of this domain (Tirosh & Stavy, 1999). Fischbein is credited with pioneering the work concerning the role of intuition in science and mathematics (Stavy & Tirosh, 1996b, 2000,; Tirosh & Stavy, 1996, 1999; Yair & Yair, 2004,; Stavy et al, 2006; Babai, 2009). According to Fischbein (as cited in Babai, 2009),intuition is an immediate cognition that exceeds the given facts. It is also a theory that implies an extrapolation beyond the directly accessible information. The intuitive rules are expressions of the natural tendency of our cognitive system to extrapolate (Stavy & Tirosh, 2000).

Often, students' intuitive knowledge of certain concepts or ideas is not in line with the accepted scientific frame works. Many students' responses to given mathematical and scientific tasks are often affected by common external features of these tasks which trigger the use of these intuitive rules. Each of these rules can be viewed as a common core to many misconceptions in science and mathematics (Tirosh & Stavy, 1999).

Most research on intuitive rules has been undertaken extensively in Israel. There has also been research on intuitive rules in some parts of the United States of America. Stavy et al (2006), tried to find how widespread intuitive rules are used by students. This was done using students in Taiwan, Australia and Israel. In their study, two of the intuitive rules; 'More A – more B' and 'Same A – same B' was investigated. In view of this, it will be desirable to investigateif the intuitive rule 'Everything can be divided into two' existsin primary and junior high school students in Accra.

Purpose of the Study

This study is aimed at investigating the existence of the intuitive rule 'Everything can be divided into two' in primary and junior high school students. It is also aimed at determining whether primary and junior high school students perceive matter as discrete or continuous in the context of subdivisions.

Research Questions

The study sought answers to the following questions:

1. To what extent does the intuitive rule 'Everything can be divided into two' exist in primary and junior high school students in selected basic schools in Accra?
2. In the context of subdivisions,do primary and junior highschool students in selected basic schools in Accra perceive matter as discrete or continuous?

Significance of the Study

The results of this study will be useful to the researcher and other interested science and mathematics teachers in many ways.

Firstly, recognizing the existence of intuitive rules such as ‘Everything can be divided into two’, may allow teachers to anticipate children’s answers and teach in a manner that will enable them abandon their preconceptions and change their view to a scientifically correct one.

Secondly knowledge of the role of intuitive rules in students’ thinking can form the basis for organizing workshops, seminars and in-service training for mathematics and science teachers. The teachers will then take this knowledge into consideration when planning sequences and strategies of instruction.

Finally, the outcome of this study could reveal the intuitive rule of the students as well as their perception of matter. This could help teachers, curriculum developers and course programme writers in developing lessons and syllabi which considerthe effect of intuitive reasoning on students’ responses to questions, and how to help students overcome them.

Delimitation of the Study

The study investigated the existence of only one intuitive rule.The existence of the other intuitive rules could not be investigated in a single study. All the basic schools in Accra could not be covered in the study. Only two schools out of a large number of basic schools in Accra were selected for the study. In the selected schools, only students in one class out of a number of classes from primary five to junior high school form three were selected for the study.

Limitations of the Study

Two basic schools out of a large number of public and private basic schools in the Ga West District of the Greater Accra Region were used for the study. Therefore generalization cannot be made to cover all the basic schools in Accra. Some of the students indicated that they had private teachers at home. The effect of what they have been taught by these private teachers on their response to the questions was not considered in the study.

Organisation of the rest of the Study

The dissertation has four additional chapters, which have been logically arranged to provide insights into the issues raised in this section and to provide answers to the research questions.

Chapter Two of the dissertation is devoted to a general review of the relevant literature on issues relating to the study, namely, characteristics of intuitive rules, origin of intuitive rules, further research on intuitive rules and more examples on the use of intuitive rules.

Chapter Three discusses the research methodology for the study. It describes the type of study and design in detail, and the rationale for the design. The strengths and weaknesses of the design are also discussed. Issues relating to population and sampling, instruments, data collection procedure, and data analysis are also discussed in detail.

In Chapter Four, the results of the study are presented and discussed according to the research questions raised. In Chapter Five, an overview of the research problem and methodology are given. A summary of the key findings is also provided. Implications and conclusions relating to the findings are also

discussed. In addition, the issues unearthed for possible future research are presented.

Definition of Terms

Intuition: Is when students solve mathematics and scientific tasks by using the external features of the tasks rather than considering the facts. It is often described as a form of cognition related to specific content domain (concentration, area, perimeter, probability, volume etc.) and students' responses are then explained in the context of this domain.

Subdivision tasks: Comprises of repeated halving tasks and serial dilution tasks.Repeated halving tasks refer to the process where a mathematical, material, or a biological object is divided into two; one half is discarded and the other half divided again into two. The process is then repeated several times. Serial dilution tasks refer to the process in which the volume of a solution (for example, sugar solution) is measured. It is divided into two and half of the solution is discarded. This is replaced with an equal volume of water and stirred. The process is then repeated several times. Serial dilution tasks can also be applied to a mixture of two solids like salt and sugar.

CHAPTER TWO

REVIEW OF RELATED LITERATURE

The chapter reviews relevant literature that provides support for the study under the following subheadings: Characteristics of intuitive rules, origin of intuitive rules, further research on intuitive rules and more examples on the use of intuitive rules.

Characteristics of Intuitive Rules

According to Fischbein as cited in (Stavy & Tirosh, 1996b, 2000,; Tirosh & Stavy, 1996, 1999; Yair & Yair, 2004,; Stavy et al, 2006; Babai, 2009), self evidence and confidence are two major characteristics of intuitive reasoning. It is self evident because students perceive statements or responses they made on the basis of the rules as being true and in need of no further justification. The rule is also used with great confidence. Other characteristics of intuitive reasoning include perseverance. The use of the rule often persists in spite of formal learning that contradicts it. The rule has the attribute of globality since students tend to apply it to diverse situations. It also has the attribute of coerciveness because alternatives are often excluded as unacceptable. Recent studies have also suggested that intuitive responses are immediate (Babai, 2009). The use of adverbs such as ‘always’, ‘never’ or ‘everything’ typify intuitive thought (Stavy & Tirosh, 1996a, 2000; Yair & Yair, 2004).

Stavy and Tirosh (2000), reports that a very distinguished physics professor, an expert in astronomy, answering a question on a television science program, was heard to explain that, 'the force that the larger star exerts on its moon is larger than the force that the moon exerts on the star'. This professor was very familiar with Newton's third law (the law of action and reaction), according to which any two bodies exert the same force on each other. The next day, he explained that because of his agitation at appearing on television, he had given an immediate, uncontrolled response.

This response is in line with the intuitive rule, 'More A (the larger the star) – more B (the more force it exerts)'. This illustrates the fact that, intuitive rules persists even after formal instruction and may dominate formal knowledge. It also shows that intuitive responses are immediate.

Origin of Intuitive Rules

The intuitive rules are expressions of the natural tendency of our cognitive systems to extrapolate (Stavy & Tirosh, 2000). The intuitive rules 'More A – more B' and 'Same A- same B' relate to comparison tasks. In many tasks embedded in everyday life and scientific situations, responses in line with these intuitive rules leads to conclusions that are accurate.For instance 'The more money you have, the more you can buy'; 'The more you study, the more you know'. Also 'Same heights of juice in two identical cups, same amount to drink'; 'Same number of candies, same price'.

For both, a repeated experience seems to reinforce it. Thus enhancing its use in other seemingly similar situations. Therefore, it is an over generalisation from successful experiences; often both in everyday life and in school situations. It is therefore reasonable to assume that children generalise

such experiences into the universal maxim: ‘More A – more B’ and ‘Same A – same B’ (Stavy & Tirosh, 1996b, 2000; Tirosh & Stavy, 1999).

The intuitive rule ‘Everything can be divided’ is a direct consequence of the natural tendency of our cognitive system to extrapolate. Here the extrapolation is from a visible stage in the process of subdividing a given object, to the next stage, and then to the next stage, and so on. That is, the extrapolation is from a given stage to the next one. Beginning at a certain age, subjects see no reason to argue that at a certain stage this extrapolation will not be valid anymore (Stavy & Tirosh, 2000).

Further Research on Intuitive Rules

Yair and Yair (2004), investigated the existence of the intuitive rule ‘Everything can be divided into two’ in elementary school children in Israel. The pupils were in grades 2 to 7 (7 to 13 years old). The results showed that an intuitive rule ‘Everything comes to an end’ meaning that subdivisions must eventually cease exists in elementary school children. However as the children mature, this rule seems to be less frequent and there is an increase in the appearance of the intuitive rule, ‘Everything can be divided into two’.

In grades 2 and 3 (7 to 9 years old), majority of the pupils have the intuitive rule ‘Everything comes to an end’. From grade 4 (9 years old), onwards there is a rise in the number of pupils who have the intuitive rule ‘Everything can be divided into two’; and a reduction in the number pupils who have the intuitive rule ‘Everything comes to an end’. Therefore the intuitive rule ‘Everything can be divided into two’ increases with age.

