UNIVERSITY OF CAPE COAST

ASSESSING LABORATORY SKILLS OF PHYSICS STUDENTS IN SELECTED SENIOR HIGH SCHOOL TOPICS IN MECHANICS AND OPTICS

SALIFU MAIGARI MOHAMMED

2010

UNIVERSITY OF CAPE COAST

ASSESSING LABORATORY SKILLS OF PHYSICS STUDENTS IN SELECTED SENIOR HIGH SCHOOL TOPICS IN MECHANICS AND OPTICS

BY

SALIFU MAIGARI MOHAMMED

Thesis submitted to the Department of Science and Mathematics Education of the Faculty of Education, University of Cape Coast, in partial fulfilment of the requirements for award of Master of Philosophy Degree in Science Education

SEPTEMBER 2010

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: Salifu Maigari Mohammed.

Supervisors' Declaration

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

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

Name: Professor Theophilus Aquinas Ossei-Anto

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

Name: Mr. Eugene Adjei Johnson.

ABSTRACT

The purpose of the study was to design and develop assessment instruments (tasks) to assess laboratory planning, performing and reasoning skills of physics students in selected senior high school topics in mechanics and optics. The accessible population was 551 SHS 3 physics students in seven schools within Sekondi-Takoradi metropolis. Two hundred and eighty nine students were sampled for the main study.

Simple and stratified random sampling methods were employed in selecting the students. The "basic skills assessment testing" method also known the "psychometric testing" approach was chosen for the study. The instruments used were three performance assessment tasks.

Scoring formats were used to score the students' responses. Another physics teacher was given 111 subsamples of the scripts to score. Inter-rater reliabilities for the tasks were: 0.93 for Task A, 0.96 for Task B, 0.94 for Task C, and 0.93 for the total tasks.

Means, standard deviations, frequencies, percentages, t-tests and ANOVAs were estimated. Results and findings from the research show that (a) majority of the students demonstrated high levels of competency in laboratory planning, performing and reasoning skills; (b) Male and female students demonstrated similar levels of planning, performing and reasoning skills. (c) Particularly, students from girls schools exhibited higher proficiency in reasoning skills than girls from mixed schools.

It was recommended that physics students should be given more opportunities to practice hands-on activities. It was also recommended that physics teachers should do more performance assessments in school.

ACKNOWLEDGEMENTS

I acknowledge the effort of my principal supervisor, Professor Theophilus Aquinas Ossei-Anto, who agreed in principle to supervise my research work long before I was officially assigned to him because the work was in his area of specialization. I sincerely express my gratitude to him for giving me good suggestions and directions, and for having time and will to supervise my research work that has enabled me to complete in time. I also acknowledge the effort of my co-supervisor, Mr. Eugene Adjei Johnson, who contributed in no small ways by advice, suggestions and directions that have enabled me to complete the study.

I hereby acknowledge the immense help of Mr. Ebenezer Kweitu of Holy Child School, Cape Coast; Mr. Isaac Mensah and Miss Martha Watts who are laboratory technicians at the Department of Science and Mathematics Education, UCC; Mr. Alfred Lumor who is a store keeper at the physics department, UCC; Messrs A. K. Ampah and Macphil Maclight of Chemistry Department, UCC; Mr. James Dadson who is an old classmate and a chemistry tutor at Takoradi Secondary School. This research would have been very difficult to complete without the immense help and support of these aforementioned people.

I also express my profound gratitude to Mr. Godwin Aboagye of the Department of Science and Mathematics Education, UCC, for his invaluable contribution toward the completion of this research. My sincere gratitude also goes to an old friend, Baba, and his wife and children, at Effiekuma night market in Takoradi, for their tremendous assistance and support that made it possible for me to carry out this research in Takoradi. My sincere gratitude once more goes to Messrs Usman Imoro and William K. Akromah of Library Basement Research Centre, UCC, as well as to Mr. Anthony Afful of Archbishop Porter's Secondary School, Takoradi.

Again, I acknowledge the contribution of the headmasters, headmistresses, assistant headmasters and assistant headmistresses, heads of science departments, physics teachers and students of the senior high schools in Cape Coast and Sekondi-Takoradi metropolises who gave me good cooperation that made this study possible. My sincere gratitude and appreciation also go to any person or group of persons who helped toward the successful completion of this research but whose names are not mentioned here.

DEDICATION

I dedicate this study to my father, Malam Mohammed Maigari, and my late mother, Fati Ali Achimfo, for laying a foundation in formal education for me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

Background to the Study

Assessment is the process for obtaining information that is used for making decisions about students, curricula and programmes, and educational policy (Nitko, 2004). Decisions about students include selecting them for educational opportunities, and credentialing and certifying their competence. According to Nitko (2004), assessing a student's competence means collecting information to help decide the degree to which the student has achieved the learning targets. Evaluation on the other hand is the process of making valued judgments about the worth of a student's product or performance (Nitko, 2004). Evaluations may or may not be based on measurements or tests results. Testing and measurements, however, reduce some of the inconsistency and subjectivity that influence evaluation; because they are more standardized and objective than other assessment techniques (Nitko, 2004).

Various types of paper-and-pencil tests are taken during years of school. The widespread use of paper-and-pencil tests in schools is largely due to their efficiency in measuring a large number of learning outcomes and the ease of scoring and recording the results (Gronlund, 1998). According to Gronlund (1998), regular assessment and feedback of the results can help students to gain insight into what they can do well, the misconceptions that they need to correct, and the extent of the skills they have in various areas.

However, disquiet has been expressed about examinations. Both universities and employers have serious reservations about the predictive reliability of examinations (Lloyd-Jones & Bray, 1986). Examinations, it is argued, test mainly trivial and irrelevant learning outcomes and memorization of factual information (Lloyd-Jones & Bray, 1986). Amuah (1996) cites Gardner and Wolf as saying that the paper-and-pencil strategy of assessing students has come under severe criticism by scholars of education. It is argued that the results obtained from paper-and-pencil tests are not wholly true reflections of students' academic achievement; and that it alienates students from participating in decisions affecting the evaluation of their academic achievements. Paper-and-pencil tests make teachers and students "examination conscious"; and that teachers tend to spend most of their valuable teaching periods to coach students in strategies for passing examinations. Teacher-developed paper-and-pencil tests and many tests and quizzes provided by textbook publishers are currently dominated by questions that ask students to recall facts and information (Cangelosi, 1990). Amuah (1996) state that the assessment of students' academic achievement has been receiving increasing attention all over the world in recent years, including Ghana.

The media coverage and parliamentary debate on the 1994 senior secondary school examination results are indices of the high premium placed on the evaluation of students' academic achievement by the Ghanaian public. Again, Amuah (1996) cites Roth as stating that the worldwide search for better strategies for assessing academic achievement has been the focus of attention of prominent educators in many countries in the world. Ghana is not left behind in this search and has initiated programmes geared towards the improvement in the quality of assessment procedures used in its schools (Amuah, 1996).

Much more importance is being placed on skills such as diagnosis and solution of problems, and implementation of solutions; scientific enquiry and the ability to work together cooperatively and to be creative. According to Achbald and Newman, and Shepard (as cited in Marzano, Pickering and Mctighe, 1993), the last decade has witnessed a growing recognition of the need for significant changes in educational assessment practices. A broad array of both academic and non-academic competencies is necessary for the modern workplace (Marzano *et al*., 1993). There are several movements calling for increase in realism in assessment by moving toward extended performance assessment and to make the traditional paper-and-pencil tests more authentic. Gronlund (1998) is of the view that education is better served by using both paper-and-pencil testing and the assessment of actual performance.

Performance assessment goes beyond multiple choice and essay tests by evaluating student behaviour in lifelike situations designed to elicit the knowledge and skills of interest (Mcdaniel, 1994). According to Marzano *et al*., (1993) performance assessment is a systematic attempt to measure a learner's ability to use previously acquired knowledge in solving novel problems or completing specific tasks. The terms alternative assessment authentic assessment, and performance assessment are all used in discussion of assessment reforms. These terms have different meaning although they are used synonymously. The evaluation of laboratory performance provides an important example of alternative or authentic assessment (De Ture, Fraser, Giddings & Doran, 1995). Anthony-Krueger (2001) cites Tamir as stating that laboratory practical examinations which require manipulation of some materials and which involve direct experience of the examinee with the materials at hand are classic examples of science performance assessment; and therefore recommends the use of practical laboratory examination as a means of assessing students' performances. According to Ossei-Anto (1996), practical work is done in the laboratories to inculcate the scientific spirit in the students which can lead to the acquisition of laboratory skills of planning, performing, observing, and reasoning.

The Trends in International Mathematics and Science Study (TIMSS) provides reliable and timely data on the mathematics and science achievement of the United States of America (U.S.) 4th- and 8^{th}-grades students compared to that of students in other countries, including Ghana (National Centre for Education Statistics [NCES], 2007). TIMSS is developed and implemented at international level by the International Association for the Evaluation of Educational Achievement (IEA), which is an international organization of national research institutions and governmental research agencies. According to NCES (2007), TIMSS data have been collected in 1995, 1999, 2003 and 2007. TIMSS is used over time to measure the mathematics and science knowledge and skills of fourth-and eighth-graders. TIMSS is designed to align broadly with mathematics and science curricula in the participating countries. The results of TIMSS therefore suggest the degree to which students have learned mathematics and science concepts and skills likely to have been taught in school. Achievement results from TIMSS are reported on a scale from 0 to1000, with an international (TIMSS) average of 500 and a standard deviation of 100. The TIMSS benchmarks describe four levels of students' achievement in mathematics and science based on the kinds of skills and knowledge students at each score cut point would need to successfully answer the mathematics and science items.

The TIMSS science assessment is designed to cover the science topics or content that the students are expected to learn and the cognitive skills students are expected to have developed. At the eighth-grade the content domains are Biology, Chemistry, Physics and Earth Science. The cognitive domains are: knowing, applying and reasoning. In 2003, the international average was 473 points on the TIMSS scale. The Ghanaian eighth-graders average was 255 points which is 218 points below the international average. With this score Ghana placed 44th out of the 45 countries that participated in TIMSS 2003 (NCES, 2007). Again, Ghana placed 47th out of the 47 countries that took part in TIMSS 2007. The Ghanaian eighth-graders average score was 303 points which is 197 points below the international average of 500 in 2007. Other African countries like Algeria, Botswana, Egypt and Tunisia that participated in TIMSS 2007 were all above Ghana in the placement (NCES, 2007).

The chief examiner's reports of the Senior Secondary Schools Certificate Examinations also point out bitterly that the performance of candidates in physics were below expectation; and that candidates did well only on questions which demanded simple recall of factual knowledge. For questions on description of experiments, candidates dwelt much on the apparatus assembled instead of explaining how (process) the experiments were performed (West African Examinations Council [WAEC], 2004).

In Ghana, there have been numerous educational reforms since independence, aimed at making education more relevant and practical to the needs of the country by increasing attention paid to problem solving, environmental concerns, prevocational training, and laboratory skills. However, very little has been achieved in the light of all these reforms, especially for schools and students in the rural communities. The results of the final senior secondary schools certificate examinations throughout the country testify to that (Eyiah, 2004). A press review report (Daily Graphic, 2007, April 12) states that Ghana's educational system so far has produced school leavers and graduates who have not been adequately prepared for the world of work. According to the report, in spite of all the educational reforms embarked upon previously, the country has failed to find answers to the problem of churning out school leavers and graduates who do not have the competencies and skills to enable them to continue their education or embark on continuous learning for self-improvement. The press review report (Daily Graphic, 2007, April 12) went on further to say that a sober look at the country's educational system has brought out major failures, which if left unresolved would hinder the efforts to propel Ghana to great future.

Therefore the goal of the 2007 education reform in Ghana is to equip the youth of the country to meet challenges of the 21st century; and that the desire and aspiration of Ghana to reach a middle-income status has made it imperative for the youth to be equipped to enable them contribute their quota towards national development. The youth are therefore consequently to be equipped to be on top of science, information and communication technology, technical training, vocational and agricultural disciplines. The reform places emphasis on mathematics, science and technology and problem solving. It also aims to heighten awareness of the environment to preserve national resources.

The press review report further states that the senior high school system is to be organised both as terminal education for entry into the world of work, and as a preparatory stage for entry into tertiary education; and that the first two years at the senior high school level, mathematics, computer science, general science, social studies and English will be compulsory subjects.

In their efforts to develop and validate instruments to assess laboratory skills of students in high school physics courses, Ossei-Anto (1996) and Johnson (2001) produced prototype instruments for assessing the level of skills possessed by physics students. The method of science laboratory procedure adopted by Ossei-Anto (1996) and Johnson (2001) consists of three categories: planning, performing and reasoning which can be likened to the pre-lab, lab and post-lab activities that take place in most of Ghana's senior high school science classes. During the planning stage, students work on designing basic experimental procedures to a problem. In the performing stage, students carry out actual experiments, manipulate objects, make observations, record data, and take decisions about a practical strategy. During the reasoning/analysis stage, students process data, explain relationships, form generalizations, discuss data accuracy, and provide sources of error and limitations to an experiment. The model for this research follows that of Ossei-Anto (1996) and Johnson (2001).

Statement of the Problem

There is a general low academic achievement of physics students in Ghana, as is evidenced in the achievements of Ghanaian 8th grades students in TMSS 2003 and 2007 (NCES, 2007); the newspaper press review report (Daily

Graphic, 2007, April 12); and the chief examiners' reports of the senior secondary schools certificate examinations (WAEC, 2000, 2001, 2003, 2004, 2005, 2006). However, it seems particularly that physics students in Ghana do not show adequate competencies in laboratory planning, performing and reasoning skills required of senior high school (SHS) students.

The chief examiner's reports specifically state that majority of candidates were unable to give correct traces of light paths through triangular glass prism; and that candidates could not accurately measure emergent angles, incident angles and angles of deviation from the light traces they produced. The reports also said that most of the traces of light paths through prism produced by candidates did not have arrows to indicate the correct directions of light rays. The reports further stated that some candidates measured wrong angles of deviation and emergent angles.

Johnson (2001) reports that most science educators currently assume that senior high school students do not show and exhibit satisfactory level of planning, performing and reasoning skills when confronted with practical issues in the laboratory. Kojima, Lynch and Web, Newman and Tamir (as cited in Ossei-Anto, 1996), also point out that students do not demonstrate adequate proficiency in skills of planning, performing and reasoning while engaged in traditional school laboratory activities.

Purpose of the Study

The purpose of this research was to design and develop assessment instruments (tasks) to assess laboratory skills competencies of some selected physics students in selected topics in mechanics and optics; to determine if they show adequate laboratory planning, performing and reasoning skills. Every science student should be able to demonstrate at least minimum competency in each of the laboratory skills of planning, performing and reasoning.

Research Questions

In the quest to find some answers to the problems confronting senior high school (SHS) physics students, regarding the proficiency of their laboratory skills, the problems were broken down into clearly defined and specific questions. These questions served as the research questions that guided the conduct of this research. The research questions are as follows:

1. To what extent do senior high school physics students engaged in laboratory work exhibit adequate competencies in the skills of:

i. Planning?

ii. Performing?

iii. Reasoning?

2. To what extent do senior high school physics students engaged in laboratory work exhibit adequate competencies in:

i. One aspect of the skills of planning, performing and reasoning?

ii. All the three skills of planning, performing, and reasoning?

3. To what extent are male senior high school physics students more skilful than their female counterparts?

4. To what extent are the skills of planning, performing and reasoning demonstrated by a student related to the type of school (boys, girls or mixed) the student attends?

In addition to the tasks that were administered to the students, opinions of students were sought about the time allocation and difficulty level of the tasks using an opinionnaire. The responses of the students to the opinionnaire helped the researcher to determine whether the performance of the students on the tasks were affected by the difficulty of the tasks, as well as the time allocated for completing the tasks.

Variables Used in the Study

The variables used in the research are:

a. Dependent Variables

i. Score for Task A

ii. Score for Task B

iii. Score for Task C

b, Independent Variables:

i. Laboratory Skills (Planning, Performing and Reasoning).

ii. Gender (male/Female).

iii. Type of School (boys, girls or mixed).

Significance of the Study

There is numerous evidence to support the fact that science instruction in senior high schools in Ghana, places much emphasis on the acquisition of knowledge and factual information; to the neglect of the development of hands-on, performance and laboratory skills in physics students. The chief examiner's reports of the senior secondary schools certificate examinations on physics recommend that students should be exposed to practical exercises that would further explain the principles of physics, as it looks as if there is too much theoretical teaching in the schools (WAEC, 2002, 2005, 2006).

According to Tamir (as cited in Ossei-Anto, 1996), students who are given the opportunity to practise more and more laboratory activities eventually score higher on achievement tests.

It is therefore hoped that as students are exposed to more and more laboratory assessments they will begin to see assessment as a teaching and learning tool. Again, as students are given the opportunity to engage in performance assessment, it will enable them to assess themselves to know their strengths and weaknesses. This in turn will enable the students to be more aware and conscious of natural phenomena and problems in their environment; and to develop and improve on their investigative attitudes and procedures to tackling these problems and phenomena. It is further believed that as physics students are given the opportunity to practise more laboratory activities, they will be motivated to see and appreciate science as a practical discipline that demands students and practitioners to place more emphasis on hands-on tasks.

Delimitations

The study was delimited to SHS 3 physics students in Sekondi-Takoradi metropolis. It was hoped that physics students at SHS 3 have acquired at least minimum competency to enable them respond to the tasks. The study was also delimited to laboratory skills of planning, performing and reasoning in machines and refraction of light through glass prism. Other approaches and methods of doing science laboratory work were not employed in this research.

Limitations

There were nine SHSs in Sekondi-Takoradi metropolis that offer physics as a subject to students. The study should have included all SHS 3 physics students in the nine SHSs in Sekondi-Takoradi metropolis. Due to the enormity of work in carrying out laboratory activities, students in seven schools were used for the study. Sekondi-Takoradi metropolis is an urban centre and findings from this study may not be applicable to students and schools in rural areas of Western Region, and students and schools in other metropolises.

This research was conducted using physics students in SHS 3, as a result of that findings from this study may not be applicable to students in SHS 1 and 2.

Organization of Rest of the Chapters

The rest of the research report consists of four chapters. Chapter Two is on the relevant literature reviewed. The details of the methodology adopted for this study is discussed in Chapter Three. Subtopics of this chapter include: design of the research, population, sample and sampling procedure and instruments (tasks) used for the study. The chapter also gives a brief report of the pilot test that was conducted prior to the main research as well data collection procedures and data analysis. Chapter Four presents the results, discussion of the results, and findings of the research. The results and discussion seek to provide answers to the research questions that guided the conduct of this study. Chapter Five, which is the last chapter of the report, gives a summary of the results and research findings, conclusions, recommendations and suggestions for further study.

CHAPTER TWO

REVIEW OF RELATED LITERATURE

Overview

This chapter gives account of the relevant literature reviewed in support of the research. The literature reviewed has been systematically organised under the following sub headings: measurement, test, assessment and evaluation; validity and reliability of tests; criticisms of paper-and-pencil tests; new trends in assessment practices; science performance assessment; validity and reliability of science performance assessment; science laboratory assessment; gender differences in science achievement; types of schools differences in science achievement; and summary of literature reviewed.

Measurement, Test, Assessment and Evaluation

Various definitions, view points and characteristics of measurement, test, assessment and evaluation, as well as how each of these concepts relate to one another have been stated by researchers, psychologists and educational practitioners. This section of the literature reviewed examines some of these definitions, view points, characteristics, relationships, similarities and contrasts among them.

Nitko (2004) defines measurement as the process for assigning numbers or scores to a specified attribute of a person in such a way that the numbers describe the extent to which the person possesses the attribute. Ebel and

Frisbie (1991), on their part state that measurement is the process of assigning numbers to an individual's characteristics or trait according to specified rules. The measured trait can be a student's knowledge, laboratory skills, comprehension or performance. They are of the view that measurements are essential for describing the quantity of certain abilities and skills an individual possesses. As such, measurements are useful for the evaluation process. Measurement by whatever means it may be accomplished, be it a carefully constructed standardized test or a rating scale designed to measure laboratory skills and students' performance, is a basic part of the evaluation process. Measurement must be seen in terms of human values and goals. An individual who has more of an educational outcome must behave differently from a person who has less of the educational outcome (Ebel & Frisbie, 1991).

Cunningham (1986) also stated that the focus of measurement in the social sciences is to provide numerical index that can be used to quantify traits which allow the study of human behaviour based on mathematical models. This provides an approach that permits an objective discussion of an attribute to be made as it enables people to talk about exact amount rather than vague approximations. Measurement of an achievement is an indirect measurement (Micheels & Karnes, 1950). The effects or outcomes of the achievement are what is measured and not the achievement itself. They maintain that achievement can be measured if adequate instruments are developed and properly used. Thus a teacher has the responsibility to use measuring instruments and devices that are as accurate as can be obtained at that time. In measuring human achievement the procedure is to determine exactly what to measure and how to measure it. Thus, Micheels and Karnes (1950) contend that measurement deals with constructing instruments for best determining relatively what a student has learned, how well he has achieved, how much he has developed, and/or to what extent he has changed. Measurement of achievement takes place in the classroom, laboratory, at the workplace and many other places. Micheels and Karnes (1950), however, contrasted that educational measurements are crude as compared to physical measurements; since human behaviour is a complex property.

A Student's achievement may be measured by a test by counting the number of test items the student answers correctly, and the same rule is used to measure the achievement of each student in the class (Ebel & Frisbie, 1991). Micheels and Karnes (1950), also opine that achievement in the classroom is often stated in terms of so many points on a test that is supposed to measure the achievement that has taken place. Nitko (2004), on his part, defines a test as an instrument or systematic procedure for observing and describing one or more characteristics of a student using numerical scale or classification scheme. Noll (1965) also states that testing involves the use of some specific instrument or set of instruments to determine a certain quality or trait or a series of such qualities. The test is usually scored by adding together the points a student earned on each question. Testing is a measurement technique and all tests that measure relevant learning outcomes are the most useful tools for evaluation process. Testing and awarding certificates/ diplomas/degrees are recognised as a legitimate basis for decision making (Cunningham, 1986). Testing makes the extent of the differences in abilities of individuals or students more obvious. Ability testing has allowed for admitting students to programmes with restricted enrolments such as medicine and law.

In the process bright students but from the lower class have risen on the socioeconomic ladder. Tests are a necessary part of any educational system. They form the basis for assignment of grades. Schools have the responsibility to award certificates to students. The schools or institutions must ensure that students have mastered certain basic knowledge and skills before they are awarded with certificates.

Nunnally and Ator (1972) also contend that tests have proved useful in making numerous decisions. If the results from tests are used with care and with an understanding of what they do and do not measure, they can be a real help for guidance counsellors, school administrators and the classroom teacher. On the other hand if the results of tests are improperly interpreted it is quite easy to make incorrect and even harmful decisions on the basis of scores from tests (Lindvall, 1967). Tests are valuable only if they are properly constructed and if the results are correctly evaluated and used to improve teaching methods. Since teachers use tests more than any other means of assessing students' progress, they must have a clear understanding of the process of writing test items and they must be proficient in using the results thereof. Karmel (1966) quotes that "the basic reason the school staff administers tests is to promote the education of a child according to his or her unique abilities" (p.1). Proper administration of teacher-made examinations is a must if examinations are to be valid and reliable measures of student progress. Tests provide for student motivation by rewarding those who are hard working and prepared, and negative consequences for those who are not. Furthermore, tests can cause a student to study more as the frequency of an individuals' behaviour is increased by reinforcement. Tests also offer useful diagnostic information to the instructor about the strengths and weaknesses of a student; provide valuable basis for the modification of the instructional programme; and informs the instructor about the effectiveness of his instruction. Tests and other instruments of evaluation are classified in several ways. However, general classification is based on the uses or the kinds of abilities that are measured, and these include:

1. Achievement tests.
2. Scholastic-aptitude (intelligence) tests.
3. Special aptitude tests.
4. Interests inventories.
5. Character or personality instruments.

Some tests are teacher made while others are standardized of aptitude or achievement. These tests require the students to select an answer (e.g., true-false, multiple choice, or matching) or to supply an answer (e.g., short answer or essay).

Assessment, on the other hand, is the process for obtaining information that is used for making decisions about students, curricula and programmes, and educational policy (Nitko, 2004). Lindvall (1967) quotes that "the teacher's major 'business' is to produce changes in students, and he can determine his degree of success only by making regular assessment of what his students have learned" (p.5). Decisions about students include selecting them for educational opportunities, and credentialing and certifying their competence. Assessing a student's competence means collecting information to help decide the degree to which the student has achieved the learning targets. A large number of assessment techniques may be used to collect this information. These include lab work, projects, and paper-and-pencil tests. Assessing students is a very important part of teaching. Good decisions about teaching need to be made. High quality information is necessary if decisions are to be accurate, valid and fair to students. It is only through high-quality assessments that high quality information is obtained (Gronlund, 1998).

Undoubtedly, the most widely used procedure for assessing student achievement is the teacher-made test. Teachers should be able to make effective use of variety of procedures and devices for determining students' achievement, but they should also learn how to avoid misusing tests and should recognize that many readily available instruments for measuring achievements are inappropriate. Properly designed and appropriately used assessment procedures can contribute to more effective instruction and greater student learning. There is a close relationship between instruction and assessment. Both require clear specifications of the learning outcomes to be achieved by students. Thus, instructional planning should be broadened to include assessment planning (Gronlund, 1998). Carefully planned assessment procedures from the beginning of instruction to the end can improve the effectiveness of many decisions by providing more objective information on which judgment can be based. Assessment can be used to ascertain the extent to which students possess the skills and abilities needed to begin instruction; enables the learning tasks on which students are making satisfactory progress to be known; and those on which they are not making satisfactory progress to be determined. Moreover, assessment allows students who are encountering learning difficulties to be identified so as to remedy the situation; as well as allows students who have mastered the learning tasks to the extent that they can proceed to the next course or unit of instruction to be identified (Gronlund, 1998).

Assessment enables the grade to be assigned each student to be determined; allows students who should be awarded certificates to be known; and motivates students by providing them with short term goals, clarifying the types of learning tasks to be learned, and providing feedback concerning their learning progress. Assessment can also aid student self-assessment by providing them with information which is a more objective basis for assessing their own strengths and weaknesses (Gronlund, 1998). Furthermore, assessment can help in evaluating instructional effectiveness, by determining the extent to which instructional objectives were realistic, whether the instructional methods and materials were appropriate, and how well the learning experiences were sequenced. “When the majority of the students do poorly on an assessment, it may be the fault of the students but the difficulty is more likely to be found in the instruction.” (Gronlund, 1998, p. 11). The various modes of assessment of educational outcomes are:

1. Formal versus informal.
2. Final versus continuous.
3. Formative versus summative.
4. Internal versus external.
5. Process versus product.
6. Convergent versus divergent.
7. Competitive versus non-competitive.

Evaluation, which is one of the accountability processes in any educational system, is defined by Nitko (2004) as the process of making valued judgments about the worth of a student's product or performance. Micheels and Karnes (1950) also characterize evaluation as a more comprehensive judgment and subjective appraisal of a student's achievement. It involves the use of strict objective techniques. Evaluations make use of all types of measuring instruments such as tests, rating scales, interviews together with exercise of judgment. Evaluations are the basis for decisions about what course of action should be followed. Gronlund (1971) also contend that evaluation is not merely a collection of techniques but a continuous process that underlies all good teaching and learning; and is inevitable in a classroom instruction, however simple or complex the consideration involved. Evaluations teachers make can have enormous influence on the lives of their students and should not be taken lightly or casually made (Gronlund, 1971). Gronlund (1971) further asserts that teachers have many varieties of sources and methods of collecting information about their students but the use of several methods and sources in combination is more common. He continued to say that the effectiveness of directing students' learning is based on the accuracy of the judgment made on them; which involves the understanding and use of principles and procedures of evaluation to make wise and intelligent decisions in guiding students' progress toward beneficial instructional and educational objectives.

Evaluative judgments are critical and necessary in guiding students toward the achievement of instructional goals, diagnosing students' learning difficulties, determining students' readiness for learning new experiences, placing students in groups for special activities, assessing students with problems of adjusting and preparing reports of students' progress for parents. It is important to clearly define the learning outcomes to be evaluated. These learning outcomes can be knowledge, understanding, thinking skills, performance skills, laboratory skills or attitudes. The educational outcomes are the results of learning stated in terms of changes in student behaviour. Lindvall (1967) also states that evaluation of students' achievement is the process of determining how well students have attained specific instructional objectives. It therefore consists of a variety of techniques and procedures, including observation of students' performance, grading classroom exercises and homework assignments, and use of several different types of achievement tests. In order to evaluate what a student has learned in a given unit or an entire course, we must know the specific ways in which he should be able to exhibit that achievement (Lindvall, 1967). There should be statements of objectives describing what he should have learned in specific and limited units of instruction. Statements of instructional goals are sometimes referred to as 'terminal behaviours' because they describe the behaviour that a student should be able to demonstrate at the termination of some period of instruction. The first step in evaluation is to translate the goals or objectives that have served as guides for teaching into a detailed listing of specific behavioural objectives that can direct evaluation efforts.

Ebel and Frisbie (1991) on their part assert that the reason for evaluation is to make a judgment about the quality or worth of an educational programme or student achievements. Evaluation seeks to describe the level of achievement, or performance of an educational programme or student. The evaluator needs to decide the kind of information needed, how to collect the information and how to synthesize it to support the value judgment. The evaluation of learning takes place in instructional context. The instructional process and the role of evaluation in it must both be understood as the background to educational measurement (Ebel & Frisbie, 1991). Schwart, Tiedeman and Wallace (1962) also characterize evaluation as the process of making judgments and coming to decisions about the value of an experience. The evaluation of the experience involves careful judgment about the adequacy or effectiveness of the experience as measured in the light of the experience set for it. Schwart *et al*. (1962) proceeded to define evaluation in education as the process of judging the effectiveness or worth of an educational experience as measured against instructional objectives. Evaluation that is based on philosophical and psychological sound objectives, and focused on the best measurements that can be obtained, is critical to securing effectiveness in the total educative process. A wide variety of evaluation techniques exist; it is the responsibility of the teacher to learn these techniques, their proper uses and limitations, and how they can be best employed to help students in their learning experiences (Schwart *et al*., 1962). The focus of teaching and evaluation in education is upon the goals, ambitions, and hopes of human beings. These goals spell out those behaviours and values that society feels will contribute most to the individual and to society.

Evaluation must be based on comprehensive and continuous measurement of all phases of individual development and toward all objectives of school. It is the means by which objective, valid, reliable and usable accounting is made of the progress of a classroom group, as the students grow academically, socially, emotionally, physically and spiritually. Evaluation must go on continuously if all the changes in the student are to be fully appraised. It must be comprehensive and encompass the entire range of the student’s activities and experiences, including the curricular, the co curricular and the nonschool. The school's appraisal and records should be focused on the individual student; they should be concerned both with his present status in terms of his capacities and achievements and with the relation of his status to expected growth patterns.

Formative evaluation is done to determine whether learning is taking place as planned. Summative evaluation, on the other hand is conducted at the end of an instructional segment to determine if learning is sufficiently complete in order to move to the next segment of instruction. Formative evaluation in the classroom provides feedback to the teacher and student which provide an opportunity for the teacher to modify instructional methods or materials to facilitate learning when things are not going well (Ebel & Frisbie, 1991).

Formative evaluation requires collection of detailed information on a regular basis. Formative evaluation usually employs teacher observation, classroom oral questioning, homework assignments, classroom exercises, quizzes and in-formal inventories. Summative evaluation demands the collection of information on a broader scope of content and skills. It normally uses classroom tests or final examinations as instruments to gather information.

From this section, it is amply clear that for instruction to be effective and efficient; and to ensure that all important learning outcomes are being taught in the classroom, there is the need to obtain credible information to arrive at these conclusions. Moreover, in order to make many educational decisions, the decisions must be based on sound, accurate and valid information and judgment. These decisions call for assessments and evaluation of students and instructional programs. The assessments in turn call for the use of tests and other techniques to measure the achievements of students. The evaluation requires the use of sound philosophical and psychological principles and procedures. Thus, it is obvious that measurement, test, assessment and evaluation are closely related and interwoven educational and psychological concepts; and are all necessary for a complete appraisal of the student, instructional and educational programmes.

Validity and Reliability of Tests

Despite the vast differences between format and construction of various evaluation devices, there are certain common standards that should be met by any measuring instrument. Each measuring instrument should be chosen carefully in terms of the criteria of: (a) validity, (b) reliability, (c) objectivity, (d) efficiency and (e) usefulness. To the extent that any of a given measurement technique satisfies these criteria are we justified in using it with students (Schwart *et al*., 1962). Nunnally and Ator (1972) also state that how well a particular test helps in making particular decisions is an indication of how valid it is. In order to use standardized tests, a teacher should be familiar with the recommended steps to be followed in constructing a standardized test and should understand the statistical procedures used in deriving scores and in assessing the reliability and validity of the instrument. Lindvall (1967) is of the view that, all informed predictions of future performance are based upon some knowledge of relevant past performance, school grade or whatever is appropriate. How well the predictions will be validated by later performances depends upon the reliability and appropriateness of the information used and the skill and wisdom with which it is interpreted. Whether to use tests, or other kinds of information, or both in making a particular decision depends upon the empirical evidence concerning comparative validity and factors such as cost (Lindvall, 1967).

Any procedure for obtaining information about students is valid to the extent that it actually provides the desired information. In the evaluation of students' achievement, the information desired is the degree to which students have achieved the specified instructional objectives. Thus the validity of any procedure for assessing students' achievement is the extent to which it provides as to whether the students have exhibited the behaviours specified in the instructional objectives. Validity is the most important quality to be sought in any evaluation process. Thus, if the specific objectives state that the student should be able to perform or demonstrate certain skills, evaluation can be made through observation of student performance. The student must be given the opportunity to display the ability described in the specific objectives and he is then assessed the extent to which he has done this (Lindvall, 1967). It must also be noted that no instrument or procedure is valid on its own only. Thus an instrument which is valid in one situation for a given purpose may be invalid in another situation for a different purpose. Nunnally and Ator (1972) went on further to state that a test is valid if it serves its functions well. It is valid only for some specific functions with specific group under specific conditions. Schwart *et al*. (1962) contend that, "the most important characteristics of any appraisal technique is validity – the extent to which the technique actually measures what it is supposed to measure" (p.75).

Four types of validity have been distinguished, namely: content validity, concurrent validity, predictive validity and construct validity. Content validity is concerned with the sampling of a specified universe of content. Concurrent validity is concerned with the relation of test scores to an accepted contemporary criterion of performance on the variable which the test is intended to measure. Predictive validity is concerned with the relation of test scores to measures on a criterion based on performance at some later date; while construct validity is the degree to which one can infer certain constructs in a psychological theory from the test scores (Nitko, 2004). Validity may be determined by curricular approaches or by statistical analysis. The content validity of a measuring instrument can be established by comparison with objectives of instruction, by comparison with expert opinion and by comparison with textbook and source materials. Consensus by experts and textbook writers concerning the important objectives and content in a particular area adequately define for teachers what these objectives should be (Schwart *et al*., 1962). To a large extent this is a sound assumption, for these people have studied the field carefully and should have a good idea of what are valid objectives. Content validity of a measuring instrument can be determined by a direct comparison of the actual behaviours involved in the stated objectives of instruction with the behaviours needed for success on the test.

According to Payne and Mcmorris (1967), validity information indicates the degree to which the test is capable of achieving certain aims. Payne and Mcmorris (1967) went on further to state that "tests are used for several types of judgment, and for each type of judgment a different type of investigation is required to establish validity" (p.77).

A data gathering procedure is reliable to the extent that it will produce consistent results in assessing the same thing (Lindvall, 1967). Reliability is important in any procedure for obtaining information about students. Reliability is a matter of degree as no procedure is absolutely reliable. Some evaluation procedures are, however, more reliable than others. Schwart *et al*. (1962) point out that the reliability of a measuring instrument refers to the consistency with which it measures, or the extent to which it can be trusted to give us the same or similar scores or description of behaviours at different times. Reliability is usually represented in terms of correlation coefficients. Test reliability is established by: test retest method, comparable (parallel) forms method, split-halves method as well as the inter-rater method. The inter-rater reliability method involves two or more raters (judges) scoring the same set of test or assessment instruments independently. Payne and Mcmorris (1967) characterize reliability as the accuracy (consistency and stability) of a measurement by a test. Any direct measurement of such consistency calls for a comparison between at least two measurements. The two measurements may be obtained by retesting an individual with the identical test or by using two raters to independently score the same test. The different methods of determining reliability take account of different errors. Reliability therefore refers to many types of evidence, each of which describes the agreement or consistency to be expected among similar observations. There are various components that may contribute to inconsistency among observations. These include response variation by the subject or student, due to changes in physiological efficiency or in psychological factors such as motivation, effort or mood. Variations in test-content or test-situation can be another source of inconsistency. So can variations in the process of observation or scoring can also be a source of inconsistency.

Criticisms of Paper-and-pencil Tests

Researchers such as Ossei-Anto (1996), Oloruntegbe (1999) and Al-Sadaawi (2007) have criticized traditional paper-and-pencil tests such as multiple-choice, fill-in-the-blank, and true or false. It is contended that traditional techniques do not prove effective for the expanded concept of learning that requires students to demonstrate higher-level thinking skills. Al-Sadaawi (2007) emphasized that traditional tests evaluate a limited number of cognitive functions and skills related only to memory, and students' ability to recall material learned out of context. Moreover, for purposes of accountability, teachers tend to tailor their instructions to students in imitation of multiple-choice questions, thus encouraging students to focus only on the options before them. Therefore "teaching to the task" has become a common practice in schools, narrowing students' potential to low-level skills, and distorting the curricula (Al-Sadaawi, 2007). Science educators claim that traditional tests cannot adequately evaluate students' ability to design and undertake experiments or assess their understanding of scientific concepts.

It is reasonable to question whether paper-and-pencil objective tests can measure the possession of many of the skills deemed essential for safe and trustworthy practice (Ebel & Frisbie, 1991). Although facts and concepts are fundamental in science education, knowledge of methods, procedures and analysis of skills that provide context is equally important (Slater, 1997). According to Slater (1997) student growth in these latter facets proves difficult to evaluate, particularly with conventional multiple-choice examinations. Historically, researchers such as Baxter, Shavelson, Goldman and Pine (1992); and Ruiz-Primo and Shavelson (1996) have shown the limitations of measuring complex cognition and inquiry process via multiple-choice and constructed-response paper-and-pencil tests. Clarke (2009) state that these tests also demonstrate limited sensitivity to discrepancies between inquiry and non-inquiry based instruction. He further went on to report that on paper-and-pencil tests, inquiry is not effectively measured. Some of inquiry abilities such as communication, and alternative explanations are tested barely or not at all. If these are valued standards, it would seem that some items should test them. Thus, despite the increasing focus of world-wide science standards on inquiry, paper-and-pencil assessments continue to demonstrate misalignment and validity issues in the measurement of this domain.