Stavy et al (2006), undertook a cross cultural study on the intuitive rules theory. The study was conducted in Australia (with aboriginal students),

in Taiwan and in Israel. In Australia students in grades 2 to 9 were used for the study. In Taiwan and in Israel, the students were in grades 2 to 10. The findings of the study was that Taiwanese and Australian Aboriginal students, much like their Israeli peers, tend to provide responses in line with the intuitive rules: 'More A – more B' and 'Same A – same B' when presented with relevant comparison tasks.

The results suggest that the developmental patterns of these two intuitive rules are similar in the three cultures and may imply that these intuitive rules are universal.

Babai (2009) did a study to consider the relationship between students' Piagetian cognitive levels and their tendency to answer in line with intuitive rules when solving comparison tasks. This study was done using students studying in the same grade level, in order to neutralize the effects of age and degree of formal education. The study used grade 7 students in junior high school(12 to 13 years old). Previous studies by Babai and Levit-Dori (as cited in Babai, 2009), showed that junior high school students in Israel manifest diverse cognitive levels mainly from early concrete operations stage to early formal operations stage.

The results suggest that intuitive reasoning is not restricted to young children. It also suggests that cognitive level influences differently the tendency to apply each intuitive rule. In the case of the intuitive rule; 'Same A – same B', the rate of correct answers decreases as the cognitive level advances. In addition, at the highest cognitive level, the tendency to respond in line with the rule was significantly higher. These observations suggest that advanced cognitive schemes reinforces the tendency to use this rule.

In the case of the intuitive rule; ‘More A – more B’, the results did not yield a clear cut picture. For example no significant differences were found among the various cognitive groups with respect to the rate or type of explanations given by the students to justify their responses. Although the results indicate an obvious positive influence to respond correctly or overcome the use of the rule with the advancement of cognitive level. However, the observed differences did not prove to be significant.

Further examples on the use of Intuitive Rules

The two match boxes, one full of sand and the other empty, were held at the same height above the ground in the same manner. They were both dropped at a certain point in time. The participants were asked if the match box full of sand will hit the ground before, or at the same time, or after the empty match box. This task was presented to various groups of students, including first-year college physics students in the United States and Israel. The researchers reported that many participants including children and many adults claimed that the full match box would hit the ground first (Stavy & Tirosh, 2000).

According to Stavy and Tirosh (2000), this response is in line with the intuitive rule ‘More A (the heavier) – more B (the faster)’. The correct response is that both match boxes will hit the ground at the same time.

According to Newtonian mechanics, the potential energy of the match box; PE =mgh (where m is the mass of the empty match box, g is the acceleration due to gravity and h is the height of the box above the ground).As the match box falls, its potential energy is transformed into kinetic energy; KE = ½ mv^2 (where v is the speed of the match box when it reaches the ground).

Form the law of conservation of energy: $mgh = \frac{1}{2} mv^2$. Consequently, $gh = \frac{1}{2}v^2$ and $v = \sqrt{2gh}$. Therefore the speed of the falling match box depends on the height at which it falls and is independent of the mass of the body.

Tirosh et al (as cited in Stavy & Tirosh, 1996a, 1996b), presented college students with several tasks concerning arithmetic operations. One such task was related to the relationship between factors and products in multiplication expressions. Many of the students argued that when one of the factors is increased, the product always increases. While this statement holds for natural numbers, it does not hold for negative numbers [for example, 2 x (-4) ≥8 x (-4)]. According to Stavy and Tirosh (1996a, 1996b), it seems that the mistaken students who made this overgeneralization were operating according to the rule 'More A (the bigger the factor) – more B (the bigger the product).

Rachmani (as cited in Stavy & Tirosh, 1996b) presented students at different age levels with several problems in an attempt to assess their understanding of percentages. One of the problems was: Joe saved 25% of his salary. Maya saved 50% of her salary. Can you determine which of them saved more money? Obviously, the answer to this problem is 'No', because the salaries of Joe and Maya are not known. Yet, many of the students responded that 'Maya saved more money, because she saved a higher percentage'. The response of the students, according to Stavy and Tirosh (1996b), is in line with the intuitive rule 'More A (a higher percentage) – more B (more money)'.

Is the size of a muscle cell of a mouse bigger than, equal to, or smaller than a muscle cell of an elephant? Explain your choice. This task was given to students in grades 7 to 12 (12 to 18 years old). Majority of the students

especially in grades 7 and 8 incorrectly claimed that an elephant which is the larger animal has a larger cell. This is in line with the intuitive rule ‘More A (larger animal) – more B (larger cell)’ (Stavy & Tirosh, 2000).

In a large survey of adolescents’ concepts of probability, Green (as cited in Stavy & Tirosh, 2000), presented students with the following task; two bags contain black and white counters. Bag J contains 3 black and 1 white, whilst bag K contains 6 black and 2 white. Which bag gives a better chance of picking a black counter? About half the students in the study incorrectly chose bag K as likely to yield a black counter, most of them reasoning that ‘there are more black ones in bag K’. This incorrect response, according to Stavy and Tirosh (2000), is in line with the intuitive rule ‘More A (more black counters in bag) – more B (the higher the probability).

Livne (as cited in Stavy & Tirosh, 2000), presented Israeli biology majors in grades 10 to 12 (15 to 18 years old) with the following task, Debby is a baby – sitter. The baby she is watching wakes up crying, and Debby wants to feed her. She realises that the milk she has heated is too hot and wants to cool it as fast as possible. She has two differently shaped bottles, a ball – shaped one and a cylinder – shaped one. Each bottle can contain 100ml. She fills each bottle with 100 ml of milk (up to the nipple), and immerses them both in ice water. Is the time needed to cool the milk in the spherical (ball – shaped) bottle equal to or not equal to the time needed to cool the milk in the cylindrical bottle? If you think that the time is not equal, in which bottle does the milk cool faster? Why?

Fewer than 50% of the students at each grade level based their answers on the ratio between surface area and volume and therefore knew that the

cylindrical bottle would cool faster. A number of students, 29% in grade10, 31% in grade 11 and 38% in grade 12 argued that 'The time needed to cool the milk in both bottles is equal because the amount of milk in each bottle is equal'. These response are in line with the intuitive rule, 'Same A (amount of milk) – same B (rate of cooling)'. These responses of mature students with a relatively high level of biology education suggest that the intuitive rule, 'Same A – same B' has coercive power.

Livne (as cited in Babai, 2009; Stavy & Tirosh, 2000; Tirosh & Stavy, 1999), asked high school students majoring in biology (grades 10 to 12), to consider two different – sized cubes and to decide if the ratio between the surface area and volume of the smaller cube was larger than, equal to, or smaller than the ratio between the surface area and volume of the bigger cube. The correct answer is that the smaller cube will have a larger ratio of surface area to volume than the bigger cube. However a substantial percentage of the students; 41% in grade 10, 45% in grade 11, and 55% in grade 12 incorrectly argued that the ratio of surface area to volume in both cubes was the same. The students explained their response by claiming that, since the cubes had the same geometrical shape, the ratio of surface area to volume in both cubes was the same regardless of their size. Such response and explanation is in line with the intuitive rule 'Same A (same shape) – same B (same ratio).

A task related to probabilistic thinking was studied by Fischbein and Schnarch (as cited in Babai, 2009; Stavy & Tirosh, 2000; Tirosh & Stavy, 1999). In their study, students were asked if the likelihood of getting heads at least twice when tossing three coins is smaller than, equal to, or greater than the likelihood of getting heads at least 200 times out of 300 tosses. The results

show that; 30% of students in grade 5, 45% of students in grade 7, 60% of students in grade 9, and 75% of students in grade 11, argued that the probabilities are equal since the ratios (2/3 and 200/300) are equal. The incorrect response of the students is in line with the intuitive rule 'Same A (same proportion) – same B (same probability).

In this case, the correct response is that, the probability of getting heads at least twice in tossing three coins is greater than getting at least 200 heads out of 300 tosses, since, as the sample size increases, the relative frequencies tend toward the theoretical probabilities (1/2).

According to Tirosh and Stavy (1999), Mathematics education literature reports that many students and adults adhere to the view that shapes with the same perimeter must have the same area. These studies interpreted students' responses as resulting from a misunderstanding of the relationship between the concepts of area and perimeter, that is, students are believed to think that shapes with the same perimeter must have the same area and vice versa. According to Tirosh and Stavy (1999), this response could be viewed in a broader perspective, as resulting from an application of the intuitive rule 'Same A – same B (same perimeter – same area; same area – same perimeter; and also same object – same perimeter or area).

A teaspoon of sugar is put into a cup of water and stirred well into it. Half of the sugar water is poured out, and half a cup of water is added to the cup and mixed thoroughly with the remaining sugar water. This is done again: Half of the sugar water poured out, half a cup of water is added, and so forth. This process is repeated. Is it possible that there is a stage at which no sugar at all will be found in the cup? Explain your answer.

This problem was presented to students in grades 7 to 11 (12 to 17 years old). Many students (around 60% in each grade level) claimed that the amount of sugar would become smaller and smaller but never reach zero because ‘Everything can be divided’. Some explained that the dissolved sugar could not disappear because ‘it mixes with water and spreads in it, and only half of both are poured out each time’. This problem refers to decreasing concentrations of sugar in solutions. The correct response is that due to the particulate nature of matter, after a large number of dilutions, the resulting solution might have a zero concentration of sugar (Stavy & Tirosh 2000).