Oloruntegbe (1999) also contends that, of the science learning outcomes – formulation of concepts, development of skills and appropriate scientific attitudes, cognitive area seems to attract more attention in school teaching, learning and assessment. Teachers seldom teach and assess skills and attitudes. Traditional multiple-choice achievement tests in science have been criticized in several ways. Despite their efficiency (economical to develop, administer, and score) they do not measure some aspects of knowledge that are valued in science education – the science processes, especially the ability to formulate a problem or carry out an investigation. Hence, multiple-choice tests are limited in capturing students' thinking and problem-solving skills (Ruiz-Primo & Shavelson, 1996). Multiple-choice tests do not look like science conducted in the laboratory or the field, and consequently may provide limited information about what students know and can do in science.

New Trends in Assessment Practices

"The increased interest in recent years in giving more curricular emphasis to higher order thinking skills has raised questions about how to measure the achievement of these skills" (Ebel & Frisbie, 1991, p.6). Currently, many of the innovative curricular teaching and assessment strategies are devised and constructed such that they are directed towards the building of problem-solving abilities. Contemporary development in cognitive and constructivist theory are of the perspective that, meaningful learning is reflective, constructive, and self-regulated; learners are seen not as merely receivers of information but as creators of their own unique knowledge structures. Learners can therefore achieve a more meaningful goal in which acquisition of knowledge, skills, and attitudes enable them to act effectively, expertly, and professionally under a teacher's guiding role. Assessment theory and practice are evolving to reflect these complexities and moving away from a narrow focus on simple tests and scoring that previously dominated teaching (Al-Sadaawi, 2007). Thus, achievement needs to be considered as a qualitative change in a person's conceptions, not simply the amount of knowledge that a person possesses. Al-Sadaawi (2007) states that, to assess students on scientific reasoning and understanding rather than simply measuring discrete knowledge, critical assessment methods have been developed, with strong preference for performance-based assessment (Al-Sadaawi, 2007).

The acknowledged weaknesses of conventional paper-and-pencil assessments have led to the recent development of alternative testing strategies. Already validated and used in many schools, one of the most commonly used of these is called performance assessment (Slater, 1997). Shavelson, Ruiz-

Primo, Li and Ayala (2003) state that, spurred by calls for education reform in the wake of international comparisons of student performance, developments in cognitive science, and augmentation of international achievement test formats, a myriad of new approaches to assess students' learning have emerged over the past decades. Some examples of these new assessments are performance assessments, concept maps, and predict-observe-explain demon-strations. Baxter *et al*. (1992) state that, three forces have spurred the demand and subsequent search for alternative measures of science achievement: (a) dissatisfaction with current multiple-choice techno-logy, (b) advances in research on cognition and instruction, and (c) reform of science curricula. Consistent with developments in cognitive research and curricula reform, and in response to critics of multiple-choice tests, studies were undertaken to develop and evaluate alternative assessments of science achievements (Baxter *et al*., 1992). The reason was that one way to assess what students know about science is to watch them do science.

According to Clarke (2009) attempts have been made to address the limitations of paper-and-pencil tests by designing hands-on and virtual performance assessments. Over the past three decades, researchers have made significant advances in methods of assessment design. Frameworks such as the assessment triangle and evidence-centred-design provide rigorous procedures for linking theories of learning to demonstrations to interpretations. To measure science learning outcomes at skills level, assessment has been modi-fied and restructured not only in form and context, but in vocabularies and nomenclatures. Thus, there have emerged different forms like alternative, authentic and performance assessments, which are labelled the most suitable for assessing science process skills demonstrated by students during science activities (Oloruntegbe, 1999). Oloruntegbe (1999) further asserts that assessment is changing for many reasons. Changes in the skills and knowledge needed for success; in understanding of how student learn; and in relationship between assessment and instruction necessitate changes in assessment strategies. The assessment strategies should be tied to design, content, new outcomes and purposes. Educators, policy makers and parents are beginning to recognize that minimums and basics are no longer sufficient, and are calling for a closer match between the skills that students learn in schools and the skills they will need upon leaving school. Attempts to assess skills have led to changing faces, nomenclature and movements in assessment. Several terminologies such as alternative assessment, authentic assessment, project based assessment and many others have been coined (Oloruntegbe, 1999).

The goals of these assessments are to provide quantitative data on students' performance particularly in science. Ruiz-Primo and Shavelson (1996) are of the opinion that performance assessments have caught public attention in the past years as a complement to multiple-choice tests. Traditional, low-inference testing, based on the assumption that knowledge could be de-contextualised is replaced by contextual assessment methodologies in science education, such as performance assessment, not on account of direct criticism, but rather on the change from behavioural to cognitive psychology, developments in the philosophy of science, and the rise of constructivism (Ellis, Jablonski, Levy & Mansfield, 2008). Ellis *et al*. (2008) contend that the global economy of the 21^{st} century requires a different set of knowledge and skills than workers needed in the past. Today students must develop the ability to access, inquire into, think critically about, and carefully analyze information. Ellis *et al*. (2008) went on further to say that performance assessments address these societal demands by providing complementary information to traditional test data; by measuring what students are able to do with the content they have learned, and not only whether they have learned the curriculum content.

Science Performance Assessments

Researchers have indentified performance-based assessment as the focus for education reforms in assessment, curriculum and instruction. The proponets of performance assessments argue that a performance-based assessment methodology provides students with meaningful paths to demonstrate their knowledge. The technique also improves students' skills by bringing into play complex functions of cognitive processing that require higher-level of thinking for problem-solving, or the development of options when an individual is confronted with new situation. Al-Sadaawi (2007) cites Baker as saying that since performance-based assessment occurs over a period of time, it provides an opportunity for students to individually achieve the highest level of learning. Performance-based assessment is authentic assessment because it involves the performance of tasks that are valued in their own right, it is situated in a real world context, and it can mirror actual tasks implemented by professionals. For students, performance assessment provides a realistic approach to science, thereby reinforcing the inquiry skills of science and assessing self-progress. For teachers, the methodology provides timely information on the learning needs of their students, and thus the teaching methods they employ. Performance assessment is therefore an appropriate strategy for assessing students' concepts and skills in science, and it prepares students for a productive future within a technologically complex world (Al-Sadaawi, 2007). Positive effects in the quality of students' learning and attitudes have been shown from empirical studies of impact of performance-based assessments. Baxter and Glaser (as cited in Al-Sadaawi, 2007) found that performance-based assessment not only supports the development of thinking and reasoning skills in the classroom, but also provides teachers with feedback that can be used to improve the classroom environment. Again, it has been found that performance-based assessment is a valid, equitable measurement of students' progress.

The International Association for the Evaluation of Educational Achievement [IEA] (1995) on its part characterize performance assessment as the use of integrated, practical tasks, involving instruments and equipment, as a means of assessing students' content and procedural knowledge, as well as their ability to use that knowledge in reasoning and solving problems. The assessment task may be as simple as the routine use of a piece of equipment or as complex as an investigation combining manipulative and procedural skills and requiring higher-order thinking and communication. Proponents of performance assessment argue that the practical nature of the tasks utilized in this mode of assessment permits a richer and deeper understanding of some aspect of student knowledge and understanding than is possible with written tests alone ([IEA], 1995). According to [IEA] (1995), these aspects include skills like weighing and measuring, the use of experimental or mathematical procedures, designing and implementing approaches to solve problems or investigate phenomena, and synthesizing knowledge, application, and personal experience into an interpretation of data. A well-designed performance assessment task, with appropriate scoring rubrics, can elicit a rich variety of student performances, and offers the possibility of deeper understanding of cognitive processes and problem-solving strategies. Detailed study of students' performance on practical tasks in life-like assessment situations, offer greater potential for understanding their achievement than paper-and-pencil tests alone. Performance assessments, used in concert with more traditional forms of assessment, are designed to provide a more complete picture of student achievement (Slater, 1997).

Slater (1997) also states that performance assessments are designed to judge students' abilities to use specific knowledge and skills. Most performance assessments require the students to manipulate equipment to solve problems or make an analysis. Slater (1997) further asserts that, rich performance assessments reveal a variety of problem-solving approaches, thus providing insight into a student's level of conceptual and procedural knowledge. Performance assessment strategies are composed of three distinct parts: a performance task; a format in which the student responds; and a predetermined scoring system (Slater, 1997). Tasks are assignments designed to assess a student's ability to manipulate equipment (laboratory equipment, computers etc.) for a given purpose. Students can either complete the task in front of a panel of judges or use a written response sheet. The student is then scored by comparing the performance against a set of criteria. When used with students with highly varying abilities, performance tasks can take maximum advantage of judging student abilities by tasks with multiple correct solutions (Slater, 1997). According to Slater (1997), students are graded on the process of problem solving using rating scales based on explicit standards. Performance assessment strategies are best utilized in concert with other forms of assessment, both factual and procedural knowledge are important components of a complete science education. The purpose of performance assessment in science is to evaluate the actual process of doing science; and examines students' actual application of knowledge to solve problems. In some cases, the solution of the problem may imply the application of a specific procedure learned in class; in others, a combination of procedures; still in others it may require a thoughtful adaptation of students' knowledge. The assessment of students' knowledge and skills focuses on the performance (process) and the result (product) [Slater, 1997].

Performance assessments are typically inappropriate for measuring students' knowledge of facts. They may be used for diagnostic purposes – to know how students solve certain types of problems; how they control variables; how they use instruments; and how they evaluate findings. Performance assessment may also be used for instructional purposes. Performance tasks simulate the authentic tasks of scientists. If the assessment task is used in such a way that the student would normally not know it is an assessment activity, it is called an embedded task (Slater, 1997). Performance assessment may also be used for monitoring purposes. The goal of science performance assessment is to judge the level of competency students have achieved in doing science. Accordingly, the assessment strategies are best used to monitor student process skills and problem solving approaches. Slater (1997) believes that the most effective performance assessments are authentic tasks that are open-ended with multiple-correct solution paths. In science, some tasks require systematic procedures that do not yield multiple-entry points or exit points. In this case, a checklist system can be appropriately used by an observer or a highly-structured student-answer sheet in which each aspect of the procedure and result is described in detail. Highly-structured assessment tasks provide students with step-by-step instructions to follow. In contrast, less structured assessment tasks give students more opportunity to make judgments in determining the procedures needed to solve the problem.

Performance assessment can be administered individually, in pairs, or collaborative groups. If it is administered in pairs or groups, students should write in their own answer/response sheet. When students solve the problem in pairs or groups, the goal and composition of the group will affect the student's individual performance (Slater, 1997). Predetermined criteria must be used to evaluate students' performance. Students' should not be scored/graded against their peers, but based on the criteria predefined. It is always useful to try to find in students' performance patterns of appropriate and inappropriate responses. This helps focus on problems observed. An authentic assessment is one in which students are required to address problem grounded in real-life contexts. Slater (1997) cites Shavelson, Baxter, and Pine; and Wiggins as stating that authentic tasks are typically complex, somewhat ill-defined, engaging problems that require students to apply, synthesize, and evaluate various problem solving approaches. Authentic tasks are clearly different in nature, form, and length from multiple-choice questions that can usually be responded to in a matter of seconds. In the evaluation of performance task, the process of performing the task is emphasized more than the final product itself.

Clarke (2009) is also of the view that performance assessments complement rather than replace existing standardized measures by assessing skills not possible via paper-and-pencil and multiple-choice tests. Science performance assessments pose a problem and put students in a mini-laboratory to solve it, evaluating the solution as to its scientific defensibility. Performance assessments are interpreted as capturing a student's scientific reasoning and procedural skills, and are believed to require the application of scientific knowledge and reasoning in simulated real-world situations as well as in situations similar to what scientists do (Ayala, Shavelson & Ayala, 2001). Oloruntegbe (1999) advocates that on-the-spot assessment of skills complements and supplements the conventional paper-and-pencil tests. Alternative, authentic and performance assessments are used interchangeably to mean the same thing. According to Oloruntegbe (1999) the characterization and categorization of traditional tests and performance assessments fall into a continuum. Events at the two ends of the continuum range from selection of response to performing a task; from contrived to real life; recall or recognition to construction and application, teachers-structured to learners structured and indirect evidence to direct evidence. Within the continuum lies four categories; tests, product or project assessments, performance assessment and process skills assessment.

All these types are useful, however, each has limitations. Therefore maintaining balance becomes of utmost importance (Oloruntegbe, 1999). Oloruntegbe (1999) went on further to state that, authentic assessment is a form of assessment in which students are asked to perform real-world tasks that demonstrate meaningful application of essential knowledge and skills. The authentic assessments engage students in applying knowledge and skills in the same way they are used in the "real world" outside school. It is performance-based assessment that requires a student to go beyond basic recall and demonstrate significant, worthwhile knowledge and understanding through a product, performance or exhibition. Students appear to learn best when they see the importance for learning and when the learning environment is familiar to them. Authentic scenarios can provide this environment and relevance to students (Oloruntegbe, 1999). Performance assessment is a form of testing that requires students to perform a task rather than select an answer from a ready-made list. Advocates say that performance assessment may be a more valued indicator of what students know, and what they are able to do (knowledge and abilities) promotes active learning and curricular-based testing. Alternative assessment includes any assessment in which students create response to a question. Oloruntegbe (1999) believes that the worth of a school graduate is a product of good teaching and assessing. Superficial effort in this way will only breed deception and defects – students parading high grades from theory and the conventional practical examination without the corresponding skill to go with them.

Performance assessments are assumed to tap higher-order thinking processes and be more directly related to what students do in the classroom and what scientists actually do – observe, hypothesize, record, infer, and generalize (Ruiz-Primo & Shavelson, 1996). Again, Ruiz-Primo and Shavelson (1996) conceive of science performance assessment as a combination of: (a) a task that poses a meaningful problem and whose solution requires the use of concrete materials that react to the actions taken by the student; (b) a format for the student's response; and (c) a scoring system that involves judging not only the right answer, but also the reasonableness of the procedure used to carry out the task. Different measurement methods can be used to collect information on students' performances. These include direct observation, notebooks, computer simulation etc. (Ruiz-Primo & Shavelson, 1996). With direct observation, an observer check-lists a student's performance and response as the student proceeds with the investigation. With notebooks, students record their procedures and conclusions as they proceed on the investigation.

Ellis, *et al*. (2008) state that science is a fertile ground for engaging students in inquiry and critical thinking; thus, it naturally lends itself to performance assessment. Ellis *et al*. (2008) further state that performance assessment is generally recognized as a form of testing that requires students to perform a task rather than select an answer from a ready-made list. The task is then scored by experienced raters, such as teachers or trained staff, who judge the quality of students' work based on agreed-upon set of criteria. Because performance assessments require students to actively engage with a task in order to show their understanding and proficiency around complex and realistic problems, they can be complicated and challenging both to develop and implement (Ellis *et al*., 2008). However, if designed properly, performance assessments can provide an indication of what students are able to do with their knowledge by requiring students to demonstrate what they know through various tasks, such as generating scientific hypothesis or conducting experiments. Ellis *et al*. (2008) state that the definition of performance assessment is comprised of two parts. First, the content to be assessed has to be the scientific inquiry or investigative process – make observations, raise questions, and formulate hypothesis; design and conduct scientific investigations; analyze and interpret results of scientific investigations; and communicate and apply the results of scientific investigations. Second, the assessment should require a student to display understanding of the scientific inquiry process via hands-on (either real or simulation) tasks.

Knowledge is an important outcome, the foundation of all learning that should be evaluated for its own sake. However, knowledge alone cannot be used as evidence of having acquired understandings, habits, attitudes and skills (Gronlund, 1971). The most effective way of ensuring that all important learning outcomes are being evaluated properly is to state the immediate objectives of instruction in a way to reflect clearly the ultimate objectives to be achieved, and then develop evaluation procedures best suited to each of them. Some educational objectives are unique to a particular course that other educational experiences have little or no direct contribution to their attainment. Learning outcomes such as understandings, laboratory skills, and performance skills are limited to educational objectives that can be derived from the specific course. Therefore the teacher should identify, select and clarify the learning outcomes for evaluation purposes. Educational objectives may also be classified in terms of the extent to which they are functional in the instructional programme. However, the stated educational objectives usually fail to contribute to the entire educational process because of inadequate attention to or improper choice of the evaluative (assessment) techniques used (Gronlund, 1971). Improved evaluative procedures must primarily contribute to improved learning. According to Ebel and Frisbie (1991), skills beyond the level of remembrance will often need to be demonstrated through performance tasks or product development in order to assess it adequately. Instruments meant to measure critical thinking and laboratory skills are not all perfect. If we expect students to learn critical thinking and laboratory skills and improve on inadequate skills they already have, then teachers need to continue to learn and to improve their measurement and evaluation practices (Schwart *et al*., 1962). To appraise the progress of the student on the basis of the various educational objectives and goals, evaluation cannot be restricted to the classroom. It must also take place in the extra class school activities, in the laboratory, on the field, in the employment and in the real world. According to Karmel (1966), schools administer many different tests because no one test can measure all the different facets of a student's ability, skills and interests.

Validity and Reliability of Performance Assessments

The international association for the evaluation of educational achievement [IEA] (I995), state that performance assessment aims to provide students with a testing environment which is more "true to life" and "authentic" than the traditional paper-and-pencil written test, and, by providing them with equipment and materials to manipulate in a realistic problem-solving situation, attempts to elicit performances or behaviours which will be a more valid indication of the students' understanding of concepts and potential performance in real life situations. Performance assessment has captured the attention of teachers, and policymakers for a variety of reasons. It reflects the current trend in many countries towards active, inquiry-oriented, hands-on teaching and learning; and it is seen as a means of assessment that is educationally valid, psychologically and developmentally appropriate, and congruent with “constructivist” pedagogies ([IEA], 1995). In most all cases, the intent of assessing learning is to go beyond ranking students on their performance to drawing inferences about what they know and are able to do with that knowledge. That is, assessments of learning are interpreted as providing information on cognitive activities (minds-on) as well as on performance (hands-on) (Shavelson, Ruiz-Primo, Li & Ayala, 2003). Shavelson *et al*. (2003) contend that assessments are intended to measure a construct. The validity of the assessment is guaranteed if the construct is clearly defined. Analysis of the “working” construct definition will describe the domain of learning and performances that the construct covers and the kinds of student responses (behaviours) to be produced to meet the demands of the construct.

Analysis of the construct should identify a range of tasks within the domain of knowledge and skills that could be presented to students; and the kinds of responses that would be expected (Shavelson *et al*., 2003). Definition of the working construct will also rule out other tasks and responses (performances) that should not be related to the construct. The definition of the construct therefore determines the tasks or situations, response demands, and scoring system that comprise a learning assessment. Again, to ensure the validity of performance assessment, logical evidence must be sought that this construct will be evoked by the assessment tasks; and empirical evidence that the construct, was, indeed evoked in a student’s behaviour (performance). Cronbach, (as cited in Shavelson *et al*., 2003) states that an assessment is a systematic procedure for eliciting, observing and describing behaviour, often with a numerical scale. The assessment is a physical manifestation of the working construct definition. It is one of many of possible manifestation of the construct in the form of an assessment that could have been produced (Shavelson *et al*., 2003). Once an assessment has been developed or selected for use, logically, its tasks and response demands are analysed to see whether if falls within the construct domain, and whether it is likely to elicit the expected behaviours or performances from a student. The task analysis involves reviewing the task and determining what kinds of thinking and skills the task might evoke in students.

According to Shavelson *et al*. (2003), this analysis posits cognitive activities that the task might evoke by examining the "opportunities and constraints" that the assessment task provides students to elicit their knowledge and skills. The validity and reliability of the performance assessment also involve collecting and summarizing students' behaviour in response to the assessment tasks. This empirical analysis focuses not only on observed and perhaps scored task performance, but also on cognitive activities and skills elicited by the task. The analysis provides evidence on a student's cognitive activities and skills that were evoked by the task as well as the student's level of performance. The analysis brings both to bear on the link between the assessment and the construct definition. Evidence from the logical and empirical analysis are put together and brought to bear on the validity of interpretations from an assessment to the construct it is intended to measure.