Amy’s mother intended to prepare a yeast cake. She mixed a certain amount of yeast with water. Amy thought that her mother used too much yeast. She poured out half of the mixture, filled the cup with water, and stirred well. Still, she thought that there was too much yeast and decided to repeat the process: She poured out half of the mixture, added water, and mixed again. She repeated the process again and again. Is it possible that she reaches a stage when there will only be pure water, with no yeast, left in the cup? Explain your answer. This question was presented to students in grades 9 to 11(14 to 17 years old).

Most of the students, 53% in grade 9, 63% in grade 10, and 70% in grade 11, argued that it is impossible to reach a stage at which no yeast will be left. They explained that ‘Everything can be halved’ or that ‘You can always halve a half ’. Some argued that that ‘the yeast is absorbed in the water and at each stage half of it will be left’. The correct response is that the process will end at the last yeast cell. According to Stavy and Tirosh (2000), the high

percentage of incorrect responses in the two tasks assuming infinity could be interpreted as an instance of the use of the rule ‘Everything can be divided’.

Students in grades 9 to 11 were presented with the following problem: In a game for two players, each player has to put a number of matches on a square game board. The rules of the game are: The first player puts the number of matches he chooses on one square. The second player has to put half the number of the first player on the next square. Each player, in his turn, puts exactly half the amount of matches of the previous stage in the next square. Dan and Ran played this game. Dan put down 32 matches, Ran continued by putting down 16 matches, and so on. Will this game come to an end? Explain your answer (Stavy & Tirosh, 1996a, 2000; Tirosh & Stavy, 1996).

In this context, matter appears in the form of discrete, discontinuous objects, and the problem refers to a defined unit (a match) in reference to a game. Therefore it is expected that, most students will argue that the game ends when the last player puts down the last match. Indeed majority of the students came up with this argument. However, some of the students; 37% in grade 9, 28% in grade 10 and 49% in grade 11 respectively, argued that the process was infinite as one could always divide by two.

A task related to the successive division of an aluminium foil was given to students in grade 12. It was found that 52% of these students gave an infinite response to this question. The majority of the students justified this incorrect response by the argument; it is always possible to continue dividing by two (Tirosh & Stavy, 1996).

Naughty John put a tablespoon of salt into a bowl of sugar and mixed it well into the sugar. Then he regretted his action. He poured out half of the salt – sugar mixture, filled the bowl up with sugar, and stirred it in thoroughly with the remaining mixture. He tasted the mixture and could still taste the salt. Therefore he repeated the action of pouring out half of the salt – sugar mixture, adding sugar, and mixing everything thoroughly. He repeated this process again and again. Is it possible to reach a stage at which no salt at all will be present in the bowl? Explain your answer (Stavy & Tirosh, 2000).

Although this process involves substances that appear to be discrete, yet the vast majority of the students who took part in the study (about 80% in the upper grades) argued that 'salt will always remain in the sugar'. Their explanations were: 'You can always halve a half' and 'Half of the salt will always be poured out with half of the sugar, and therefore salt will always be left. This process is a serial dilution of salt and the method used by John could eventually lead to separating the salt out of the sugar (Stavy & Tirosh, 2000).

Students' answers and explanations to the three tasks described above is an indication of the coercive effect of the intuitive rule 'Everything can be divided' on students' responses.

To sum up, the intuitive rules are 'More A – more B', 'Same A – same B', 'Everything can be divided' and 'Everything comes to an end'. Children and adults responses to various science and mathematics tasks show the persistence of intuitive reasoning in their responses. Children's responses to Piagetian tasks which was interpreted as errors may be due to intuitive reasoning (Stavy & Tirosh, 1996a, 1996b, 2000; Tirosh & Stavy, 1999).

CHAPTER THREE

METHODOLOGY

This chapter describes and explains how the study was conducted. It discusses the research design, population, sample and sampling technique, research instrument, data collection procedure, and data analysis.

Research Design

This study used the survey design to investigate the existence of the intuitive rule ‘Everything can be divided into two’ in primary and junior high school students, using questions on the subdivision of mathematical and material objects, and to find out based on their answers if they perceive of matter as discrete or continuous.

A self-report survey questionnaire with a closed ‘yes’ or ‘no’ section and an open section where the students were asked to describe and explain their choice of answer was used. The rationale for the survey design is that it is useful for investigating a variety of educational problems including assessment of attitudes and opinions; Gay (as cited in Amponsah, 2009). Also results are procured faster. Sampling has a greater scope regarding the variety of information required and further, it allows for higher quality of work as more accurate data can be provided under suitable conditions.

Sample surveys have some disadvantages when it comes to eliciting basic information required for every unit of the population. Errors due to sampling also tend to be higher for small sample sizes.However considering

the homogeneity and magnitude of the target population for this research, the advantages in the sample survey far outweigh its disadvantages. For this study in particular, the data were collected using questionnaires, which have their own weaknesses such as personal biases and ambiguities or misunderstanding of items on the questionnaires. Therefore every question was explained verbally by me and demonstrated by a visual example to minimize this problem.

Population

According to Best and Kahn (as cited in Aboagye, 2009) a population is any group of individuals or objects that have one or more characteristics in common that are of interest to the researcher. It is the larger group about which generalization is made. The target population was students in primary 5 and 6; and junior high school; forms 1, 2 and 3 in private and public schools in the Tantra Hill and Ofankor area of the Ga West District of the Greater Accra Region.

Sample and Sampling Technique

Two schools were selected based on convenience. All students in class 5 to junior high school form 3 in the two selected schools were used. The two schools are a public school at Tantra Hill where the income levels of the parents are perceived to be relatively low (school A) and a private school at Ofankor where the income levels of the parents are perceived to be relatively high (school B). This is because we wanted students that represented the various income levels in the city.Students in primary classes 5 and 6, and junior high school forms 1, 2 and 3 in the selected schools were used. Where a class or form had more than one stream; then one class was selected. This was

done by simple random sampling using the shuffling method. All the students in the selected class were used. Previous research by Stavy and Tirosh (1996, 2000), Yair and Yair (2004), Stavy et al (2006), shows that students' responses to questions on subdivisions depends on their age. Table 1 shows the different classes of students that took part in the study; their class sizes and age statistics.

As indicated in Table 1, the mean age, which is the usual average age of the students in primary 5, was 10.66.The oldest in the class (or maximum age) was 14 and the youngest in the class (or minimum age) was 8. The total number of students in primary 5 for both schools who took part in the study was 90.

For primary 6; the mean age for the students was 11.37 years. The oldest in the class was 14 years and the youngest was 10 years. The total number of students in primary 6 for both schools who took part in the study was 68.

For junior high school form 1, the mean age was 12.44 years. The oldest in the class was 16 years and the youngest was 11 years. The total number of students in junior high school form 1 for both schools who took part in the study was 87.

For junior high school form 2, the mean age was 13.44. The oldest student was 15 years and the youngest was 11 years. The total number of students in junior high school form 2 for both schools who took part in the study was 75.

For junior high school form 3, the mean age was 14.26. The oldest student was 17 years and the youngest was 13 years. The total number of

students in junior high school form 3for both schools who took part in the study was 88.

All in all, 408 students in class 5 to junior high school form 3 with a mean age of 12.43 took part in the study.

Instrument

A self- report survey questionnaire with close-ended and open-ended questions were designed and administered to pupils from class 5 to junior high school form 3 in two schools in Tantra hill and Ofankor. One was a public school (school A) and the other was a private school (school B).

The questionnaire consisted of two sections. Section A discussed personal data of the students. Section B consisted of five questions about the repeated division of material and geometrical objects. It also included five questions on science and mathematics which were not about subdivisions. The questionnaire contained a total of 14 questions. This was given to each student in every class or form. Every question in section B on subdivision was explained verbally by me and demonstrated by a visual example. The rest of the questions were explained verbally, and time given for answering before the next one was presented. The instrument was developed with the help of Professor Yoav Yair of the Open University of Israel. Professor Yoav Yair and a colleague have a publication titled “Everything Comes to an End”: An Intuitive Rule in Physics and Mathematics.

Data Collection Procedure

Data for the study were collected in December 2011, at the two selected schools. School A at Tantra Hill and school B at Ofankor in the Ga West district of the Greater Accra Region. A total of eight visits were made,

four to each school. The first visit was to see the head teacher of school A and the principal of school B to introduce myself. The second visit was to meet teachers of the selected classes to find out what the students had been taught. The third and fourth visits were to administer the instruments to the students in classes 5 to Junior high school form 3. The questions were explained verbally by me and demonstrated by a visual example. The students were given time to answer each question before the next question was presented. Their responses to the answered questionnaire were then collected. This procedure by the researcher ensured 100 % retrieval of the questionnaire, while saving a lot of time.

Data Analysis

For questions under successive division, the students were given options to choose from. That is, yes (the process will come to an end) or no (the process can go on indefinitely). The students were then asked to give reasons for their choice. The arguments for the various responses were grouped and classified as finite or infinite. For example ‘When something is divided over and over again, at a point there will be no more of it’ or ‘it becomes small and cannot be divided anymore’ are finite answers. However, ‘It will never end because you can always divide it’ or ‘It will be left with smaller particles which can also be cut into other particles’ are infinite answers.

The research questions were analysed based on the students’ responses and explanations to the five questions on successive division. If students give four or five same responses to the questions on subdivisions, it means their answers were not affected by the nature of the objects. That is, whether

material or geometrical. Students who give four or five infinite answers to the questions on subdivisions have the intuitive rule, ‘Everything can be divided into two’ and perceive of matter as continuous in the context of subdivisions.

The tools used in the data analysis were descriptive statistics.The descriptive tools that were used included a table; showing the various classes, age range, mean age and number of students in each class. For the five questions on subdivisions, a bar chart showing the response of students in each class to each question was drawn. This was used to differentiate between the responses of the various classes and the trends. A graph, showing the percentage of children that gave identical answers to four or five of the questions on subdivision, as a function of class was used.This was to determine the persistence of students’ answers to questions of successive divisions. Distinction was made between finite and infinite answers on the graph. The reasons for the various responses were analysed to determine the intuitive rule the students have. The data were processed and analysed using the computer software, Statistical Package for Social Sciences (SPSS).

CHAPTER FOUR

RESULTS AND DISCUSSIONS

In this chapter, the findings from the study are presented and discussed in relation to the two research questions.

Preliminary Analysis of the trend of the responses of Students to the Five Questions on Subdivisions

In order to answer the two research questions, it is necessary to look at the trend of the response of students to the five questions on subdivisions in the questionnaire.

The first question on subdivisions in the questionnaire was on repeated division of a cube of sugar. The students were asked whether the repeated division of a cube of sugar will end. They were given options to choose from; ‘yes’, meaning it will come to an end or ‘no’, meaning it will go on indefinitely. They were then asked to give reasons for their answer. Their reasons were classified as ‘finite’ or ‘infinite’. Where the students did not give any explanation, it was classified as ‘no comment’.

Figure 1 shows the response of students according to class on the division of a cube of sugar.

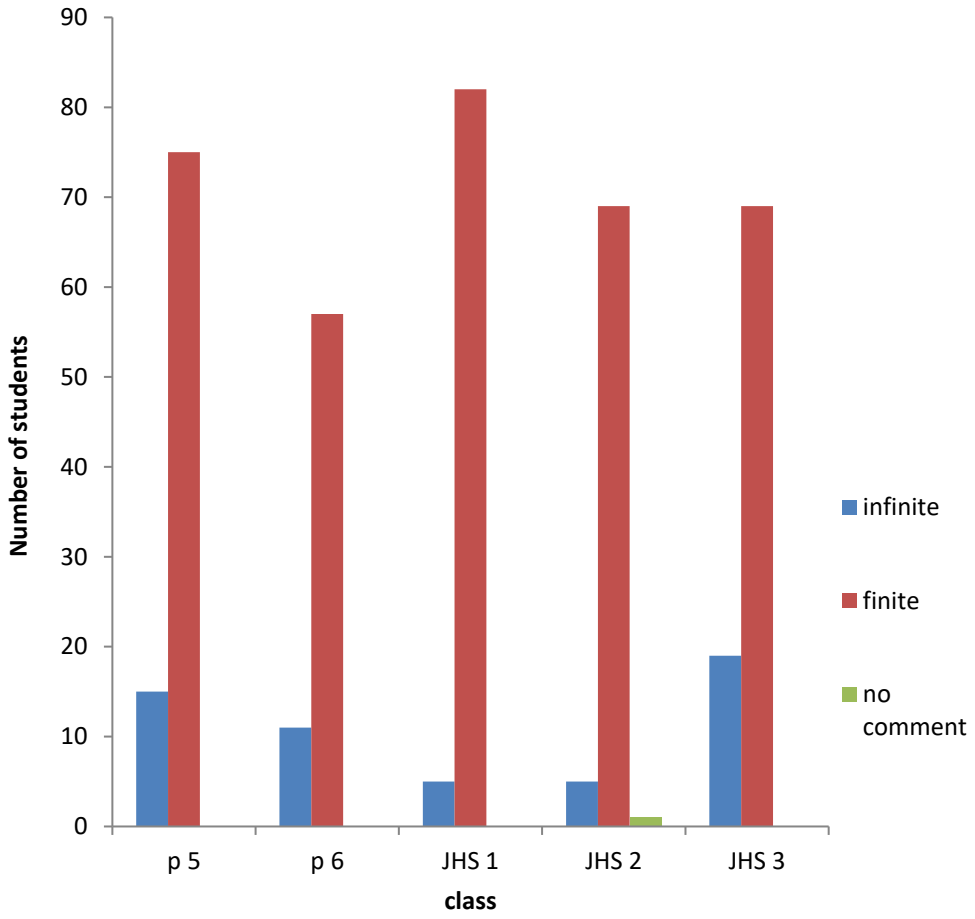


Figure 1: Response of students to the repeated division of a cube of sugar according to class

Most of the students from primary 5to junior high school form 3 explained that the repeated division of the cube of sugar will end. That is, the division of the cube of sugar is finite. However a number of students also explained that the repeated division of the cube of sugar will not end. That is, the division of the cube of sugar is infinite.

A few of the students in junior high school form 2 did not give any explanation; and were classified as no comment. The number of infinite answers decreased from primary 5 to junior high school form 1. From junior

high school form 2 to form 3, there was a rise in the number of infinite answers.

The second question on subdivisions was on serial dilution of sugar solution. The responses of the students were classified as finite, infinite and no comment. This is shown in Figure 2.

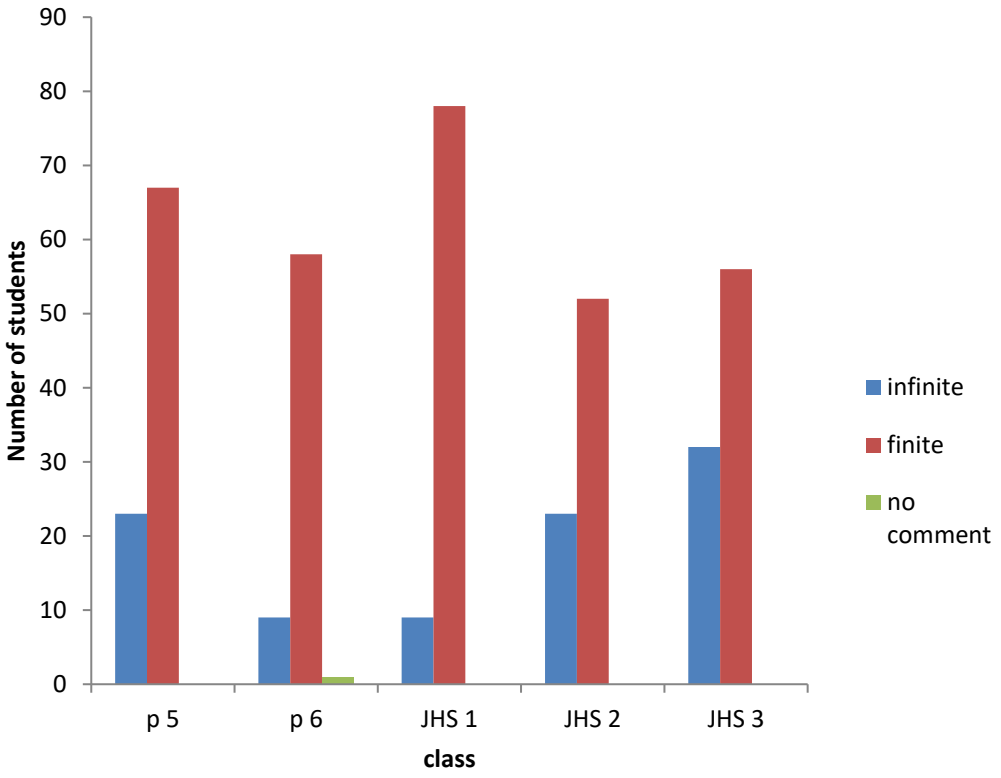


Figure 2: Response of students to the serial dilution of sugar solution according to class

The trend in the dilution of the sugar solution was similar to the trend in the division of the cube of sugar. However as compared to the question on the division of a cube of sugar, the number of infinite responses for each class increased considerably. A few of the students in primary 6 did not give any explanation and were classified as no comment.

The third question on subdivisions was on repeated division of a copper wire. The students were given options to choose from and give a

reason for their choice. Their reasons were classified as finite, infinite and no comment. This is shown in Figure 3.