Clarke (2009) contends that performance assessment potentially provides greater construct validity for science inquiry over paper-and-pencil tests. The goal of an assessment is to provide valid inferences related to particular expectations for students. Science cannot be understood as content separated from the process that create that content; and therefore assessments must be developed that cover students' ability to do scientific inquiry and understandings about scientific inquiry within the domain of science. Clarke (2009) further contends that validity is a central issue in test construction. Validity is an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment. In order to provide evidence that performance assessment tasks demand students' ability to do inquiry, series of validity studies must be conducted to provide evidence on construct validity (Clarke, 2009). Determining the cognitive validity of performance assessments with respect to scientific reasoning they elicit is paramount; since these assessments are touted as tapping higher order thinking skills and mimicking what scientists do (Ayala *et al*., 2001). A performance assessment is content rich if it requires specific content knowledge to succeed; and it is process open if students in order to complete the assessment have to come up with their own procedures rather than follow a procedure.

Ruiz-Primo and Shavelson (1996) on their part assert that the validity of performance assessment depends on the task; the occasion for administering the task; and the method used to assess the performance. Results show that measurement methods seem to tap different aspects of science achievement. Each method may provide a different insight into what students know and can do. The technical qualities of performance assessments must therefore be more carefully examined before the scores are reported to students, parents, and policy makers. The utility of performance assessment refers to the extent to which performance scores provide useful information for monitoring instruction and student progress. Practicality of performance assessment refers to the extent to which performance assessments can be used in a classroom without excessive cost, effort, or disruption. Ruiz-Primo and Shavelson (1996) are therefore of the view that for science performance assessment to be useful for teaching, they need to be linked directly to instructional units; and have a well-designed scoring system that clearly reflects what students know and can do. Ellis *et al*. (2008) cite Messick as warning that two of the major threat to validity of performance assessments are: construct under-representation, where an assessment is too narrowly focused; and construct-irrelevant variances, where assessments tap knowledge and skills not relevant to the content around which the assessment was designed.

Studies have found that if the criteria for scoring performance assessment are clear, and that examples are available to show levels of competency, performance assessments are highly consistent across different evaluators - inter-rater reliability (Slater, 1997). Ruiz-Primo and Shavelson (1996) also state that research findings on rater sampling, or consistency of scores across raters, are positive. Raters can be trained to reliably evaluate student performance. However, there are indications that students perform inconsistently from one performance task to the next (inter-tasks-reliability). This suggests that students' grades will be most reliably determined from a number of performance assessments in concert with other forms of assessment. To assess the reliability of performance assessments, series of generalizability studies must be conducted. Clarke (2009) cites Shavelson and Webb as stating that generalizability theory (g-theory) is a statistical theory that allows decision makers to study the dependability of behavioural measures and procedures. It is a commonly used technique for making decisions and drawing conclusions about the dependability of performance assessments. Ruiz-Primo and Shavelson (1996) report that larger numbers of tasks are needed before generalizable measures of the achievement can be made. They went on further to say that performance assessments can be managed as efficiently as hands-on science instruction, and that scoring can be easy and quick to learn and use. Teachers who opt to use an activity-based curriculum in their classroom can administer performance assessments with no more difficulty than that associated with inquiry science.

Aspects of reliability that are particularly important to performance assessments include the reliability of the instrument itself, inter-rater reliability, where to cut-off scores, and how to deal with scores that fall near those cut-off points (Ellis *et al*., 2008). Generalizability of a performance task allows inferring beyond the task itself to a broader set of skills and abilities related to the performance assessment and, thus, are intricately related to the content representative and construct-related validity of the assessment. Because of the extensive time required for the typical performance task, there is a conflict in performance assessment between time-intensive depth of examination and the breadth of domain coverage needed for generalizability of construct interpretation. In other words, a central tension exists between the depth and breadth of knowledge and skills that are tapped in a specific performance task and the extent to which generalizations regarding those knowledge and skills can be made (Ellis *et al*., 2008).

Science Laboratory Assessments

The evaluation of laboratory performance provides an important example of alternative or authentic assessment (De Ture, Fraser, Giddings & Doran, 1995). Tamir (as cited in Anthony-Krueger, 2001), states that laboratory practical examinations which require manipulation of some materials and which involve direct experience of the examinee with the materials at hand, are classic examples of science performance assessment; and therefore recommends the use of practical laboratory examination as a means of assessing students' performances.

The laboratory has been central component of science instruction since the early 20^{th} century. It has been used to teach experimental methods and techniques that clarify and/or validate existing scientific principles and theories. Hofstein and Lunetta (2003) assert that the laboratory has been given a central and distinctive role in science education, and science educators have suggested that rich benefits in learning accrue from using laboratory activities. Most of the assessment of students' performance in the science laboratory continues to be confined to conventional, usually objective, paper-and-pencil measures. More sensitive measures of students' understandings of laboratory methodologies, the hypotheses and questions they generate from the lab experiences, and practical skills they exhibit have all too often been neglected (Hofstein & Lunetta, 2003). Eubanks (1997) also states that for almost seventy years multiple-choice has been very nearly synonymous with the idea of national chemistry examinations. As a consequence, many high school and college chemistry teachers use multiple-choice exams without seriously considering their strengths and weaknesses. According to Hofstein and

Lunetta (2003), methodologies for research and assessment that have been developed in the last 20 years can help researchers seeking to understand how science laboratory resources are used, how students' work in the laboratory is assessed, and how science laboratory activities can be used by teachers to enhance intended learning outcomes. Current researches suggest that, in fact, students know more than they are able to demonstrate on multiple-choice exams (Eubanks, 1997).

In the context of science laboratory, students are graded on the perform-anace of manipulating variables, using scientific apparatus, identifying hypo-theses, making measurements and calculations, organizing and managing data, and the communication of results (Slater, 1997). Graded laboratory perform-ances go far beyond grading a final field report – this strategy considers the process that becomes the laboratory report as well (Slater, 1997). Assessments of students' performance and understandings associated with the science laboratory should be an integral part of the laboratory work of teachers and students (Hofstein & Lunetta, 2003). Assessment tools should examine the students' inquiry skills, their perceptions of scientific inquiry, and related scientific concepts and applications identified as important learning outcomes for investigation or the series of investigation. Knowledge about how to assess learning in the school science laboratory has increased substantially, and new techniques and media that can support the assessment of students' practical skills and associated understanding have been developed. According to Hofstein and Lunetta (2003), authentic and alternative assessment methods have been developed and validated to measure outcomes of school science programmes, including inquiry and activity in the laboratory. The laboratory assessments require students to design an investigation, collect and analyze data, and formulate findings. The students' visual representation and interpretation of their quantitative data is incorporated in the analysis. Criteria are used by researchers and teachers to unobtrusively observe and rate each student during normal laboratory activities. Students are assessed according to the following broad phases: planning and design; performance; analysis and interpretation; and application.

Hofstein and Lunetta (2003) further contend that recent developments in the use of new technology tools being used in science classrooms have high potential to help researchers and teachers to monitor students' work and ideas. The new practical assessment resources and strategies can be used by researchers and teachers to assess learning associated with inquiry and laboratory performance. The most promising efforts in assessment reform are those that address directly the relationship of assessment and instruction, specifying precisely how assessment can be used to support improved instructional practice. If we truly value the development of knowledge, skills and attitudes that are unique to practical work in science laboratories, appropriate assessment of these outcomes must be developed and implemented continuously by teachers in their own laboratory classrooms (Hofstein & Lunetta, 2003). Ossei-Anto (1996) also confirms that teachers trust and value tests and performance tasks they have developed because these serve their needs most. All the students' learning experiences should be assessed and the assessment should be authentic. Researchers, teachers, and testing jurisdictions whose goal is to assess comprehensively the learning that takes place in school science generally, or in the school laboratories more specifically, should use appropriate assessment tools and methodologies to identify what the students are learning (conceptual as well as procedural). According to Eubanks (1997) students who are most successful in examinations are those who use similar problem-solving techniques as the individual who wrote the items. Laboratory assessment should be based on the notion that laboratory knowledge should be demonstrated rather than through questions that can often be answered without actually having had to do any laboratory work at all (Eubanks, 1997).

Gender Differences in Science Performance Assessments

Ossei-Anto (1996) reported that, statistical differences were found with gender on the performance tasks, where males scored higher on the refraction of light planning task, while females scored higher on the reflection of light performing and reasoning tasks. Ossei-Anto (1996) went on further to assert that this finding was both surprising and encouraging for girls to outperform boys on a physics laboratory activity. According to him some questions to ponder about is whether girls did better than boys on the tasks that involved mirrors because girls, in everyday life experiences, use mirrors than boys.

Ruiz-Primo and Shavelson (1999) also reported from a study that, results show that on 5 of the 10 performance assessments used in the study no gender effect was found. Differences on the other 5 assessments seemed to be related to the particular science content that was assessed. They therefore concluded that students' prior experiences, their interaction with instruction, and teachers' preconceptions all play an important role in testing students' performances. In yet another study in California, it was found that girls tended to have higher overall mean scores than boys on the performance measures, but boys tended to score higher than girls on certain types of questions within a performance task (Klein *et al*., 1997). Proponents of education reform recommend replacing traditional multiple-choice tests with performance assessments. A factor cited in support of this recommendation is that females usually score lower than boys on traditional multiple-choice tests (Klein et al., 1997).

Several hypotheses have been advanced to explain gender differences in test scores. One theory is that the multiple-choice format itself favours males. It is believed that multiple-choice tests reward students who guess, and boys are more willing to guess than girls (Klein *et al*., 1997). Again, it has been postulated that students do better on test items that deal with objects or events that are drawn from their own "sphere of experience"; and that boys would have advantage on tasks that are sensitive to experiences with science-related activities. It has also been postulated that gender-related differences in specific cognitive abilities may lead boys to perform differently on certain items. Proponents of alternative assessments have suggested that performance assessments will reduce differences among groups by reinforcing appropriate curriculum changes and by providing students with hands-on opportunities to demonstrate their knowledge and understanding of scientific principles, not simply by recalling facts, but by constructing solutions (Klein *et al*., 1997). These measures emphasize the process by which students generate solutions, not just the correctness of the solution itself. The underlying theory is that individuals approach problem-solving differently because of varying styles and not because of different abilities. Therefore performance assessments are expected to narrow the differences in scores among groups because they are designed to allow for individual variation and they put less emphasis on guessing, exposure to science-related activities outside the classroom, testwiseness, and other extraneous presumed factors.

Science achievements in several content domains were explored with performance assessments, and it was found that males and females had similar mean scores. The few significant differences that emerged depended on the specific science content domain assessed. It was found that girls tended to score slightly higher than boys on the performing assessments (Klein *et al*., 1997). However, although certain types of performance tasks favour girls, other types favour boys. It is believed these performance patterns are related to the emphasis a question or task places on certain cognitive abilities or skill experiences. Therefore differences between boys and girls on performance assessments are sensitive to the specific types of tasks (Klein et al., 1997). Ssempala (2005) cites Chou that in a study in Taiwan girls consistently performed better than boys in all achievement variables. Ssempala (2005) found in a study conducted in Kampala District in Uganda that, there was no statistically significant differences between girls and boys in their ability to manipulate the apparatus/equipment, take observation, report/record results correctly, and compute/interpret/analyze results during chemistry practical.

The Programme for International Student Assessment (PISA) is a triennial survey of knowledge and skills of 15-year-olds. It is the product collaboration between participating countries and economies through the Organization for Economic Cooperation and Development (OECD), and draws on leading international expertise to develop valid comparisons across countries and cultures (OECD, 2007). More than 400 000 students from 57 countries making up close to 90% of the world economy took part in PISA 2006. The focus was on science. One of the main findings of PISA 2006 is that: males and females showed no difference in average science performance in the majority of countries, including 22 of the 30 OECD countries. In 12 countries, females outperformed males on average, while males outperformed females in 8 countries. Most of these differences were small (OECD, 2007).

According to Spelke (2005), long standing claims have been made that seek to explain gender differences in scientific and mathematical performance of males and females. One claim asserts that males and females are predisposed from birth to learn about different things: male infants learn about objects and their mechanical relationships, whereas female infants learn about people, emotions, and personal relationships (Spelke, 2005). From the beginnings, boys are more apt than girls to develop the knowledge and skills required by mathematics and science. Another claim is that genetically males have a better command than females on specific cognitive systems that give rise to effective reasoning in science and mathematics. A third claim is that males show greater variability in inherent mathematical and scientific talent than females, therefore they predominate in the pool of highly talented students from which future mathematicians and scientists will emerge. According to Spelke (2005) these claims have been reviewed in the light of research on the developmental and cognitive foundations of mathematical and scientific thinking. A review of evidence from studies of infants, children and adults yields little support for these claims.

Infants show few cognitive differences and no male advantage in the processing of objects, space or number. Although research on older children and adults has revealed differences between the performance of males and females on specific cognitive tasks, this research provides no evidence for sex differences in overall aptitude for mathematics or science at any point in development (Spelke, 2005). Spelke (2005) asserts that studies suggest that human talent for mathematical and scientific thinking has a considerable genetic basis in a set of core systems for representing objects, space and number. These systems emerge early in infancies, remain present throughout life, are harnessed by children when they learn mathematics, and are used by adults when engaging in mathematical and scientific thinking. Evidence suggests that these core systems are equally available to males and females (Spelke, 2005). Males and females show somewhat different cognitive profiles when confronted with complex tasks that can be solved by multiple strategies, but they show equal performance on tasks that tap the core foundations of mathematical thinking. Moreover, males and females show equal abilities to learn advanced, college-level mathematics. Insofar as mathematical ability is central to students' progress in the sciences, males and females would seem to be equally capable of learning and performing in science (Spelke, 2005).

An evolutionary account of sex differences in mathematics and science supports the conclusion that, although sex differences in math and science performance have not directly evolved, they could be indirectly related to differences in interests and specific brain and cognitive systems (Halpern *et al*., 2007). According to Halpern *et al*. (2007), experience alters brain structures and functioning, so causal statements about brain differences and success in math and science are circular. A wide range of sociocultural forces contribute to sex differences in mathematics and science achievement and ability – including the effects of family, neighbourhood, peer, and school influences; training and experience; and cultural practices. Halpern et al. (2007) are of the view that there are no single or simple answers to complex questions about sex differences in science and mathematics.

Shaw and Nagashima (2009) undertook a study in which they examined student learning in science as measured by performance assessments embedded within inquiry-based units of instruction. Using mean scores as the basis for comparison, results showed the majority of students achieving at the proficient level as defined by initiative-developed rubrics. According to Shaw and Nagashima (2009) depending on the performance assessment and the student subgroup (male or female), potential factors related to performance include science discipline and access to economics-related resources such as computers. While findings from various performance assessments indicate mixed achievement gaps by gender (male and female), the combined findings from these various studies indicate females outperforming males (Shaw & Nagashima, 2009). They also found in their study that the mean scores for females in all three performance tasks were consistently above the mean scores for the total sample; while the mean scores for the males on all three performance tasks were consistently below the mean scores for the total sample. This means that the mean scores for females on all three performance tasks were consistently above those for males; which imply that females consistently outperformed their male counterparts on all three performance tasks. Shaw and Nagashima (2009) also assert that the over-performance of females in relation to their male counterparts have been found in studies by Geier and colleagues, Johnson and colleagues, and Lee and colleagues. They contend that gender gaps have been shown to be sensitive to assessment type and content orientation. While the general pattern is one of girls outperforming boys; it has been found that boys outperform girls on multiple-choice tests; and girls outperform boys on performance assessments. In yet another study on performance assessment, it was found that females perform comparably with males on physical science tasks while girls outperform boys on the life science tasks. Shaw and Nagashima (2009) on their part, found in their study on performance assessment that females outperform males on both life and physical science. They assert that these apparent contradictions might be explained by the strong connection of the particular physical science content to the life science.

Types of Schools Differences in Science Performance Assessments

According to Downs (2007) the level of performance of a student depends on the type of school he/she attends. In a study, Downs (2007) found that students attending virtual schools do have greater success in mastering the important subjects and skills of science and mathematics than any public schools. Stranahan, Borg and Borg (2002) reported that education in the state of Florida in 1997 relied heavily on standardized test scores as a measure of school effectiveness. In this education reform, known as School Recognition Programme, letter grades from “A” through “F” were given to schools as a measure of their effectiveness. The primary criteria by which these letter grades are assigned are individual students’ scores on standardized assessment tests given in a year. According to Stranahan *et al*. (2002) the grade that a student receives has a major financial impact on the funding of public schools. Stranahan *et al*. (2002) conducted a study to determine if a school's grade depends on the intrinsic qualities of the school or on the qualities of the individual students who go to that school. They found that students attending a school with higher rates of teacher turnover have lower scores on assessment tasks. In these schools a higher proportion of the teachers are new to the school. It was also found that the performance of students on the assessment tasks depend on the levels of education (advanced degrees) of the teachers as well as their years of experience in teaching. They also found that the performance of students on performance assessments in a school depend on the class size as well as the effectiveness of the leadership in the school. Students in schools with smaller class sizes tend to perform better on assessment tasks.

Underwood and McCafferey (1990) reported in the findings of their study that, when single-gender group worked together there was an increase in the level of activity in response to the task that was assigned to them, and this was associated with an increase in the number of correct completions. When mixed-gender group worked, however, there was no noticeable increase in the number of attempts or number of correct completions. The report went on further to say that single gender groups of boys and girls improved in their number of attempts and in their correct attempts, but mixed groups did not. Single gender groups show an improvement relative to the performance of the same children (students) when working individually, but groups composed of boys and girls show no improvement. There was indeed a tendency for the mixed-gender group to perform at a level slightly poorer than their previous individual level (Underwood & McCafferey, 1990). Underwood and McCafferey (1990) state that, in general, the single-gender groups worked by discussion and agreement, with each member of the group contributing to the decision. Children working in mixed gender groups tended not to work by negotiation to achieve joint problem-solving, but co-operated by instruction. There was very little discussion of alternative solutions in the mixed -gender groups. Harvey (1985) also reported that there is significant difference in the performance in physics between boys taught in single-sex and mixed schools. Also, there is significant difference in performance in physics between girls taught in mixed and single-sex schools. However, there was no significant difference in performance in physics between boys taught in mixed schools; likewise there was no significant difference in performance in physics between girls taught in mixed schools.

Summary of Major Findings of Review of Related Literature

1. To ensure that instruction goes on in the classroom as planned, there is the need for regular assessment and evaluation of teaching and learning.

2. In order to make educational decisions, there is the need to obtain high quality information; and for sound subjective judgment to be made.

3. Making educational decisions, evaluations and assessments call for the use of tests and other techniques to measure educational achievement.

4. Thus tests, measurements, assessments and evaluation are closely connected and relevant for educational practice.

5. To measure educational achievement, the learning targets or terminal behaviours of instruction must be clearly defined.

6. Tests and other measuring instrument must be valid and reliable to ensure that inferences and interpretations based on their scores are accurate and appropriate.

7. Paper-and-pencil tests have been criticized for not measuring all and other important learning outcomes.

8. New trends in assessment practices call for the use of performance and authentic assessments to measure and tap hands-on and minds-on skills.

9. Science performance assessments involves assessment of students as they are engaged in scientific investigations in real-life or simulated situations.

10. Laboratory assessment is a typical example of science performance assessment.

11. Factors affecting the achievements of students in science performance assessments include: gender of the students, type of school the student attends, type of task assessed (physical science, biological science etc.).

CHAPTER THREE

METHODOLOGY

Overview

The chapter covers the procedures taken to carry out the research. The chapter highlights the research design employed for the study. It also talks about the population, sample and sampling procedure, the instruments (tasks) used and how it was developed for the study. The chapter finally gives account of how the data was collected and analysed.