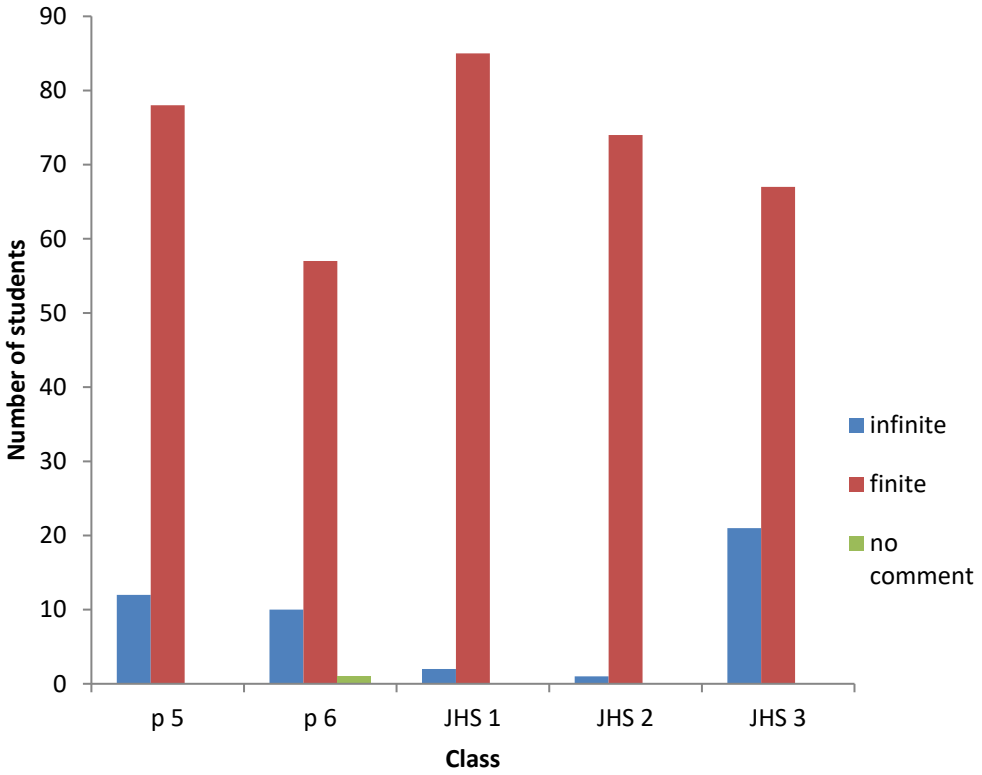


Figure 3: Response of students to the repeated division of copper wire according to class

The trend in the division of the copper wire was similar to that of the division of the cube of sugar. But generally, the infinite responses reduced a bit. A few of the students in primary 6 did not give any explanation; and were classified as no comment.

The fourth question on subdivisions was on repeated division of a line segment. The students were given options to choose from and to give a reason for their choice. Their reasons were classified as finite, infinite or no comment. The response for each class is shown in Figure 4.

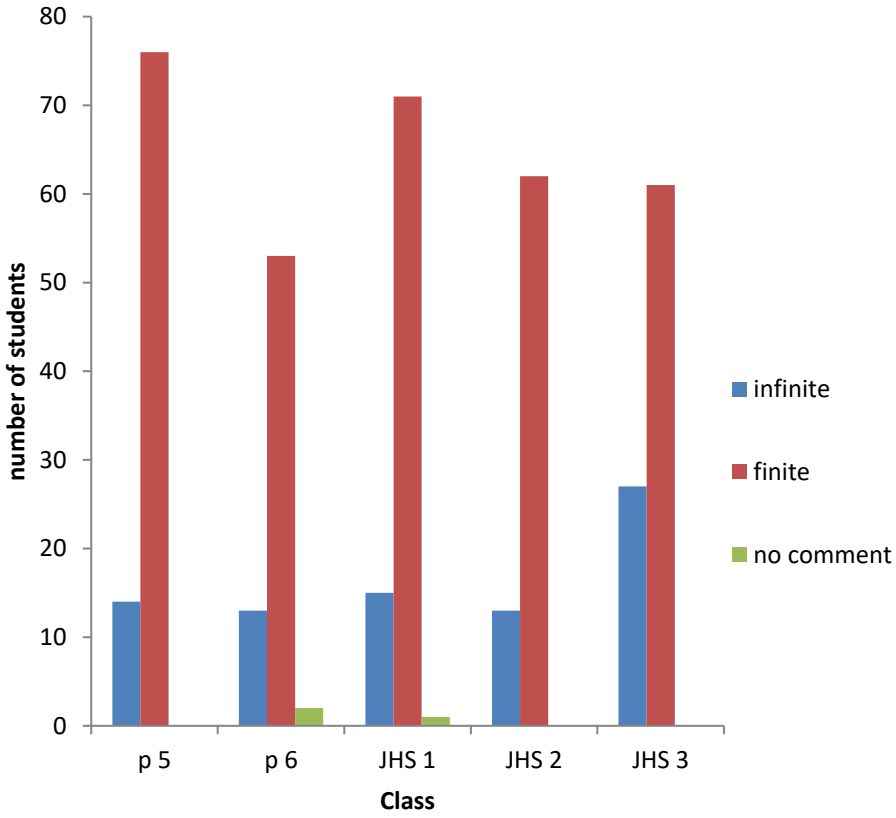


Figure 4: Response of students to the repeated division of a line segment according to class

In the division of the line segment, the trend was the same with majority of the students explaining that the division of the line segment is finite. However as compared to the division of the cube of sugar, the number of infinite responses for each class increased considerably. A few of the students in primary 6 and junior high school form 1 did not give any explanations; and were classified as no comment.

The fifth question on subdivisions was on repeated division of a rectangle. The students were also given options to choose from and give a

reason for their choice. Their reasons were classified as finite, infinite or no comment. The response for each class is shown in Figure 5.

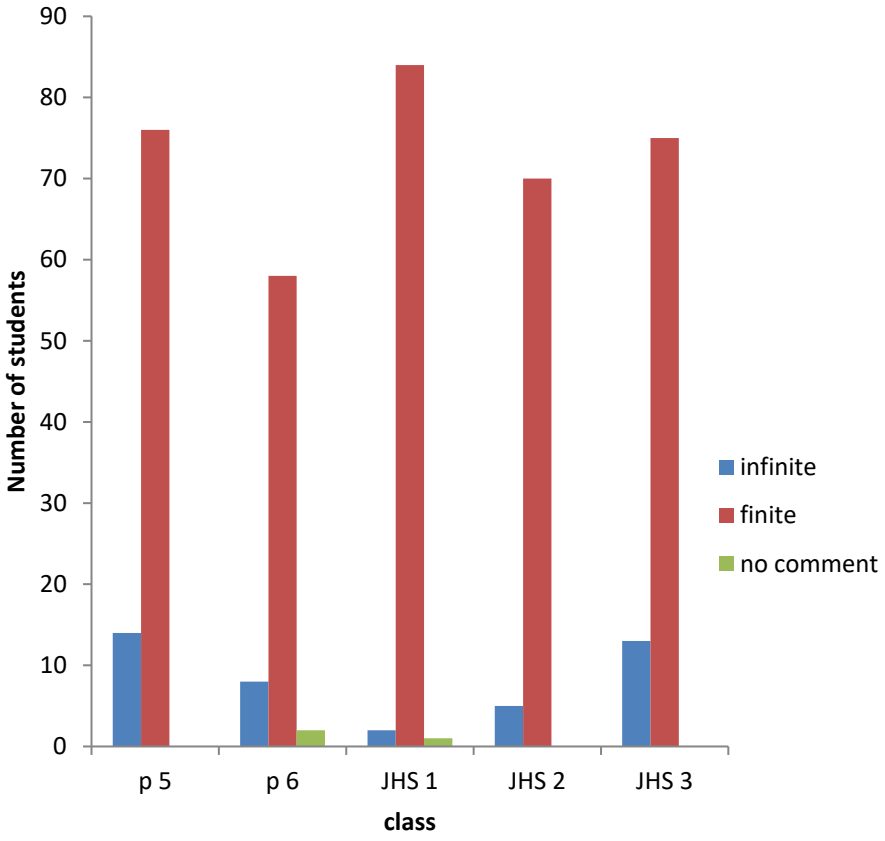


Figure 5: Response of students to repeated division of a rectangle according to class

In the division of the rectangle, the number of finite responses was high. That of the infinite responses decreased from primary 5 to junior high school form 1 and rose from form 2 to form 3. Compared to the division of the line segment, the number of infinite responses decreased considerably in all the classes. A few of the students in primary 6 and junior high school form 1 did not give any explanations; and were classified as no comment.

Research Question One

Does the intuitive rule ‘Everything can be divided into two’ exist in primary and junior high school students?

According to Stavy and Tirosh (2000), many of the responses described in literature as alternate conceptions could be interpreted as evolving from several common intuitive rules. Three such rules have been identified: ‘More A – more B’; ‘Same A – same B’; and ‘Everything can be divided’ (Stavy & Tirosh, 2000).

In this work, the existence of the intuitive rule ‘Everything can be divided into two’ in primary and junior high school students in the Ga West district of the Greater Accra Region was investigated.

The presence of intuitive rules can be investigated by using questions on the subdivision of material and geometric objects. An analysis of the student’s responses; relating it to the characteristics of intuitive rules and the intuitive rule theory will enable one to determine the intuitive rule the students have.

The questionnaire used for the research included five questions on subdivisions. Two of the questions were on the subdivision of material objects (a cube of sugar and a copper wire). Two were on the subdivision of geometric objects (a line segment and a rectangle). One was on the serial dilution of sugar solution.

The expected answer for the material objects is that the repeated division will end when the atomic level is reached. The expected answer for the geometric objects is that the repeated division will not end. Whilst the expected answer for the serial dilution of the sugar solution is that due to the

particulate nature of matter, after a large number of dilutions, the resulting solution might have a zero concentration of sugar (Stavy & Tirosh, 2000).

In order to investigate the intuitive rule of the students it is important to determine whether their answers to the five questions on subdivisions is affected by the nature of the objects or not. That is, whether material or geometrical. The percentage of students who gave same responses to four or five of the questions on subdivisions according to class is shown in Figure 6.

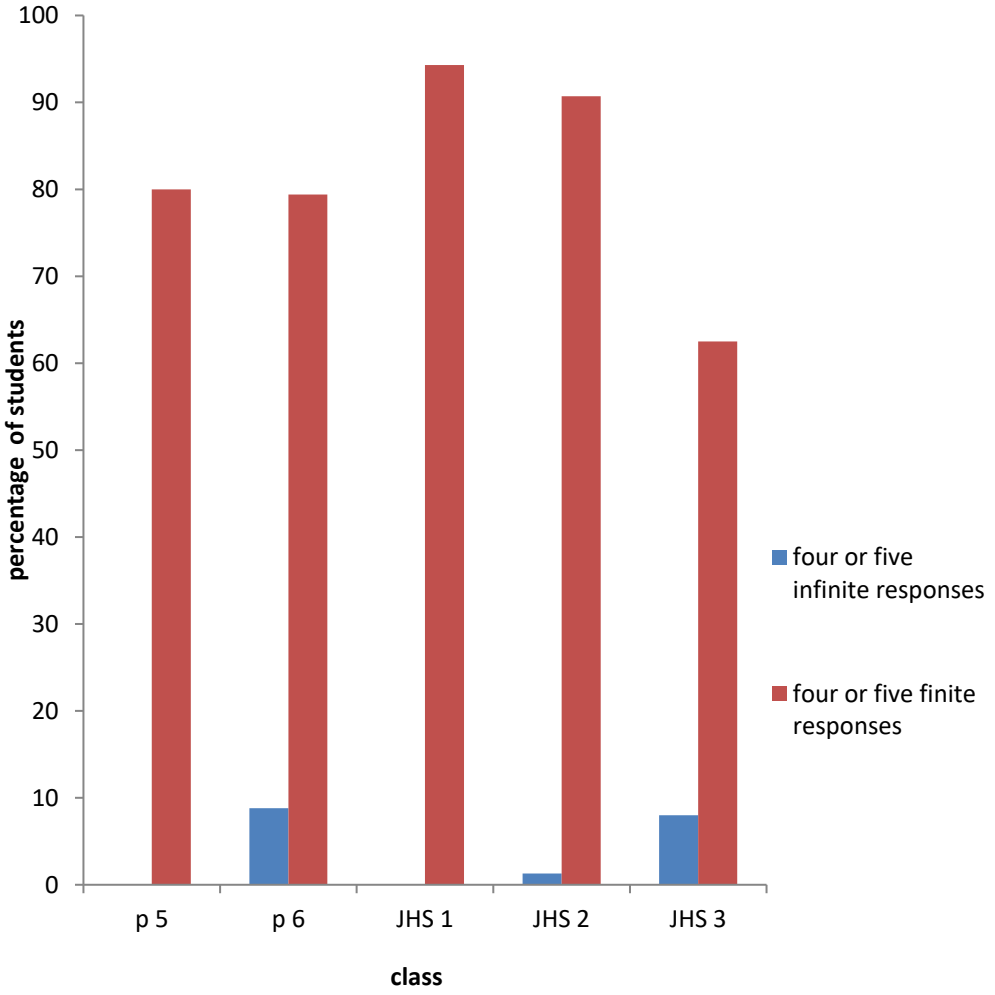


Figure 6:Percentage of students who gave four or five finite or infinite responses to the questions on subdivision according to class

Majority of the students gave four or five finite responses to the questions on subdivisions. However, a minority; 8.8% in primary 6, 1.3% in junior high school form 2, and 8% in junior high school form 3 gave four or five infinite responses.

Reasons for Answers

Primary 5

The total number of the students who took part in the study for primary 5 was 90. As shown on Figure 6, 80% of the students in primary 5 gave four or five finite responses to all the questions on subdivisions. This means that, their answers were not affected by the nature of the objects. Whether the object was material or geometrical, their answer was the same. The remaining 20% of the students gave mixed responses to the questions.

Some of the finite responses were;

R1: 'The more you divide, the shorter it becomes and comes to an end'.

R2: 'Everything has an end; therefore as you keep dividing the line, it will come to the end'.

R3: 'It will become small and we can't divide it again'.

R4: 'It will become smaller and smaller and there won't be any more to cut'.

This shows that majority of the students in primary 5 perceive that all divisions will end. The most dominant explanation was

R5: 'We cannot divide anymore because it will become small'.

The responses given by the students arein line with the characteristics of intuitive rules by Fischbein (as cited in Stavy et al, 2006; Yair & Yair, 2004). Therefore, there exists the intuitive rule 'All divisions will come to an end'in 80% of primary 5 students.

The remaining 20 % of the students gave mixed responses. Some of the infinite responses were;

R6: ‘This is because no matter how many times you divide it; there will be some left for you to divide’.

R7: ‘If you divide it again and again it will give an answer’.

R8: ‘Because a line never comes to an end’.

R9: ‘Because a rectangle is a horizontal and vertical line and never ends’.

Godfred, an 11 year old student in school A wrote;

R10: ‘It will not come to an end because when cutting up a carrot and you throw all away there will be pieces of the carrot left’.

Primary 6

The total number of students who took part in the study for primary 6 was 68 students. As shown on chart 6, 79.4% of the students gave four or five finite responses to the questions on subdivision. The percentage of students who gave four or five infinite responses to the questions on subdivisions was 8.8%.

Some of the finite responses were;

R1: ‘This is because as you divide it, it becomes smaller and smaller, till it becomes small and cannot be divided’.

R2: ‘It will become smaller and smaller and there won’t be any more to cut’.

R3: ‘Because as you keep on cutting it into pieces, it reduces and will definitely come to an end’.

R4: ‘The more you divide it, the more it gets smaller so it will surely come to an end’.

These are some of the common finite responses. Other finite responses were;

R5: ‘It becomes small that you cannot divide again because it is not going to last forever’.

Some of the infinite responses were;

R6: ‘The line will become small but will never end’.

R7: ‘It will not come to an end because the more you divide into half continuously, there will be the remaining half with you which can also be divided’.

R8: ‘As you cut, it will be very small, but it will not end. It can also be cut’.

On the serial dilution of the sugar solution, Maame, an 11 year old student in school B, wrote R9: ‘This is because as you mix and pour the water, there are still particles of sugar in the water, until you decide to pour all the water away; there will always be sugar in the water’.

From the response of the students and according to the characteristics of intuitive rules, 79.4% of the students in primary 6 have the intuitive rule: ‘All divisions will come to an end’, whilst 8.8% of the students in primary 6 have the intuitive rule, ‘Everything can be divided into two’.

Junior High Form 1

The total number of students who took part in the study for form 1 was 87 students. As shown on chart 6, 94.3% of the students, gave four or five finite responses to the questions on subdivision.Some of the finite responses were;

R1: ‘When you repeat this again and again, it will come to an end, because everything has a beginning and an end’.

R2: ‘As you keep dividing it, it will become a small particle which is very difficult to divide’.

R3: ‘It will become shorter and shorter and finally comes to an end’.

R4: ‘As you keep on cutting it into pieces, it will definitely come to an end’.

R5: ‘If you do it repeatedly, it will come to a point that it will be small and you cannot cut it’.

R6: ‘The more you divide it into halves, the smaller it becomes and when it is left with very small half, it cannot be divided so it comes to an end’.

R7: ‘When the beaker is re-filled several times, the sugar particles are gradually taken out so it will surely come to a stage where there will be no sugar particles in the water’.

R8: ‘The sugar particles will dissolve in the water. When we pour the mix of sugar water with ordinary water a lot of sugar particles will be poured away. The process repeated many times, it will end’.

From their responses 94.3% of the students in junior high school form 1 have the intuitive rule; ‘All divisions will come to an end’. The remaining students in form 1 gave mixed responses to the questions on subdivisions.

Some of the infinite responses were;

R9: ‘This is because when it is being cut for several times, it will be left with small particles which can also be cut into other particles’.

R10: ‘This is because half is always left so it can’t come to an end’.

R11: ‘It will not come to an end because a line has no end and you can continue dividing’.

Junior High Form 2

The total number of students who took part in the study for form 2 was 75 students. As shown on chart 6, 90.7% gave four or five finite responses to the questions on subdivisions.

Some of the finite responses were;

R1: 'This is because it is decreasing in size so it will come to a point when there will be no more sugar, because when something is divided over and over again at a point there will be no more of it'.

R2: 'Because the copper wire will become shorter and shorter and we will reach a stage where it will be too thin to be able to cut'.

R3: 'It will end because as we divide it, it becomes smaller and reduces in size till it gets finished'.

R4: 'Because if you keep on cutting it, by all means it will come to an end'.

R5: 'It will end because nothing lasts forever'.