Research Design

The "basic skills assessment" method also known as the "psychometric assessment testing" approach was employed in this research to assess laboratory skills proficiencies of physics students in selected topics in mechanics and optics, to determine if they exhibit adequate laboratory skills of planning, performing and reasoning. Psychometric tests were designed to produce effective and consistent measures of the levels of planning, performing and reasoning skills possessed by the physics students. Performance tasks were administered to the students that were scored right or wrong. The research design allowed the levels of skills possessed by the students to be quantified in numerical terms, and were analyzed using statistical methods. The "psychometric assessment testing" approach also enabled relationships between the independent and dependent variables to be determined.

The research questions were specific, measurable and capable of rigorous statistical analysis (Somekh & Lewin, 2005). The respondents were provided with laboratory equipment, problems were posed to them, and they were allowed to use those resources to generate solutions to the tasks.

Weaknesses of the research design include difficulty in controlling the abilities of the students used for the study. Again, it was difficult to manipulate the independent variables to randomly assign a student to any type of school (boys, girl or mixed); and to randomly assign a student to any gender (male or female).

Population

There were nine SHSs in Sekondi-Takoradi metropolis that offer physics as a subject to students. The targeted population for the research was all SHS 3 physics students in the nine schools within the metropolis. The accessible population, however, was all SHS 3 physics students in seven schools in Sekondi-Takoradi metropolis.

There were a total of 551 physics students in SHS 3 in the seven selected schools. SHS 3 physics students were used for the research because they had spent at least two years studying physics in their schools. It was therefore hoped that after two years of receiving instruction in physics laboratory (practical) activities, they may have acquired at least minimum laboratory skills and competencies to be able to respond to the tasks on the instrument.

Sample and Sampling Procedure

The seven schools used for the main research comprised of two boys, one girls, and four mixed schools. Only one girls school was involved in the study because that was the only girls school within Sekondi-Takoradi metropolis which offer physics as a subject to females. The types of schools chosen for the study were guided by the research questions. Sekondi-Takoradi metropolis was used for the main research because there were nine senior high schools within the metropolis that made it feasible for the study to be carried out in the city. Computer (MS Excel) generated random numbers were used for sampling the students.

Probability sampling methods were used in sampling the students. In this way it was ensured that each student in the population had equal chance of being selected for the study (David & Sutton, 2004). Stratified and simple random sampling procedures were used. Many schools put students of different abilities in different classes. Therefore the stratified random sampling method was used to produce more representative samples of the students for the research. There were two SHS 3 physics classes in school A. Stratified random sampling was used to sample 40 students from the school. Twenty students were randomly selected from each class in school A. In school B, there were four SHS 3 physics classes in the school. Stratified random sampling was used to sample 10 students from each of the four classes in school B. Thus, a total of 40 students were selected from the school. However, there was only one SHS 3 physics class in each of schools C, D, E and F. Simple random sampling was used in schools C, D, and E to select 40 students from each of the schools; while all the 49 students in school F were used for the study. There were two SHS 3 physics classes in school G, and stratified random sampling was used to select 40 students from the school. Twenty students were selected from each class in school G. In all a total sample size of 289 SHS 3 physics students was used for the main research. This represents52.5% of the population of the (551) SHS 3 physics students accessible for the research. The notion of large sample size was to allow the researcher to state, with a certain level of confidence that findings using the sample would also be found in the population (David & Sutton, 2004).

Instruments

The instruments for the research were performance (practical) tasks that were administered to the students. The instruments were designed by the researcher to assess laboratory skills of physics students in selected topics in mechanics and optics. There were three performance tasks or instruments. Task A was a planning task, which demanded the respondents to design a laboratory procedure or state the steps they would take to move a drum full of engine oil from the garage onto a raised platform in the compound of a construction firm. Task B was a performing task that required the respondents to carry out actual experiments, manipulate materials, make observations, record data, and to take decisions on a practical strategy to determine the relative density of sand. Task C was a reasoning task which demanded the respondents to make measurements, record data, process the data, explain relationships, form generalizations, discuss data accuracy and to state sources of errors and limitations on passage of light rays through a triangular glass prism. The tasks were designed in such a way that they were independent of each of other – knowledge and information on one task was not needed to respond to the other tasks.

Validity and Reliability of Instruments

Experts opinions from the team of supervisors were sought, as well as references from textbooks were made to determine the face, content and construct validity of the instruments. Consensus from the experts' opinions of the team of supervisors and textbook writers concerning important objectives, skills, and content in performance assessment, adequately defined the validity of the instruments (Schwart *et al*., 1962). This was a sound conclusion to make, for these experts have studied the field carefully and have a good idea of what are valid objectives, skills and content. Pilot-testing of the instruments was done in three selected senior high schools within Cape Coast metropolis prior to the main research to check the reliabilities of the instruments. Reliabilities of the instruments were improved through the use of diagrams to clarify the tasks. Clear directions on the tasks administration were given to students to remove ambiguities and help improve reliabilities of the instruments. The researcher personally administered the tasks to respondents to ensure that each student was given fair and adequate time and resources to complete the tasks. This further improved reliabilities of the instruments greatly. Scoring formats (schemes) were used to score the responses of the respondents to ensure uniformity in scoring which also improved reliabilities of the instruments.

The main research was done using a sample of 289 SHS 3 physics students in Sekondi-Takoradi metropolis. Responses of the students were scored by the researcher. Another physics teacher was given 111 sub-samples of the students' responses to score independently. The two sets of the students' scores, from the researcher and the physics teacher, were correlated to estimate the inter-rater reliabilities of the instruments. Statistical package for social sciences (SPSS) programme was used to calculate the reliability coefficients. Cronbach alpha coefficients of the tasks were determined to find the internal consistency of the items on each instrument. Moreover, inter-tasks correlations of the instruments were estimated to ensure that the tasks were independent from each other. The inter-rater reliabilities were: 0.93 for Task A, 0.96 for Task B, 0.94 for Task C, and 0.93 for the Total Tasks. The inter-tasks correlations were: 0.073 between Task A and Task B, 0.015 between Task A and Task C, and 0.000 between Task B and Task C. The Cronbach alpha reliabilities were: 0.83 for Task A, 0.90 for Task B, 0.88 for Task C, and 0.76 for the Total Tasks.

An opinionnaire was also administered to the sample of students to solicit their opinions on the difficulty of the tasks they performed, as well as the adequacy of the time given to them to complete the tasks. Samples of the tasks or instruments for the research are shown in appendices B, C, D and E.

Pilot Test of Instruments

Pilot-testing of the instruments was done prior to the main research to test the appropriateness of the research questions; to check the intended statistical and analytical procedures; and to carry out evaluation of the research instruments before the main data collection. Pilot-testing of the instruments was done in Cape Coast metropolis of the Central Region. There were a total of seven senior high schools in Cape Coast metropolis that offer physics as a subject to students. However, three schools: one boys, one girls and one mixed schools were used for piloting the instruments. Stratified random sampling method was used to select 10 students from each of the three schools. A total sample size of 30 students was involved in the pilot test. Experts' opinion from the team of supervisors and textbook references were sought to determine the face, content and construct validities of the instruments prior to the test. Permission was sought from the headmasters, assistant headmasters and physics teachers to do the pilot test in their schools. The instruments were administered to the students on specific dates and their responses were independently scored by the researcher and another physics teacher. Scores obtained by the respondents were analysed to determine and estimate the validity, reliability and usability of the instruments. The inter-rater reliabilities were: 0.82 for Task A, 0.88 for Task B, 0.85 for Task C and 0.88 for the Total Tasks. These high inter-rater reliabilities indicated that the scoring formats could produce consistent measures of the students' responses, and that there was no bias in scoring the students' responses. The Cronbach alpha values were: 0.66 for Task A, 0.71 for Task B, 0.83 for Task C and 0.74 for the Total Tasks. The high Cronbach alpha values also suggested that the items on the instruments were internally consistent; and that the instruments were good for assessing laboratory planning, performing and reasoning skills of the students. Inter-tasks correlations were: 0.05 between Task A and Task B; 0.18 between Task A and Task C; and 0.24 between Task B and Task C. The low inter-tasks correlations showed that the tasks were independent from each other, and that information on one task was not needed to respond to the other tasks. Other analyses of the scores were done to determine the suitability and appropriateness of the research questions. After analysis of students' responses from the pilot test, and with suggestions from the team of supervisors, minor corrections were made to the research instruments to ensure high validity and reliability before the instruments were used for the main data collection.

Data Collection Procedure

The researcher obtained a letter of authorization from the Department of Science and Mathematics Education of the University of Cape Coast to carry out the research in the selected schools. The researcher went to the selected sc-hools to meet the headmasters/headmistresses and physics teachers to ask for permission and their cooperation to do the study. The researcher paid regular visits to the schools prior to the research to familiarize himself with the schools, and established rapport with the staff and students for the study.

After sampling the students for the main research, the researcher had discussions with the physics teachers in the seven schools to arrange for convenient dates for administration of the instruments. The researcher explained the nature of the tasks (instruments) to the students and teachers. The students were told in advance to bring calculators, rulers and mathematical sets along with them on the actual dates set for administration of the tasks. The researcher made advance arrangements to obtain laboratory equipment for administration of the instruments. Materials obtained by the researcher for the research include relative density bottles, electronic balances and beakers, measuring cylinders, beam balances, water and sand.

On the actual days for administration of the tasks, the sampled students were confined in classrooms. Answer booklets, graph sheets, laboratory equipment and the tasks were given to them. The researcher spelt out the rules and regulations guiding the conduct of the tasks to the students; and allowed them to ask questions for clarification of any point regarding the rules and regulations that was unclear to them. Five minutes and a further 25 minutes were given to the students to read through and to respond to each task. A total of1:30 minutes was used to respond to the performance tasks in each school. The researcher collected the answer booklets after the students had finished responding to the tasks. The researcher then distributed the opinionnaire to the students for them to respond to it. The researcher collected the opinionnaire from the students after they had responded to it. The researcher cleaned and packed the laboratory equipment used for the main study at their appropriate places after completion of the tasks administration. The researcher then thanked the headmasters/headmistresses, physics teachers and the students and left the schools. Responses of the students were scored using marking rubrics. Samples of the marking rubrics are shown in Appendixes F, I and K. Scores obtained by the students were input into SPSS programme for analysis. Again, responses of the respondents to the items on the opinionnaire were input into SPSS and coded for analysis.

Data Analysis

Data from the main research were analysed using the research questions as guide. The data were organized, coded and input into SPSS programme for analysis. Means, standard deviations, correlation coefficients, frequencies and percentages of the data were determined. ANOVA and t-test of the data were also determined. Most of the statistical tools used to analyse the data are parametric techniques. Parametric techniques are generally more powerful than nonparametric techniques and hence much more likely to reveal a true difference or relationship if one really exist (Fraenkel & Wallen, 2000). Parametric techniques, however, make various kinds of assumptions about the nature of the population from which the samples involved in the study are drawn. The analysis of the data was presented using the following parameters:

i. Age.

ii. Skills of Planning, Performing and Reasoning.

iii. Gender.

iv. Types of school.

Quantitative analysis rather than qualitative analysis of the data was done. This sort of analysis gave precise magnitudes of outcomes of the research. The quantitative analysis allowed objective discussions of the results to be made rather than vague approximations.

CHAPTER FOUR

RESULTS AND DISCUSSION

Overview

This chapter gives full account of the statistical and analytical procedures used for the research. The results and discussions of the research have been organised along the lines of the research questions. The results have been presented under the following subheadings: (a) the scoring of the booklets; (b) distribution of sample for the study; (c) students' characteristics; (d) inter-rater reliabilities for tasks; (e) Alpha reliability coefficients for the tasks; (f) inter-tasks correlations; (g) levels of competency exhibited by the physics students in laboratory skills of planning, performing and reasoning; (h) differences in laboratory skills of planning, performing and reasoning shown as a result of differences in gender (males and females); (i) differences in laboratory skills of planning, performing and reasoning shown as a result of differences in the types of schools (boys, girls or mixed) of the respondents; and (j) students' comments to the performance tasks administered.

Overview of Statistical Procedures

Statistical Package for Social Sciences (SPSS) programme was used for the statistical analysis. Responses of the students were scored independently by the researcher and another physics teacher using the scoring formats shown in Appendices F, H and J. The two sets of scores from the two raters were correlated to find the inter-rater reliabilities of the tasks. The results obtained and the discussions on the results are shown below. Means, standard deviations, frequencies, percentages, t-tests and ANOVAs of the data were determined.

Distribution of Research Sample

The distribution of the sample for the main research is as shown in Table 1 and Figure 1.

Table 1
Distribution of Research Sample by Type of School and Gender

School	Type of School	Gender		Number of Respondents
		Male	Female	
A	Boys	40	0	40
B	Boys	40	0	40
C	Mixed	19	21	40
D	Mixed	23	17	40
E	Mixed	14	35	49
F	Mixed	22	18	40
G	Girls	0	40	40

N = 289.

From Table 1 and Figure 1, the sample for the main research was drawn from seven senior high schools in Sekondi-Takoradi metropolis of Western Region of Ghana. The respondents were 289 senior high school 3 (SHS 3) physics students from the seven schools in Sekondi-Takoradi. The sample size represents 52.5% of the population of 551 SHS 3 physics students accessible for the research. The respondents were male and female students sampled from boys, girls and mixed schools within the metropolis.

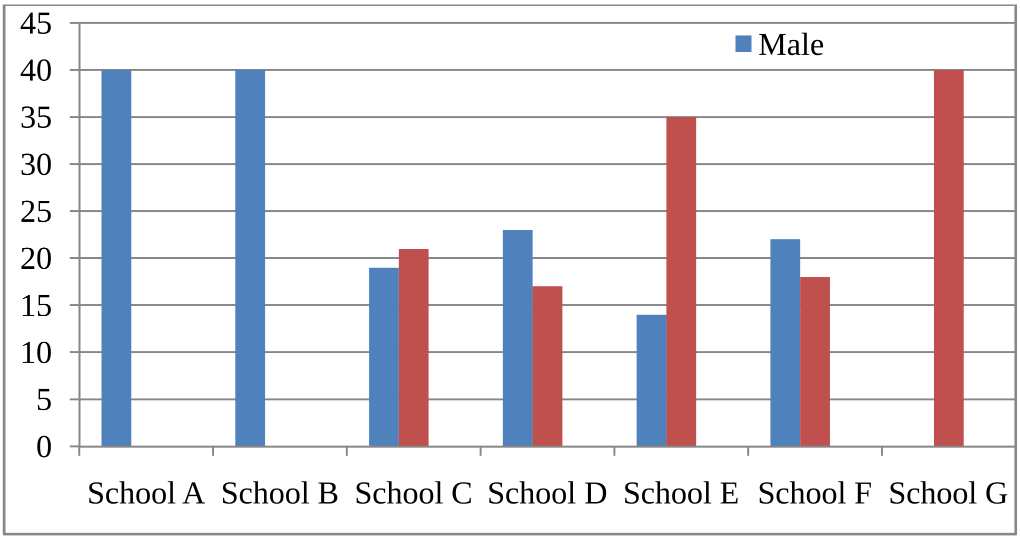

Figure 1. Distribution of Research Sample by Types of Schools and Gender

Table 2 shows distribution of the research sample by schools.

Table 2

Distribution of Research Sample by Schools

School	Number of Respondents	Percentage (%)
A	40	13.84
B	40	13.84
C	40	13.84
D	40	13.84
E	49	16.96
F	40	13.84
G	40	13.84

N = 289.

From Table 2, 40 students representing 13.84% of the sample were randomly sampled from each of the six schools – A, B, C, D, F and G. However, all the 49 students in school E, representing 16.96% of the sample were used for the study.

Table 3 and Figure 2 show the distribution of the research sample type of school.

Table 3

Distribution of Research Sample by Types of Schools

Type of School	Number of Students	Percent (%)
Boys	80	27.7
Girls	40	13.8
Mixed	169	58.5

N = 289.

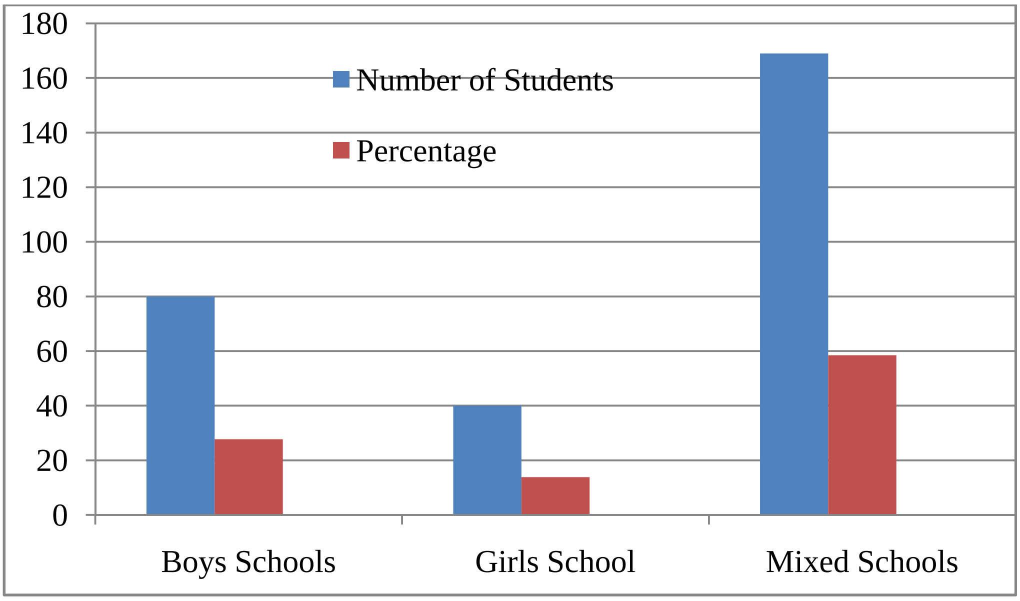

Figure 2: Distribution of Research Sample by Types of Schools

Results from Table 3 and Figure 2 show that 80 students representing 27.7% of the sample were selected from the two boys schools – A and B; 40 students representing 13.8% of the sample were selected from the girls school - G; while 169 students representing 58.5% of the sample were sampled from the four mixed schools – C, D, E, F.

Table 4

Distribution of Research Sample by Gender

Gender	Number of Students	Percent (%)
Male	158	54.7
Female	131	45.3

N = 289.

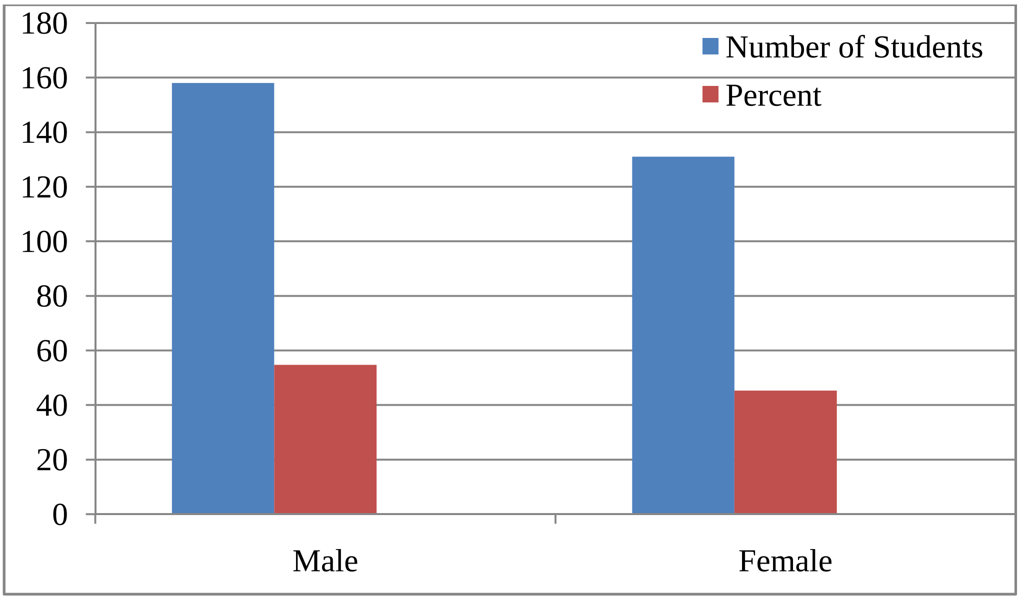

Figure 3: Distribution of Research Sample by Gender

From Table 4 and Figure 3, there were 158 boys representing 54.7% of the research sample; and 131 girls representing 45.3% of the research sample. The number of girls sampled for the main research was slightly less than the number of boys. This was due to fact that the mixed schools selected for the research had many girls offering elective science. Moreover, the sampling was not done with the view to give advantage to any gender over the other.

Table 5 and Figure 4 show distribution of ages of the research sample.

Table 5

Distribution of Ages of Research Sample

Age(years)	Number of Respondents	Percentage (%)
16	14	4.8
17	141	48.8
18	98	33.9
19	29	10.0
20	7	2.4

N = 289.

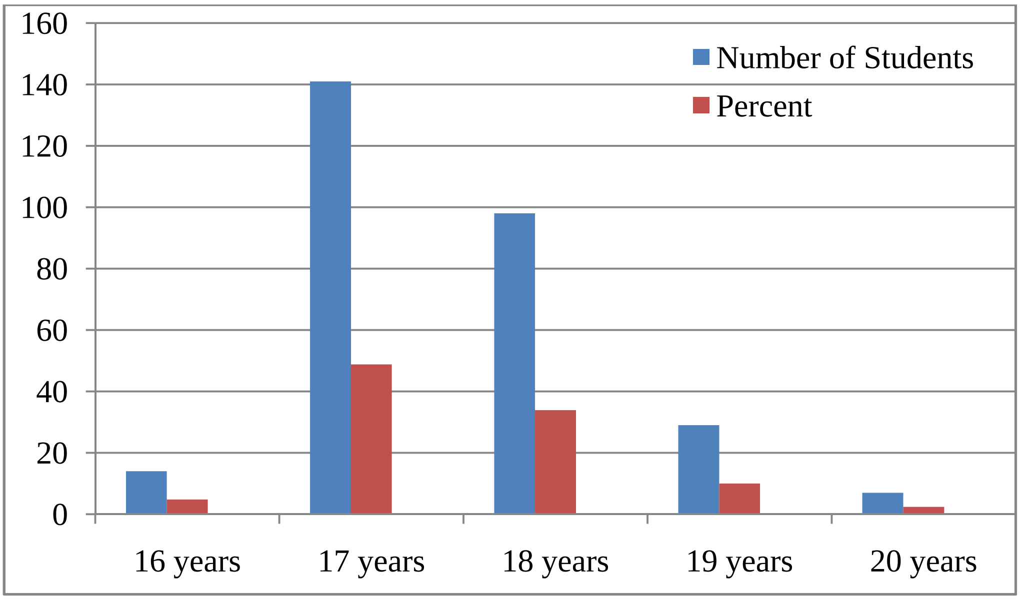

Figure 4: Distribution of Ages of Research Sample

From Table 5 and Figure 4, ages of the respondents ranged from 16 to 20 years. The mean age of the respondents was 18 years and the standard deviation was 1year. Majority of the students, 141 of them, representing 48.8% of the sample were 17 years old. Seven students representing 2.4% of the sample were 20 years old; while 14 students representing 4.8% of the sample were 16 years old. This shows that all the students were at the formal operational stage, and potentially had the abilities to engage in laboratory investigations and to respond the tasks in the study.

Students Characteristics

The students (respondents) were drawn from senior high schools (SHSs) that offer physics as elective science subject to students within Sekondi-Takoradi metropolis. The respondents were SHS 3 physics students who had spent at least two years studying physics. The students were supposed to have acquired at least minimum skills and competencies in physics laboratory (practical) work. At SHS 3 the students know how to handle science equipment, make observations, record data, and process data to look for patterns, establish relationships between variables and to make generalizations.

After scoring the students' responses to the tasks, the scores were coded and input into SPSS software for analysis. There were three performance tasks for the students to respond to.

Tasks and Inter-Tasks Reliabilities and Correlations

Table 6 shows the inter-rater reliabilities for the individual and total tasks.

Table 6

Inter-Rater Reliabilities

Task	Correlation Coefficient
Total Tasks	0.93
Task A (Planning)	0.93
Task B (Performing)	0.96
Task C (Reasoning)	0.94

N = 289.

As shown for the two raters (researcher and the other physics teacher who scored 111 subsample of students performance tasks booklets), the inter-rater reliability was 0.93 for the total performance tasks with 35 maximum number of credits. The proportion of agreement for the planning task with 8 credit points was 0.93. The proportion of agreement for the performing task with 11 credit points was 0.96; while the proportion of agreement for the reasoning task with 16 credit points was 0.94. The various correlations were at a significant level of 0.01. The inter-rater reliabilities determined by the researcher and the other physics teacher were very high. These show that the scoring formats (schemes) used for scoring the three performance tasks were very reliable and could be used to produce consistent and accurate measures of students' responses to the tasks. The reliability coefficients also showed that there were no biases in scoring the students' responses.

The values of correlation coefficients (Pearson) between the performance tasks are as shown in Table 7.

Table 7

Correlation Coefficient (Pearson) Values between Tasks

	Task A	Task B	Task C
Task A	-	0.073	0.015
Task B	0.073	-	0.000
Task C	0.015	0.000	-

N = 289.

Correlation coefficient between Task A (Planning Skills) and Task B (Performing Skills) was 0.073. The correlation coefficient between Task A (Planning Skills) and Task C (Reasoning Skills) was 0.015; while the correlation coefficient between Task B (Performing Skills) and Task C (Reasoning Skills) was 0.000. The low values of the correlation coefficients between the tasks indicated that they were independent from each other; and that information on one task was not needed to respond to the other tasks. The low values of the reliabilities also point to the fact that the level of competency demonstrated by a student in one task is not dependent on or does not affect the level of competency demonstrated by the student on the other tasks.

Table 8 shows Cronbach alpha reliability values for the individual and the total tasks.

Table 8

Cronbach Alpha Reliability Coefficients for Individual and Total Performance Tasks

Task	Points/Items	Alpha Coefficient
Task A (Planning)	8	0.83
Task B (Performing)	11	0.90
Task C (Reasoning)	16	0.88
Total Tasks	35	0.76

N = 289.

Alpha reliability coefficients of the individual tasks (Task A, Task B and Task C) as well as the total tasks were determined. Values of the reliability coefficients for the tasks were high. Cronbach Alpha reliability value for the total tasks was 0.76 as shown in Table 8. The reliability coefficient was 0.83 for the planning skills which had 8 items. The performing skills with 11 items had a reliability coefficient value of 0.90; while the reasoning skills with 16 items had reliability value of 0.88. These high values of Cronbach alpha indicate that the items on the tasks were internally consistent. The high values also show that the tasks on the instrument were good for assessing laboratory skills of planning, performing and reasoning of physics students in the selected topics in mechanics and optics.

Response Patterns

Details of the response patterns to the performance tasks are shown on Table 9 through Table 16.

Research Question 1:

To what extent do senior high school physics students engaged in laboratory work exhibit adequate competencies in the skills of:

i. Planning?

ii. Performing?

iii. Reasoning

Means and Standard Deviations of Individual and Total Performance Tasks

The mean scores and standard deviations for the performance tasks are as shown in Table 9.

Table 9

Mean Scores and Standard Deviations of Performance Tasks

Task	Number of Items	Mean	Standard Deviation
Task A (Planning)	8	5.8	2.36
Task B (Performing)	11	9.6	2.55
Task C (Reasoning)	16	13.3	3.50
Total Tasks	35	28.7	5.78

N = 289.

Using the scoring formats provided for the planning, performing and reasoning tasks, as are shown in Appendices F, H and J, the median scores were 4 for the planning skills; 5.5 for the performing skills; 8 for the reasoning skills and 17.5 for the total performance tasks. Scores at the median marks show average levels of proficiency in laboratory skills. Scores above the median marks show high levels of proficiency, while scores below the median marks show low levels of proficiency. The overall attainment of the respondents in the research was high. The mean scores for the tasks were: 5.8 for the planning skills representing 72.5% of the total marks of 8; 9.6 for the performing skills representing 87.3% out of the total marks of 11; and 13.3 for the reasoning skills representing 83.1% out of the total marks of 16. The total performance tasks had a mean score of 28.7 representing 82.0% out of the total marks of 35. The values of the mean scores for the planning skills, performing skills and reasoning skills from Table 7 were above the median scores of 4, 5.5 and 8 for the planning, performing and reasoning skills respectively. These values therefore suggest that the students exhibited high levels of proficiency in the skills of planning, performing and reasoning in the selected topics in mechanics and optics. Again, the high mean score for the total performance tasks of 28.7, which is above the median mark of 17.5 further suggests that on the whole the students showed high levels of competency on the performance tasks. It can also be inferred from the mean scores of 9.6 (87.3%) for the performing skills; 13.3 (83.1%) for the reasoning skills and 5.8 (72.5%) for the planning skills that the students exhibited the highest level of proficiency in the performing skills. Reasoning skills was the next high level of proficiency showed by the students, while the level of proficiency showed in the planning skills was the least among the three laboratory skills. The standard deviation for the planning skills was 2.36; standard deviation for the performing skills was 2.55; standard deviation for the reasoning skills was 3.50; while the standard deviation for the total performance tasks was 5.78. These low values of standard deviations suggest that there were no much differences in the levels of proficiencies exhibited by the individual students in each of the skills of planning, performing and reasoning in the main research; and that individually and collectively the respondents exhibited high proficiencies in laboratory skills of planning, performing and reasoning.

After scoring the students' responses to the tasks, the responses were class-ified into two. One group was responses that received full credit of one point on an item, and the other group was responses that received no credit on the same item. Each item on a task was scored dichotomously. A response to the item either received full credit of 1 point for correct and appropriate answer, or no credit for wrong and inappropriate answer.

Planning Skills (Task A)

The distribution of credits for Task A is as shown in Table 10 and Figure 5.

Table 10
Distribution of Credits for Planning Skills (Task A)

Component Assessed[b]	Full Credit (%)	No Credit (%)
1 General Strategy	88.9	11.1
2 Sequential Plan	83.0	17.0
3 Detailed Plan	72.0	28.0
4 Workable Plan	65.1	34.9
5 Materials	65.4	34.6
6 Outlines/Diagrams	64.4	35.6
7 Safety Procedure (S1)	76.1	23.9
8 Safety Procedure (S2)	66.4	33.6

N = 289.

b. Details of items are in scoring format (Appendix F).

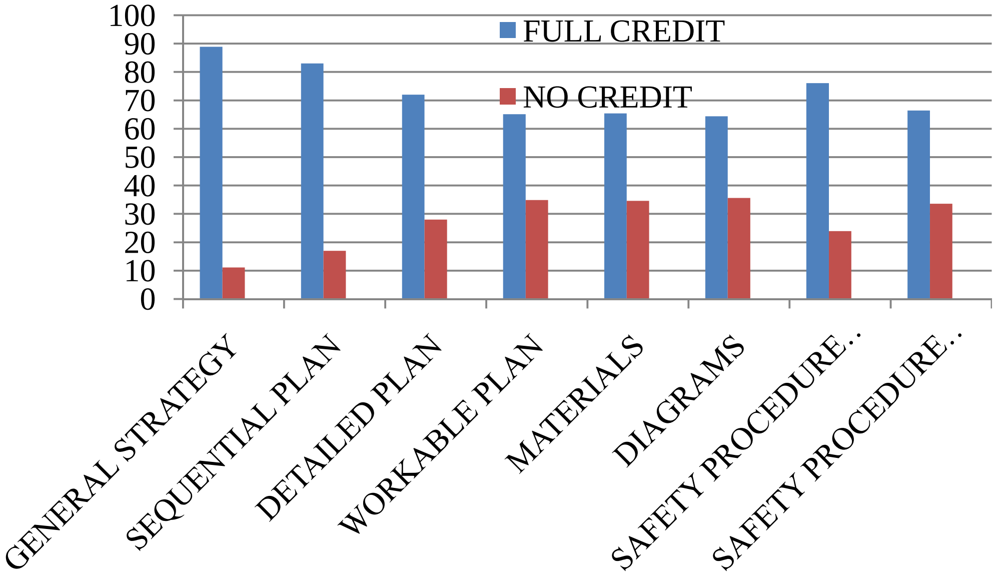

Figure 5: Distribution of Credits for Planning Skills (Task A)

The planning skills (Task A) required the students to provide detailed, step by step procedure on how three workers in a construction firm intend to move a drum full of engine oil from the garage onto a raised platform in the compo-und of the firm. Eight items were assessed on the planning skills (Task A). Two hundred and fifty seven out of the 289 students representing 88.9% of the respondents had full credit on general strategy, while 240 out of the 289 students representing 83.0% of the respondents had full credits on sequential plan. These results suggest that majority of the students demonstrated excellent levels of competency in general strategy and sequential plan.

More than 185 out of the 289 students, representing 64% of the respondents had full credits on all the eight items assessed under planning skills. This indicates that majority of the students showed high levels of proficiency in laboratory planning skills. On the other hand 103 students representing 35.6% of the respondents could not draw appropriate diagrams and outlines for their plans; while 100 out of the 289 students could not choose the relevant materials out of the list provided for their plans. These showed that many of the students exhibited some inadequacies on some aspects of the planning skills. This means that the students need more practice on laboratory planning skills for them to improve on these inadequacies.

Two examples of students' wrong and inappropriate answers as well as two examples of students' correct and appropriate answers on the planning skills (Task A) are shown below.

Examples of students' wrong and inappropriate answers

1. Arrange the steel rollers in a straight line with the car with the jack. The platform is raised in such a way that it is arranged in between the steel rollers and the car. The wooden board is made to incline to the platform raised. The drum containing the engine oil is tied with a rope attacked to the jack of the car. The drum is then roll on the 4 steel rollers and the car jack then folded the rope tied to the car. This causes the drum containing the engine oil to roll on the wooden board and finally onto the platform.
2. I roll the heavy drum of engine oil out of the garage. Arrange the wooden board on the platforms for the person to roll the heavy drum of engine oil in it well. Use the rope to tie the drum of engine oil very well. Roll the drum of engine oil from the floor gradually on the wooden board and the truck. Use the rope to left it from the wooden board on the truck gentle.

Some students provided either incorrect or incomplete answers to the planning skills (Task A) as shown in the examples above.

Two examples of responses provided by students that were found to be appropriate and correct are listed below.

Examples of students' appropriate and correct answers

1. The wooden board is placed inclined to the truck where the drum of oil is. The drum of oil is pushed on the incline plane onto the truck. The truck is pulled to the compound where the platform is. The wooden board is again placed inclined to the truck and the drum of oil is rolled down to the ground. The wooden board is then placed inclined to the raised platform. The drum of oil is pushed on the incline plane to the top of the platform.
2. In the first place I will put the drum of engine oil on the four steel rollers on the ground. Now I will gradually push the drum of oil slowly with the help of the steel rollers out of the garage till it gets to the raised platform. The wooden board will now be inclined to the raised platform. I will push the drum of engine oil on the incline plane till it gets to the top of the platform.

Table 11 and Figure 6 show the total scores obtained by the respondents on the planning skills (Task A).

Ninety four out of the 289 students representing 32.5% of the respondents scored the total marks of 8; while 61 students representing 21.1% of the respondents scored a total of 7 marks as shown in Table 11 and Figure 6. In all, 155 out of the sample of 289 students representing 53.6% of the respondents scored 7 and 8 marks. It can be inferred from these results that many of the students exhibited excellent levels of proficiency in laboratory skills of planning. Two hundred and fourteen out of the 289 respondents representing 74.1% of the research sample scored a total marks of 5 and above. This shows that majority of the students exhibited high levels of competency in laboratory planning skills in the main research.

Table 11

Distribution of Total Scores for Planning Skills (Task A) by Sample

Total Score	Number of Students	Percent (%)
0	9	3.1
1	12	4.2
2	21	7.3
3	13	4.5
4	20	6.9
5	25	8.7
6	34	11.8
7	61	21.1
8	94	32.5

N = 289.

a: Total marks for Task A is 8.

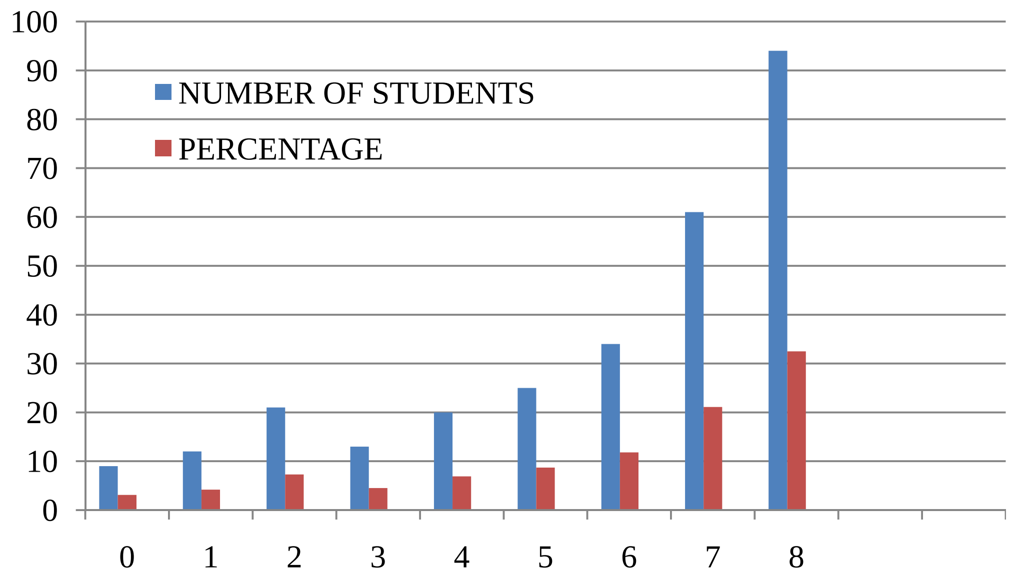

Figure 6: Distribution of Total Scores for Planning Skills (Task A) by Sample

On the other hand, 55 students representing 19% of the research sample scored a total marks of 3 and below. This result suggests that many of the students exhibited some inadequacies on the planning skills.

Performing Skills (Task B)

Table 12 and Figure 7 show distribution of credits for the performing skills. Eleven items were assessed under the performing skills.

Table 12

Distribution of Credits for performing Skills (Task B)

Component Assessed[b]	Full Credit (%)	No Credit (%)
1. Value of m_1	94.8	5.2
2. Value of m_2	88.9	11.1
3. Value of m_3	91.7	8.3
4. Value of m_4	94.5	5.5
5. Value of $(m_2 - m_1)$	86.2	13.8
6. Value of $(m_3 - m_2)$	86.5	13.5
7. Value of $(m_4 - m_1)$	91.0	9.0
8. Value of $(m_4 - m_1) - (m_3 - m_2)$	83.7	16.3
9. Relative Density of Sand	83.0	17.0
10. Safety Procedure (S1)	81.3	18.7
11. Safety Procedure (S2)	76.1	23.9

N =289.

b: Details of items are in scoring format (Appendix H).

From Table 12 and Figure 7, 274 students representing 94.8%; 273 students representing 94.5%; 265 students representing 91.7% and 263 students representing 91.0%, out of the research sample size of 289 students had full credits for the values of m_1, m_4, m_3 and $(m_4 - m_1)$ respectively.