R6: 'Everything has an end; you can cut and cut until it is left with the last piece that cannot be cut anymore'.

R7: 'It will come to an end because when cutting the copper wire, it reduces in size, so as long as it reduces in size, the process will end'.

R8: 'Yes, it will end, because when you continue breaking it, it becomes smaller (atom) which cannot be divided (indivisible)'.

R9: 'Because everything has a beginning and an end, that is, the more you divide it, the more it gets finished'.

R10: 'Because it will come to a time when there will be no sugar left'.

R11: 'If an even number is divided many times, it will come to zero and cannot be divided anymore'.

From the characteristics of the intuitive rules and the responses of the students, 90.7% of students in Junior high school form 2 have the intuitive rule: 'All divisions will come to an end'.

The percentage of students who gave four or five infinite responses was 1.3%.The remaining 8% gave mixed responses.

Some of the infinite responses were;

R12: ' Because as we keep on mixing and pouring water and adding water, part of the sugar particles still remains in the water in the beaker'.

R13: 'It will not come to an end even though the copper wire will be very small, the wire can be cut from ½ to ¼ to ⅛ to $^1/_{16}$ to $^1/_{32}$ and will go on and on until the wire may not be seen again'.

R14: 'Because when you keep dividing it into halves, you keep getting more halves so it does not end'.

R15: 'Because the line has no end, so it will not come to an end'.

R16: 'The sugar was stirred repeatedly; it mixes with the water making its particles not visible. It dissolves in the water making every quantity of water in the beaker have sugar in it'.

In Junior high school form two, 1.3% of the students have the intuitive rule: 'Everything can be divided into two'.

Junior High form 3

The total number of students who took part in the study in form 3 was 88. As shown on chart 6, 62.5% gave four or five finite responses to the questions on subdivision; 8.0% of the students gave four or five infinite responses to the questions on subdivisions; and the rest of the students; that is, 29.6% gave mixed responses.

Some of the finite responses were;

R1: ‘Yes; if you keep on repeating the process many times it will come to an end because it will become tiny pieces in which you cannot cut again’.

R2: ‘It reduces in size which makes it come to an end’.

R3: ‘It will end because the atom is the basic unit of matter and cannot be broken down to any other smaller unit’.

R4: ‘Because the basic unit of matter is an atom so eventually it will remain with an atom and it can’t be divided into any smaller unit’.

R5: ‘Because the sugar will get to the smallest unit called the atom where it cannot be broken down again’.

R6: ‘The copper wire reduces, becomes small and cannot be cut again’.

R7: ‘Because when you keep on dividing, it will reach a time that the cube of sugar will become like an atom, which is very small, and this will make it impossible to be divided or it won’t be divisible again’.

R8: ‘It will come to an end because when the cube is been divided, it reduces gradually and cannot be cut anymore’.

R9: ‘Everything comes to an end in this world, so it will end’.

R10: ‘It becomes small but nothing in this world has no end so at long last it will come to an end’.

R11: ‘It will become small and difficult to cut so it will surely come to an end’.

R12: ‘The more you divide it, the more it becomes smaller till it finishes up’.

R13: ‘Anything that decreases in size has an end’.

R14: ‘It will come to an end because the rectangle has a given area and with time the area will get finished’.

R15: ‘The process would come to an end because everything in life has an end and whatever we do also would have an end cause breaking the sugar into pieces, only particles would be left and you can’t divide’.

R16: ‘Because as the water is poured on the sugar, it becomes a dilute solution and as more sugar solution is poured out and ordinary water is added, the sugar dissolves along with water and gradually no sugar is left’.

In line with the characteristics of intuitive rules and the responses of the students, 62.5% of the students in form 3 have the intuitive rule ‘All divisions will end’.

Some of the infinite responses were;

R17: ‘Because if you keep on cutting, it will get to smaller particles which you can still cut into two’.

R18: ‘Even though the wire gets smaller we can still cut it over and over again no matter how small it is’.

R19: ‘It will become small but it maybe still broken into halves’.

R20: ‘A line is a distance between two points so it will never end’.

R21: ‘There will still be sugar in the beaker because the sugar dissolved in the water so it formed a solution. I.e. it formed part of the water so, so far as all the water is not thrown away, there will still be sugar in the water’.

R22: ‘No, because there will still be small or micro particles in the beaker and the presence of sugar will still be in water’.

R23: ‘Because lines are made up of points and points can be divided severally’.

R24: ‘Because, the sugar dissolved in the water and also a mixture of water and sugar cannot be separated by pouring away an amount of water’.

R25: ‘No, because we can divide the line into two parts no matter how small it is’.

R26: ‘It will never end because every half that is left becomes a whole so even if you cut the smallest piece, the half will becomes a whole’.

R27: ‘When the line segment is divided, a new line segment is created which also can be divided into two halves’.

R28: ‘A line has no end and can be divided continuously without ending’.

In form 3, 8% of the students have the intuitive rule: ‘Everything can be divided into two’.

From all the discussions, the majority of students used for the study, have the intuitive rule: ‘All divisions will end’. However a minority; 8.8% in primary 6, 1.3% in junior high school form 2, and 8.0% in junior high school form 3 have the intuitive rule: ‘Everything can be divided into two’.

Research Question Two

In the context of subdivisions, do primary and junior high school students perceive of matter as discrete or continuous?

The usual approach in studies of children’s concepts on the continuity of matter is to ask questions on subdivisions such as whether it is possible to repeatedly divide a certain material or mathematical object. When the answer to such a question is that the process terminates, that is, the number of subdivisions is finite; this may point to an atomistic, discrete, perception of the object in question. An answer that claims that the process is infinite, that is, the number of subdivisions is unlimited points to a continuous perception of matter (Yair & Yair, 2004).

In primary 5, majority of the students, which is 80%, gave 4 or 5 finite responses. Their answers were not affected by the nature of the questions whether material or geometric. This implies that in the context of subdivisions, the students perceive matter as discrete.

In primary 6, 79.4% of the students gave 4 or 5 finite responses. The students had previously been taught about matter, states of matter and change of states of matter. However their answers were very similar to those in primary 5. Therefore it implies that in the context of subdivisions, the students perceive matter as discrete. However, 8.8% of the students gave 4 or 5 infinite responses to the questions on subdivision. Their answers were not dependent on whether the substance was material or geometric, therefore they perceive of matter as continuous in the context of subdivisions.

In junior high school form 1, 94.3% gave 4 or 5 finite responses to the questions on subdivisions. This implies that in the context of subdivisions, the students perceive matter as discrete.

In junior high school form 2, 90.7 % of students gave 4 or 5 finite responses. The students had previously been taught about elements, compounds and the atomic structure. Therefore 90.7% of the students in junior high school form 2 perceive matter as discrete. However, 1.3% of the students gave 4 or 5 infinite responses and hence perceives matter as continuous.

In junior high school form 3, 62.5% of the students gave 4 or 5 finite responses. Therefore they perceive matter as discrete. However, 8.0% of the students gave 4 or 5 infinite responses and hence perceive matter as continuous.

From the foregoing discussion, majority of the students in the study perceive of matter as discrete in the context of subdivisions. However a minority; 8.8% in primary 6; 1.3% in junior high form 2, and 8.0% in junior high form 3 perceive matter as continuous.

CHAPTER FIVE

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

Overview

This study was conducted to investigate the existence of the intuitive rule 'Everything can be divided into two' in primary and junior high school students. It was also aimed at determining the students' perception of matter in the context of subdivisions. The study sought answers for the following research questions:

1. Does the intuitive rule 'Everything can be divided into two' exist in primary and junior high school students?
2. In the context of subdivisions, do primary and junior high school students perceive mater as discrete or continuous?

In all, 408 students in class 5to junior high school form 3 in two schools on the Ga west district of the Greater Accra Region were used for the study. The two schools were: a public school (school A), where the income levels of parents are perceived to be relatively low and a private school (school B), where the income levels of the parents are perceived to be very high.

The study used a self report survey questionnaire with close–ended and open –ended questions which were administered to the students. It contained fourteen questions. Five of the questions were on subdivisions and students were asked to give a reason for their choice. The responses of the students were analysed to determine the intuitive rule they had and to give the

researcher a better understanding of their reasoning processes when solving these tasks.

Summary of Key Findings

1. Analysis of students' response on the existence of the intuitive rule 'Everything can be divided into two' revealed the following:
    A. In primary 5 where the mean age of the students was 10.6, there exists the intuitive rule 'All divisions will come to an end' in 80% of the students. This means that 80% of the students in primary 5 perceive that subdivisions must eventually cease.
    B. In primary 6 where the mean age of the students was 11.37, there exists the intuitive rule 'All divisions will come to an end' in 79.4% of the students. This means that 79.4% of the students in primary 6 perceive that subdivisions must eventually cease.However, 8.8% of the students in primary 6 have the intuitive rule 'Everything can be divided into two. This means that 8.8% of students in primary 6 perceive that subdivisions will never cease.
    C. In Junior High School form 1 where the mean age of the students was 12.44, 94.3% of the students have the intuitive rule 'All divisions will come to an end'.
    D. In Junior High School form 2, where the mean age of the students was 13.44, 90.7% of the students have the intuitive rule ' All divisions will come to an end' whilst 1.3% of the students in Junior High School form 2 have the intuitive rule 'Everything can be divided into two'

E. In Junior High School form 3 where the mean age of the students was 14.2, 64.5% of the students have the intuitive rule ‘All divisions will end’, whilst, 8% of the students have the intuitive rule ‘Everything can be divided into two’.