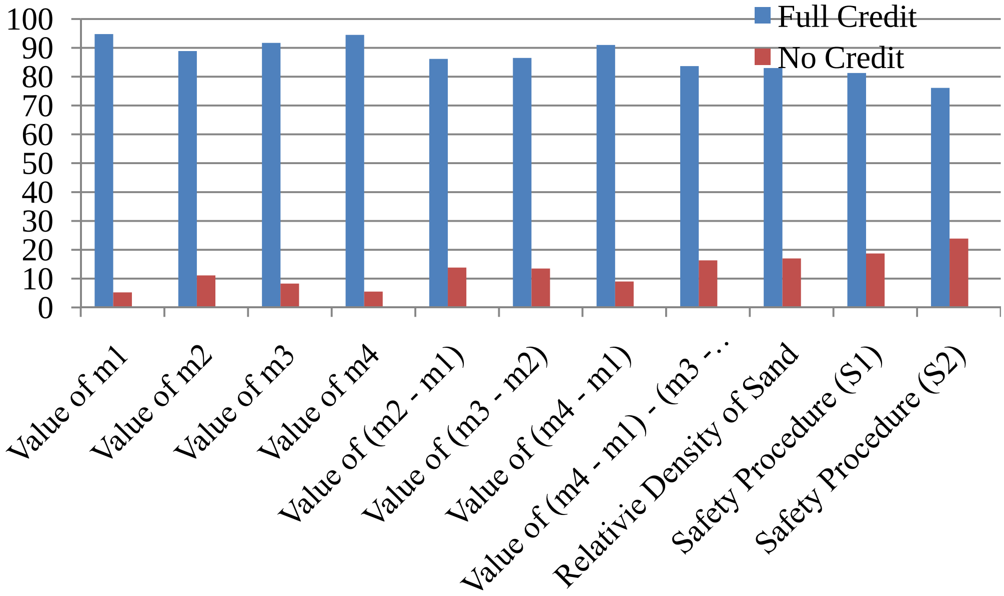

Figure 7: Distribution of Credits for Performing Skills (Task B)

Again, 257 students representing 88.9%; 250 students representing 86.5%; 249 students representing 86.2%; 242 students representing 83.7%; 240 students representing 83%; and 235 students representing 81.3%, out of the sample of size of 289 students had full credits for the values of m_2, $(m_3 - m_2)$, $(m_2 - m_1)$, $(m_4 - m_1) - (m_3 - m_2)$, relative density of sand and first safety procedure respectively. These results indicate that majority of the students showed adequate levels of proficiency in responding to items 1, 4, 3, 7, 2, 5, and 9 respectively. In all, about 220 students representing 76.1% of the sample had full credits on all the eleven items that were assessed. It can therefore be deduced from Table 12 and Figure 7 that majority of the respondents exhibited high levels of competency in the performing skills. On the other hand, 69 students representing 23.9% of the sample had difficulty stating the second safety procedure correctly.

Examples of students' incorrect measurements and calculations on the performing skills (Task B) are shown in a and b below. Few students involved in the research made this type of incorrect measurements and calculations.

Examples of Students' Incorrect Measurements and Calculations

a.

Mass of empty bottle (m_1)	Mass of bottle +Sand (m_2)	Mass of bottle+ Sand + Water (m_3)	Mass of Water only (m_4)	Relative Density (RD)
30.0g	70.0g	100.0g	70.0g	4

b.

Mass of empty bottle (m_1)	Mass of bottle +Sand (m_2)	Mass of bottle+ Sand + Water (m_3)	Mass of Water only (m_4)	Relative Density (RD)
20.0g	70.0g	105.0g	100.0g	1.8

Examples of students' correct measurements and calculations on the performing skills (Task B) are shown in c and d below. Majority of the respondents involved in the research had the measurements and calculations correct within the range of the marking scheme prepared by the researcher.

Examples of Students' Correct Measurements and Calculations

c.

Mass of empty bottle (m_1)	Mass of bottle +Sand (m_2)	Mass of bottle + Sand + Water (m_3)	Mass of Water only (m_4)	Relative Density (RD)
29.8g	65.0g	101.5g	81.2g	2.36

d.

Mass of empty bottle (m_1)	Mass of bottle +Sand (m_2)	Mass of bottle+ Sand + Water (m_3)	Mass of Water only (m_4)	Relative Density (RD)
30.0g	66.2g	101.5g	80.1g	2.45

Table 13 and Figure 8 show the distribution of total scores obtained by the students on the Performing Skills (Task B).

Table 13

Distribution of Total Scores for Performing Skills (Task B) by Sample

Total Score	Number of Students	Percentage (%)
0	0	0.0
1	5	1.7
2	5	1.7
3	8	2.8
4	5	1.7
5	10	3.5
6	11	3.8
7	3	1.0
8	1	0.3
9	10	3.5
10	68	23.5
11	163	56.4

N = 289.

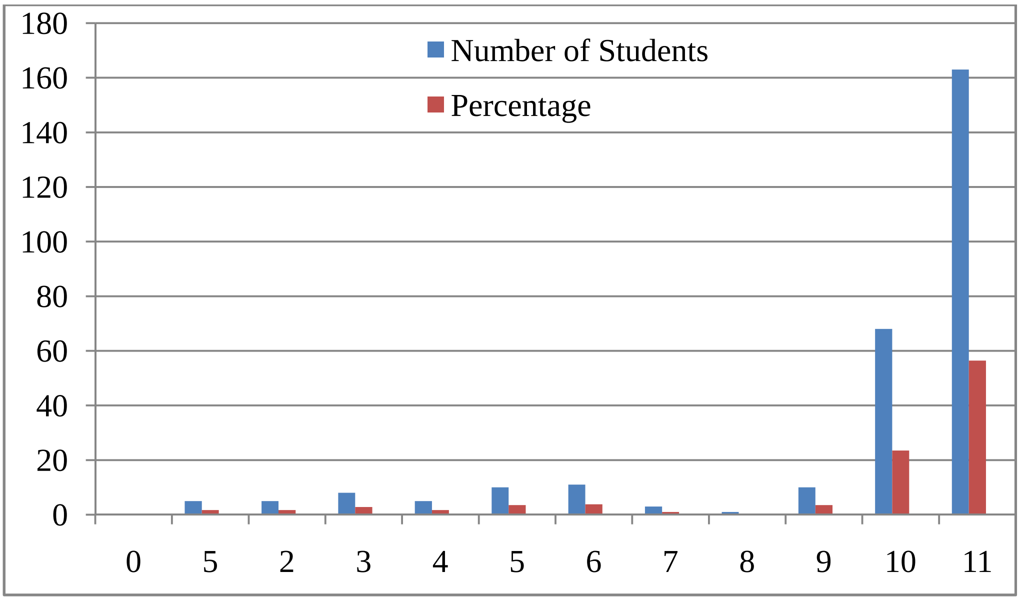

Figure 8: Distribution of Total Scores for Performing Skills (Task B) by Sample

One hundred and sixty three respondents representing 56.4% of the sample scored the total marks of 11 on the performing task as is shown in Table 13 and Figure 8. Again, 68 students representing 23.5% of the sample scored a total of 10 marks. In all 231 students representing 80.0% of the sample scored a total of 10 and 11 marks. These results indicate that many of the students exhibited excellent levels of proficiency in the performing skills (Task B). A total of 256 students representing 88.5% of the research sample scored 6 marks and above. It can therefore be inferred from Table 13 and Figure 8 that majority of the respondents in the research demonstrated high competency levels in laboratory performing skills. On the other hand, 33 students representing 11.4% of the research sample scored a total of 5 marks and below. This result suggests that few students exhibited low levels of competency in laboratory performing skills.

Reasoning Skills (Task C)

The reasoning skills (Task C) required the students to measure incident angles and the corresponding angles of deviation of rays of light passing through triangular glass prism; record their measurements and use the values to plot graphs; and answer questions based on the graphs drawn.

Table 14 and Figure 9 show distribution of credits for the reasoning skills (Task C). Sixteen items were assessed under the reasoning skills.

Two hundred and sixty six students representing 92%; 265 students representing 91.7%; and 264 students representing 91.3%, out of the research sample size of 289 students scored full credits on items 1, 3 and 2 respect-ively as shown in Table 14 and Figure 9. Again, 256 students representing

88.6%; 252 students representing 87.2%; 248 students representing 85.8%; 245 students representing 84.8%; and 242 students representing 83.7%, out of the research sample of 289 students scored full credits on items 4, 10, 11, 9 and 6 respectively. Furthermore, 242 students representing 83.7%, and 242 students representing 83.7% of the sample scored full credits on items 7 and 8 respectively.

Table 14

Distribution of Credits for Reasoning Skills (Task C)

Component Assessed[b]	Full Credit (%)	No Credit (%)
1. Pair values of i_2 and D_2	92.0	8.0
2. Pair values of i_3 and D_3	91.3	8.7
3. Pair values of i_4 and D_4	91.7	8.3
4. Pair values of i_5 and D_5	88.6	11.4
5. Pair values of i_6 and D_6	79.9	20.1
6. Horizontal axis labelled	83.7	16.3
7. Vertical axis labelled	83.7	16.3
8. Scale Chosen	83.7	16.3
9. i_1 and D_1 plotted	84.8	15.2
10. i_2 and D_2 plotted	87.2	12.8
11. i_3 and D_3 plotted	85.8	14.2
12. i_4 and D_4 plotted	81.3	18.7
13. i_5 and D_5 plotted	71.6	28.4
14. i_6 and D_6 plotted	82.7	17.3
15. Description of Curve	69.9	30.1
16. Deduced Relationship	69.2	30.8

N = 289.

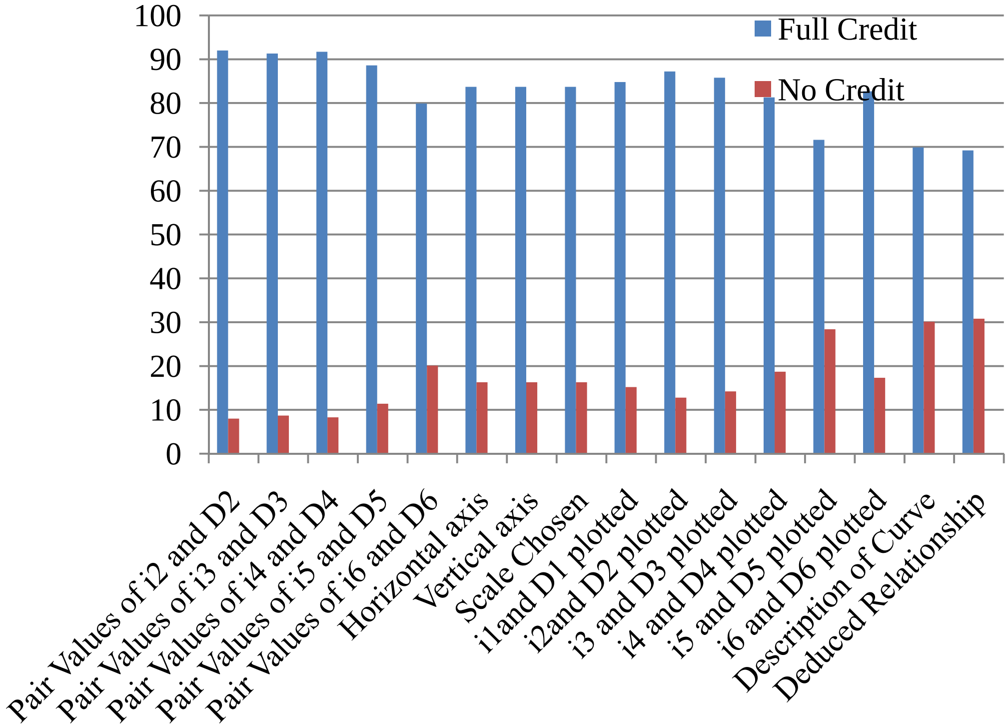

Figure 9: Distribution of Credits for Reasoning Skills (Task C)

Two hundred and thirty nine students representing 82.7% and 235 students representing 81.3% of the sample scored full credits on items 14 and 12 respectively. These results indicate that majority of the students exhibited adequate levels of proficiency in responding to items 1, 3, 2, 4, 10, 11, 9, 6, 7, 8, 14 and 12 respectively. About 200 students representing 69.2% of the sample scored full credits on all the 16 items assessed under the reasoning skills. It can therefore be inferred from the results that majority of the students showed high competency levels in laboratory reasoning skills in the research.

On the other hand, about 87 students representing 30.1% of the sample had difficulties in describing the shape of the curve drawn from their plots; and 89 students representing 30.8% of the sample could not deduce correct relationship between the incident angles and the angles of deviation. These results indicate that many of the students exhibited some inadequacies on the reasoning skills. It further indicates that many of the students need more practice on laboratory reasoning skills for them to improve on these inadequacies.

Table 15 and Figure 10 show distribution of total scores for Task C by sample.

Table 15

Distribution of Total Scores for Reasoning Skills (Task C) by Sample

Total Score	Number of Students	Percentage (%)
1	5	1.7
2	3	1.0
3	2	0.7
4	2	0.7
5	5	1.7
6	4	1.4
7	4	1.4
8	8	2.8
9	7	2.4
10	5	1.7
11	4	1.4
12	23	8.0
13	26	9.0
14	55	19.0
15	45	15.6
16	91	31.5

N = 289.

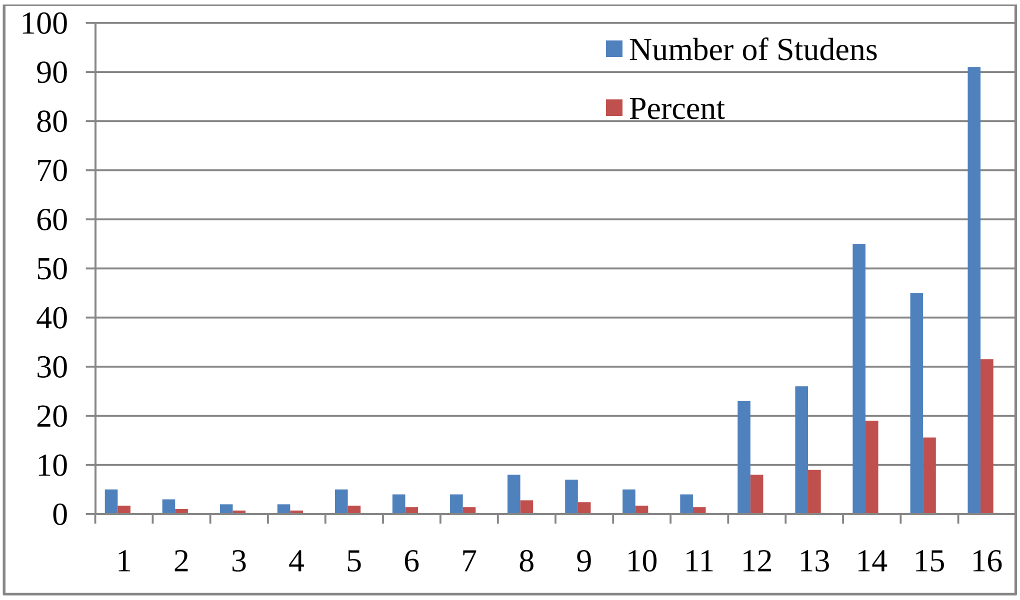

Figure 10: Distribution of Total Scores for Reasoning Skills (Task C) by Sample

Ninety one students representing 31.5% of the research sample scored the total marks of 16 as shown in Table 15 and Figure 10; while 45 students representing 15.6% of the sample scored 15 marks. In all 136 students representing 47.0% of the sample scored 15 and 16 marks. This indicates that many of the respondents exhibited excellent levels of proficiency in laboratory reasoning skills in the research. From Table 15, 256 students representing 88.6% of the research sample scored a total of 9 marks and above. It can therefore be deduced from Table 15 that majority of the students showed high proficiency levels in reasoning skills in the research. On the other hand 25 students representing 8.6% of the sample scored 7 marks and below. This shows that few of the students exhibited some inadequacies on the reasoning skills.

Total Performance Tasks

Table 16 shows the distribution of total scores obtained by the respondents on the total performance tasks.

Table 16

Distribution of Total Scores for Total Performance Tasks by Sample

Total Score	Number of Students	Percentage (%)
1 – 5	0	0.0
6	2	0.7
7	1	0.3
8	0	0.0
9	1	0.3
10	1	0.3
11	1	0.3
12	1	0.3
13	2	0.7
14	1	0.3
15	2	0.7
16	1	0.3
17	0	0.0
18	7	2.4
19	6	2.1
20	3	1.0
21	5	1.7
22	5	1.7
23	7	2.4
24	9	3.1
25	9	3.1
26	14	4.8

(Table 16 Continued).

27	9	3.1
28	13	4.5
29	23	8.0
30	24	8.3
31	30	10.4
32	28	9.7
33	36	12.5
34	28	9.7
35	20	6.9

N = 289.

Twenty students representing 6.9% of the research sample scored the total marks of 35 as shown in Table 16. Twenty eight students representing 9.7% of the sample scored 34 marks; while 36 students representing 12.5% of the sample scored 33 marks. In all, 84 students representing 29.1% of the sample scored 33 marks and above. These results show that many of the respondents exhibited excellent levels of proficiency in the total performance tasks. Moreover, 263 students representing 91.0% of the sample scored 18 marks and above. The median score is 17.5. It can be inferred from Table 16 that generally, majority of the students exhibited high levels of competency in laboratory planning, performing and reasoning skills in the research.

Research Question 2:

To what extent do senior high school physics students engaged in laboratory work exhibit adequate competencies in:

1. One aspect of the skills of planning, performing and reasoning?

2. All the three skills of planning, performing and reasoning?

Planning Skills (Task A)

From Table 9, the mean score obtained by the students on the planning skills is 5.8. This result represents 72.5% of the total marks of 8. It can therefore be deduced that generally the students exhibited high competency levels in laboratory planning skills in the main research. One hundred and eighty five out of the 289 students, representing 64.0% of the respondents had full credits on all the eight items assessed under the planning skills as shown in Table 10. Moreover, results from Table 11 show that 214 students out of the 289 respondents, representing 74.1% of the sample scored a total of 5 marks and above on the planning task. Results from Tables 10 and 11 therefore suggest that the students exhibited high levels of proficiency in planning skills in the research.

Performing Skills (Task B)

From Table 9, the mean score obtained by the students on the performing skills is 9.6. This result represents 87.3% of the total marks of 11. The result shows that on the whole the students exhibited high proficiencies in laboratory performing skills in the research. Two hundred and twenty students out of the sample of 289, representing 76.1% of the respondents scored full credits on all the 11 items that were assessed under the performing skills as is shown in Table 12. Again, results from Table 13 show that 256 students, representing 88.5% of the sample scored a total of 6 marks and above on the performing task. It can be inferred from the results in Tables 9, 12 and 13 that the students demonstrated high proficiencies in performing skills in the study.

Reasoning Skills (Task C)

From Table 9, the mean score obtained by the students on the reasoning skills is 13.3. This represents 83.1% of the total credits of 16. The result therefore indicates that the students demonstrated high levels of competency in laboratory reasoning skills in the research. Results from Table 14 show that about 200 students representing 69.2% of the sample had full credits on all the 16 items assessed under reasoning skills. Moreover, Table 15 shows that 256 students representing 88.6% of the sample scored a total of 9 marks and above on the reasoning task. It can therefore be deduced from the results in Tables 9, 14 and 15 that majority of the students exhibited high levels of competency in laboratory reasoning skills in the study.

Total Performance Task

From Table 9, the mean score obtained by the students on the total performance tasks is 28.7. This value represents 82.0% of the total marks of 35. The mean score therefore indicates that the students exhibited high levels of competency in the total performance tasks in the study. Two hundred and sixty three students out of the sample size of 289 students, representing 91.0% of the respondents scored a total of 18 marks and above as shown in Table 16. These results suggest that the students exhibited high levels of competency in the total performance tasks.