2. Analysis of the students response on their perception of matter in the context subdivisions revealed the following:
   A. In primary 5, 80% of the students perceive matter as discrete in the context of subdivisions.
   B. In primary 6, 74.9% of the students perceive matter as discrete, whilst 8.8% of the students perceive matter as continuous, in the context of subdivisions.
   C. In Junior High School Form 1, 94.3% of the students perceive matter as discrete in the context of subdivisions.
   D. In Junior High School Form 2, 90.7% of the students perceive matter as discrete whilst, 1.3% of the students perceive matter as continuous in the context of subdivisions.
   E. In Junior High School Form 3, 62.5% of the students perceive matter as discrete whilst, 8% of the students in form three perceive matter as continuous in the context of subdivisions.

Conclusions

Based on the findings from this study, the following conclusions can be drawn:

Majority of the students used for study have the intuitive rule ‘All divisions will end’, meaning that subdivisions must eventually cease. However, a minority, 8.8% in primary 6, 1.3% in Junior High School

Form 2 and 8.0 % in Junior High School Form 3, have the intuitive rule, ‘Everything can be divided into two’.

Majority of the students perceive matter as discrete in the context of subdivisions. However a minority; 8.8% in Primary 6, 1.3% in Junior High School form 2 and 8% in Junior High School form 3 perceive matter as continuous.

Recommendations

The following recommendations have been made for educational practice:

1. In teacher education, it is important to raise teachers’ awareness of the role of intuitive rules in students thinking, so as to enable the teachers to foresee possible cognitive obstacles.
2. In teaching the topic ‘matter’, it is necessary for the teachers to find out the students’ perception. The lesson can then be structured to enable students change their perception to the scientifically correct one.
3. The intuitive rule is ‘triggered’ by subdivision tasks such as repeated halving or decreasing series. Therefore teachers should be trained on how to determine the intuitive rule in students, when teaching topics which involve these tasks so that they can plan their instruction to enable their students overcome the effects of the intuitive rule.

Suggestions for Further Research

1. The research should be replicated in Senior High Schools to find out the existence of the intuitive rule.
2. This research should also be replicated in Primary and Junior High Schools in other districts and regions in Ghana to determine how widespread the intuitive rule is used by students.

3. The other two intuitive rules not covered in this study can also be investigated in Primary, Junior High and Senior High School students.

REFERENCES

Aboagye, G. K. (2009).*Comparison of learning cycle and traditional teaching approaches on students' understanding of selected concepts in electricity.* Unpublished Master of Philosophy Thesis, University of Cape Coast.

Amponsah,K.D. (2009). *KNUST Female students' participation in Science and Technology- based careers: Factors, Perceptions and Constraints.* Unpublished Master of Philosophy Thesis, University of Cape Coast.

Anderson, B. (1986). The experiential gestalt of causation: A common core of pupil's preconceptions in science. *European Journal of Science Education,8,* 155-171.

Babai, R. (2009).Piagetian cognitive levels and the tendency to use intuitive rules when solving comparison tasks.*International Journal of Science and Mathematics Education,8*, 203 -221.

diSessa, A. A.(1983). Phenomenology and the evaluation of intuition. In D.Genter,& A. L, Stevens (Eds.),*Mental modules* (pp 15-33) Hillsdale, NJ: Erlbaum.

Driver, R. (1994).*Making sense of secondary science*. London: Routledge.

Driver, R., Guesne, E.,&Tiberghien, A. (Eds.).(1985). *Children's ideas in science.* Milton Keynes, England: Open University Press.

Erickson, G. (1979). Children's conceptions of heat and temperature.*Science Education,63*(2), 221-230.

Erickson, G., &Tiberghien, A. (1985).Heat and temperature. In R. Driver, E.Guesne, & A.Tiberghien (Eds.),*Children's ideas in science* (pp 52-84). Milton Keynes, England: Open University Press.

Gutierrez, R.,&Ogborn, J. (1992).A causal framework for analysing alternative conceptions. *International Journal of Science Education,14*, 201-220

Nussbaum, J. (1985). The particulate nature of matter in the gaseous phase. In R. Driver, E. Guesne, & A. Tiberghien (Eds.), *Children's ideas in science* (pp 124-144). Milton Keynes, England: Open University Press.

Piaget, J. (1965).*The child's conception of number*.New York; Norton.

Piaget, J. (1968).*On the development of memory and identity*.Barre, M A: Clark University Press.

Piaget, J. (1969).*The child's conception of Time.* New York: Basic Books.

Piaget, J.,&Inhelder, B. (1974).*The child's construction of quantities.*London: Routledge.

Piaget, J., Inhelder, B.,&Szeminska,A. (1960).*The child's conception of geometry.* London: Routledge&Kegan Paul.

Pine, K., Messer, D.,& St. John, K. (2001). Children's misconceptions in primary science: A survey of teachers' views. *Research in Science and Technological Education, 19*(1), 79-96.

Stavy, R., Babai, R., Tsamir, P., Tirosh, D., Lin, F. L., & McRobbie, C. (2006). Are intuitive rules universal? *International Journal of Science and Mathematics Education,4*, 417-436.

Stavy, R., & Tirosh, D. (1993). Subdivision processes in mathematics and science. *Journal of Research in Science Teaching,30*(6), 579-586.

Stavy, R., & Tirosh, D. (1996a).The role of intuitive rules in science and mathematics education.*European Journal of Teacher Education,19*(2), 109-119.

Stavy, R., & Tirosh, D. (1996b). Intuitive rules in science and mathematics: The case of ‘more of A -more of B’. *International Journal of Science Education,18*(6), 653-667.

Stavy, R., & Tirosh, D. (2000). *How students (mis-)understand science and mathematics: Intuitive rules*. New York: Teachers College Press.

Tirosh, D. (1990). Inconsistencies in students’ mathematical constructs.*Focus on learning problems in mathematics, 12*, 111 – 129.

Tirosh, D., & Stavy, R. (1996). Intuitive rules in science and mathematics: The case of ‘Everything can be divided by two’. *International Journal of Science Education,18*(6), 669-683.

Tirosh, D., & Stavy, R. (1999). Intuitive rules: A way to explain and predict students’ reasoning. *Educational Studies in Mathematics,38,* 51-66.

Vosniadou, S., & Brewer, W. F. (1987).Theories of knowledge restructuring in development.*Review of Educational Research,57,* 51-67.

Yair, Y., & Yair, Y. (2004). ‘Everything comes to an end’: An intuitive rule in physics and mathematics. *Science Education,88,* 594 - 609.

APPENDIX A

QUESTIONNAIRE

This questionnaire is part of a university study. For its success, it is important that you answer all the questions. Please write your answers with a pen.

Thank you for your cooperation!

SECTION A

(i) Age ……

(ii) Do you have a private teacher at home? Mark: [yes] or [no]

(iii) If yes, what subjects does he/she teach you?

……

…….....

(iv) Form …….

SECTION B

1. What is matter?

   .....

   ……

2. List the states of matter

   …….....

   ……

3. On the table is a cube of sugar. We break it into two halves, and take away one half. We break again the remaining half, again taking away half. We repeat this process many times. Will this come to an end?

   Mark: [yes] or [no]andexplainyour answer.

   ……

……

……

……

4. We take a beaker filled with water (about 200 mls) and add to it a cube of sugar and mix them very well. From the mixture of sugar and water, pour away half. We then add only water to fill the beaker. We re-mix it very well, and pour away half. We add more water and re-mix. Pour away half and repeat the process again and again. Will there be a stage when there will no longer be sugar in the beaker? Mark: [yes] or[no] and explain your answer.

   ……

   ……

   ……

5. What happens to sugar when it is mixed with water?

   ……

   ……

6. You have a copper wire. We cut it in two identical parts with a pair of pliers. We take one of the parts and cut it again in half. We repeat this again and again. Will this process come to an end? Mark: [yes] or[no] and explain your answer.

   ……

   ……

   ……...

7. What is a line?

   ……

   ……

8. How many points are there in a line?

   ……..

9. On the sheet of paper provided, draw a line. Divide it in half. Again divide one of the halves in two. Repeat this again and again. Will this process of dividing the line segment come to an end? Mark: [yes] or[no] and explain your answer.

   ……

   ……

   ……

10. On the sheet of paper provided, draw a rectangle. Divide it to half. Again, divide one of the halves in two. Repeat this again and again. Will this process of dividing the rectangle come to an end? Mark: [yes] or[no] and explain your answer.

    ……

    ……

    ……

    …….....

APPENDIX B

Table 1:Age description of Students

	Primary 5	Primary 6	JH Form 1	JH Form 2	JH Form 3
Number of Students	90	68	87	75	88
Minimum Age	8	10	11	11	13
Maximum Age	14	14	16	15	17
Mean Age	10.66	11.37	12.44	13.44	14.20