It can therefore be inferred from the results and discussions above that the students demonstrated high levels of competency in all the three laboratory skills of planning, performing and reasoning.

Research Question 3:

To what extent are male senior high schools physics students more skilful than their female counterparts in planning, performing and reasoning skills?

The scores of boys and girls from the performance tasks were compared and the results are as shown in Tables 17 through 20.

Planning Skills (Task A)

Table 17 shows the independent samples t-test for gender difference on the planning skills.

Table 17

Independent Samples t-test for Gender Difference on Planning Skills (Task A)

Variable	Group	N	Mean	SD	t	df	*p*
	Male	158	5.88	2.270	0.498	287	
Gender							0.619
	Female	131	5.74	2.480	-0.498	287	

Not Significant, $P > 0.05$

From Table 17, boys had a higher mean score of 5.88 as against a lower mean score of 5.74 for girls in the planning skills. This appears to indicate that boys outperformed girls in laboratory planning skills in the research. Again, the standard deviation for boys was 2.270 while the standard deviation for girls was 2.480. This suggests that scores obtained by the girls on the planning skills were much spread out than the scores of boys.

From the t-test for equality of means shown on Table 17, there is a difference of 0.14 in the mean scores for boys and girls. However, the *P*-value of 0.619 which is greater than the statistically significant level of 0.05, suggests that the difference in mean scores for boys and girls is not statistically significant. It can therefore be deduced from table 17 that both boys and girls exhibited similar levels of competency in laboratory planning skills in the study. (M=5.74, SD=2.48), t(287) = 0.498, $P > 0.05$ (2-tailed), d=0.14.

Performing Skills (Task B)

Table 18 shows the independent samples t-test for gender difference on the performing skills.

Table 18

Independent Samples t-test for Gender Difference on Performing Skills (Task B)

Variable	Group	N	Mean	SD	t	df	*p*
	Male	158	9.33	2.876	-1.802	287	
Gender							0.073
	Female	131	9.87	2.066	1.802	287	

Not Significant, $P > 0.05$

From Table 18 girls had a higher mean score of 9.87 as against a lower mean score of 9.33 for boys in the performing skills. The mean scores apparently suggest that girls outperformed boys in laboratory performing skills. A possible reason for the apparent better performance of girls than boys on the performing skills may be due to the relatively smaller sample of girls in the research sample. Again, girls had a lower standard deviation of 2.066 as against a higher standard deviation of 2.876 for boys. This shows that scores obtained by boys were much spread out than scores obtained by girls.

From the independent samples t-test for equality of means, there is a difference of -0.54 for the mean scores of boys and girls. However, the *P*-value of 0.073 which is greater than the statistically significant level of 0.05 shows that the difference in the performing skills exhibited by boys and girls is not statistically significant. This further suggests that both boys and girls demonstrated similar levels of competency in laboratory performing skills in the research.

(M = 9.33, SD = 2.88), t(281.524) = -1.857, $p > 0.05$ (2-tailed), d = -0.54.

Reasoning Skills (Task C)

Table 19 shows the independent samples t-test for gender on the reasoning skills.

Table 19
Independent Samples t-test for Gender Difference on Reasoning Skills (Task C)

Variable	Group	N	Mean	SD	t	df	*p*
	Male	158	13.05	3.867	-1.225	287	
Gender							0.221
	Female	131	13.56	2.995	1.225	287	

Not Significant, $P > 0.05$

From Table 19 girls had a higher mean score of 13.56 as against a lower mean score of 13.05 for boys in the reasoning skills. These results apparently suggest that girls outperformed boys in laboratory reasoning skills. Again, girls had a lower standard deviation of 2.995 as against a higher standard deviation of 3.867 for boys. These values indicate that scores obtained by boys were much spread out than scores obtained by girls.

From the independent samples t-test on Table 19, there is a difference of -0.51 in the mean scores of boys and girls. The *P*-value of 0.221, however, is greater than the statistically significant level of 0.05, and it can be inferred from Table 19 that there is no statistically significant difference in the reasoning skills of boys and girls in the research. Thus boys and girls exhibited the same proficiency levels in laboratory reasoning skills in the study. (M = 13.05, SD = 3.87), t(287) = -1.225, $p > 0.05$ (2-tailed), d = -0.51.

Total Performance Tasks

Table 20 shows the independent samples t-test for gender on the total performance tasks.

Table 20

Independent Samples t-test for Gender Difference on Total Performance Tasks

Variable	Group	N	Mean	SD	t	df	*p*
Gender	Male	158	28.27	6.133	-1.313	287	0.190
	Female	131	29.17	5.309	1.313	287	

$P > 0.05$

From Table 20 girls had a higher mean score of 29.17 as against a lower mean score of 28.27 for boys in the reasoning skills. The values appear to suggest that girls outperformed boys in the total performance tasks. Again, girls had a lower standard deviation of 5.309 as against a higher standard deviation of

6.133 for boys. This shows that the scores obtained by boys were much spread out than scores obtained by girls.

Table 20 shows that there is a difference in mean scores of -0.90 between boys and girls. However, the *P*-value of 0.190 is greater than the statistically significant level of 0.05. This indicates that the difference in the performance of boys and girls is not statistically significant. It can be inferred from Table 20 that both boys and girls exhibited the same competency levels in laboratory performance skills in the research.

Research Question 4:

To what extent are the skills of planning, performing and reasoning demonstrated by a student related to the type of school (boys, girls or mixed) the student attends?

Planning Skills (Task A)

Table 21 shows the means and standard deviations for the planning skills by type of school.

Table 21

Means and Standard Deviations for Planning Skills (Task A) by Type of School

Type of School	n	Mean	Standard Deviation
Boys	80	5.83	2.343
Girls	40	5.88	2.493
Mixed	169	5.80	2.357
Total	289	5.82	2.364

From Table 21, students from the girls school had the highest mean of 5.88 followed by students from the boys schools with a mean of 5.83, while students from the mixed schools had the lowest mean among the three types of school with a score of 5.80. These results apparently suggest that there were differences in planning skills exhibited by students from the three different types of schools in the research. Students from the girls school appeared to show the highest level of planning skills followed by students from the boys schools; while students from the mixed schools appeared to show the lowest planning skills.

Table 22 shows ANOVA for types of schools for the planning skills.

Table 22

ANOVA for Planning Skills (Task A) by Types of Schools

Source	df	Sum of Square	Mean Square	F	*P*
Between Groups	2	0.196	0.098	0.017	0.983
Within Groups	286	1609.1	5.626		
Total	288	1609.3			

Not Significant, *P* > 0.05

The *P*-value of 0.983 for F(2, 286) from the analysis of variance is greater than the statistically significant level of 0.05. This shows that the differences in the mean scores of the three types of schools are not statistically significant. It can therefore be inferred from Table 22 that students from the three types of schools (boys, girls and mixed) exhibited similar levels of proficiency in laboratory planning skills in the study. $F(2, 286) = 0.017$, $p > 0.05$.

Performing Skills (Task B)

Table 23 shows the means and standard deviations for the three different types of schools on the performing skills (Task B).

Table 23

Means and Standard Deviations for Performing Skills (Task B) by Type of School

Type of School	n	Mean	Standard Deviation
Boys	80	9.83	2.305
Girls	40	9.68	2.443
Mixed	169	9.43	2.688
Total	289	9.57	2.551

From Table 23, students from the boys schools had the highest mean score of 9.83, followed by students from the girls school with a mean score of 9.68, while students from the mixed schools had the lowest mean score of 9.43 among the three different types of schools. These mean scores apparently suggest that there were differences in levels of performing skills demonstrated by students from the different schools involved in the study. It further appears to indicate that students from the boys schools exhibited the highest level of performing skills in the research; and they were followed by students from the girls school; while students from the mixed schools appeared to have exhibited the lowest level of performing skills.

Table 24 shows ANOVA for the different types of schools on the performing skills. The *P*-value of 0.508 for F(2, 286) from the analysis of variance is greater than the statistically significant level of 0.05. This indicates that the differences in the mean scores of students from the three types of schools are not statistically significant. It can be deduced from the results that students from the three types of schools (boys, girls and mixed) exhibited the same levels of proficiency in performing skills in the study. $F(2, 286) = 0.679$, $p > 0.05$.

Table 24

ANOVA for Performing Skills (Task B) by Types of Schools

Source	df	Sum of Square	Mean Square	F	*P*
Between Groups	2	8.858	4.429	0.679	0.508
Within Groups	286	1865.8	6.524		
Total	288	1874.7			

Not Significant, $P > 0.05$

Reasoning Skills

Table 25 shows the means and standard deviations for students in the three different types of schools on the reasoning skills (Task C).

Table 25

Means and Standard Deviations for Reasoning Skills (Task C) by Type of School

Type of School	n	Mean	Standard Deviation
Boys	80	13.54	3.628
Girls	40	14.35	2.558
Mixed	169	12.91	3.558
Total	289	13.28	3.502

N = 289.

From Table 25 students in the girls school had the highest mean score of 14.35, followed by students in the boys schools with a mean score of 13.54, while students in the mixed schools had the lowest mean score of 12.91 among the three different types of schools. These mean scores apparently show that there were differences in levels of reasoning skills demonstrated by students in the different types schools used for the study. The results further appear to suggest that students in the girls school exhibited the highest level of reasoning skills in the research; and they were followed by students in the boys school; while students in the mixed schools appeared to have exhibited the lowest level of reasoning skills.

Table 26 shows ANOVA for students in the different types of schools on the reasoning skills.

Table 26

ANOVA for Reasoning Skills (Task C) by Types of Schools

Source	df	Sum of Square	Mean Square	F	*P*
Between Groups	2	74.825	37.412	3.095	0.047*
Within Groups	286	3457.5	12.089		
Total	288	3532.3			

*Significant, $P < 0.05$

The P-value of 0.047 for F(2, 286) from the analysis of variance is less than the statistically significant level of 0.05. This indicates that the differences in the mean scores of students in the three types of schools are statistically significant. It can therefore be inferred from Table 26 that students in the three types of schools (boys, girls and mixed) exhibited different levels of proficiency in reasoning skills in the study. $F(2, 286) = 3.095$, $p < 0.05$.

Tukey HSD post-hoc tests were done for the pair wise comparison of means.

Table 27 shows Tukey HSD pair wise comparison of means between the types of schools.

Table 27

Post Hoc Tests on Reasoning Skills (Task C) by Types of Schools

Type of School (I)	Type of School (J)	Means Difference (I – J)	*p*
Boys	Girls	-0.8125	0.450
Boys	Mixed	0.6322	0.374
Girls	Boys	0.8125	0.450
Girls	Mixed	1.4447*	0.049
Mixed	Boys	-0.6322	0.374
Mixed	Girls	-1.4447*	0.049

N = 289. * The mean difference is significant at the 0.05 level.

From Table 31, the mean difference between students in the boys and girls school is -0.8125 and the *p*-value is 0.450. This value is greater than the 0.05 level for statistical significance, and suggests that the difference between the mean scores of students in boys and girls schools is not statistically significant. It can be deduced from Table 27 that students in both boys and girls schools demonstrated similar levels of proficiency in reasoning skills in the research. Again, the difference in mean scores between students in boys and mixed schools is 0.6322 and the *p*-value is 0.374. Since this result is greater than the 0.05 level for statistical significance, the difference in mean scores between students in boys and mixed schools is not statistical significant. It can therefore be inferred from Table 27 that students in both boys and mixed schools exhibited the same levels of competency in reasoning skills in the study. The difference in mean scores between students in girls and mixed schools is 1.4447 and the *p*-value is 0.049. Since this value is less than the statistically significant level of 0.05, the difference in mean scores between the students in girls and mixed schools is statistically significant. It can be deduced from Table 27 that students in the girls school exhibited higher levels of competency in reasoning skills than students from the mixed schools.

Total Performance Tasks

Table 28 shows the mean scores and standard deviations for students in the three different types of schools on the total performance tasks.

Table 28

Means and Standard Deviations for Types of Schools on Total Performance Tasks

Type of School	n	Mean	Standard Deviation
Boys	80	29.21	5.474
Girls	40	29.90	5.118
Mixed	169	28.14	6.029
Total	289	28.68	5.782

From Table 28 students in the girls school obtained the highest mean score of 29.90. Students in the boys schools had a mean score of 29.21, while students in the mixed schools obtained the lowest mean score of 28.14 among the three different types of schools. This appears to indicate that students in the girls school demonstrated the highest proficiencies in laboratory skills, while students in the mixed schools exhibited the lowest levels of proficiency in laboratory skills among the three types of schools used for the study. Again, the standard deviation for students in the girls school is the lowest with a value of 5.118, while the standard deviation for students in the mixed schools is the highest with a value of 6.029. This shows that scores obtained by students in the mixed schools were much more spread out than scores of students in the boys schools; which in turn are much spread out than scores of students in the girls school.

Table 29 shows ANOVA for students in the three different types of schools on the total performance tasks.

Table 29

ANOVA for Total Performance Tasks by Types of Schools

Source	df	Sum of Square	Mean Square	F	*P*
Between Groups	2	1322.2	66.108	1.991	0.138
Within Groups	286	9494.9	33.199		
Total	288	9627.1			

$P > 0.05$

The *p*-value from the analysis of variance is 0.138, which is greater than the statistically significant level of 0.05. This shows that the differences in the mean scores of students in the girls, boys and mixed schools are not statistically significant. It can therefore be inferred from Table 29 that generally, students in the girls, boys and mixed schools demonstrated similar levels of competency on the total performance tasks in the research.

Students' Comments on the Performance Tasks

Responses of students to the items on the opinionnaire were coded and input into SPSS programme and analyzed.

Table 30 and Figure 11 show responses of the respondents on the time provided for completing the Planning Skills (Task A).

Table 30

Distribution of Students' Responses on Time Provided for Planning Skills (Task A)

Response	Number of Students	Percentage (%)
No Response	7	2.4
Not Enough	32	11.1
About Right	227	78.6
Too Much	23	8.0

N = 289.

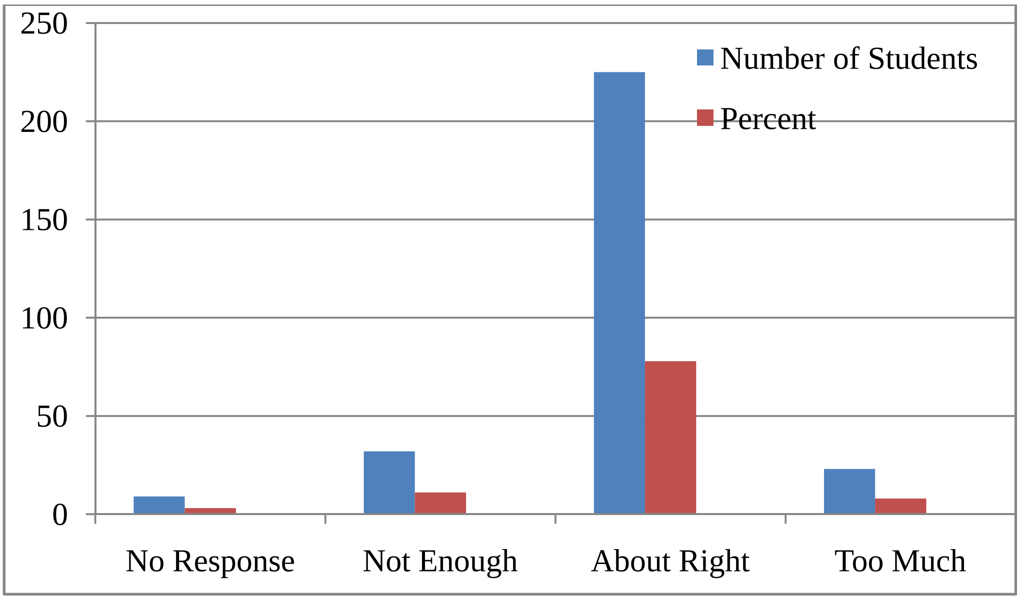

Figure 11: Distribution of Students Responses on Time provided for Planning Skills (Task A)

From Table 30 and Figure 11, majority of the students, 227 out of the 289 respondents, representing 78.6% of the research sample agreed that the time provided for completing Task A was about right. Twenty three students representing 8.0% of the sample said that the time provided for completing Task A was too much. In all, 250 out of the 289 respondents representing 86.6% of the students were of the view that time provided for completing Task

A was sufficient for them. Only 32 respondents representing 11.1% of the sample said that time provided for completing Task A was not enough, while another seven students representing 2.4% of the sample did not respond to the item.

Table 31 and Figure 12 show the students' responses on the difficulty level of the Planning Skills (Task A).

Table 31

Distribution of Students' Responses on Difficulty Level of Planning Skills (Task A)

Response	Number of Student	Percentage (%)
No Response	14	4.8
Easy	78	27.0
About Right	181	62.6
Difficult	16	5.5

N = 289.

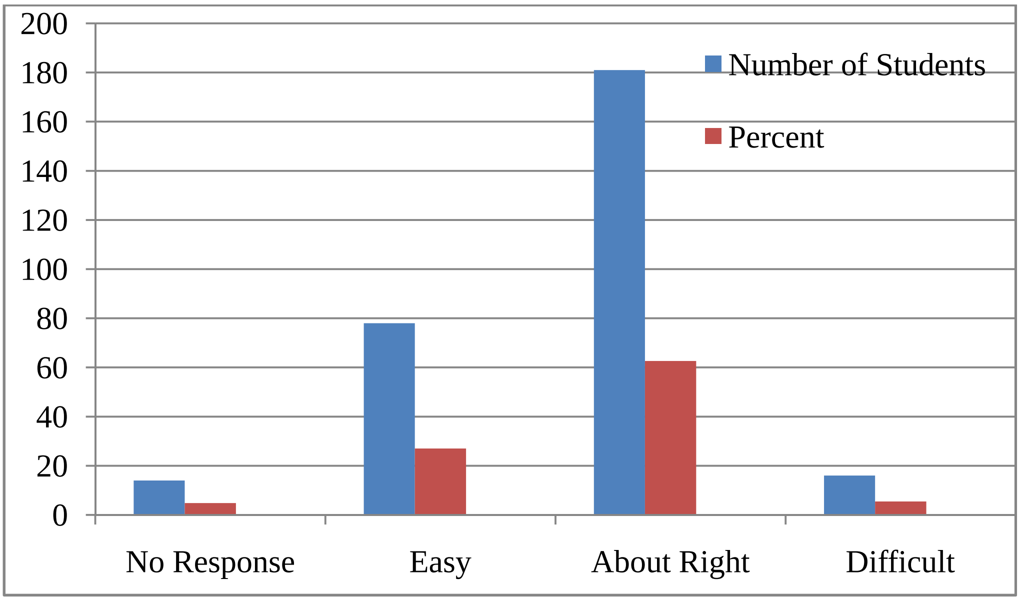

Figure 12: Distribution of Students Responses on Difficulty Level of Planning Skills (Task A)

From Table 31 and Figure 12, majority of the students, 181 out of the 289 respondents, representing 62.6% of the research sample said the difficulty level of the Planning Skills (Task A) was about right. Seventy eight students representing 27.0% of the sample said Task A was easy. In total, 259 out of the 289 students representing 89.6% of the sample agreed that the difficulty of Task A was within their reach. Only 16 students representing 5.5% of the sample said that Task A was difficult, while 14 students representing 4.8% of the sample did not respond to the item.

Table 32 and Figure 13 show responses of the students on the time provided for completing the Performing Skills (Task B).

Table 32

Distribution of Students' Responses on Time Provided for Performing Skills (Task B)

Response	Number of Students	Percentage (%)
No Response	11	3.8
Not Enough	43	14.9
About Right	193	66.8
Too Much	42	14.5

N = 289.

From Table 32 and Figure 13, majority of the students, 193 out of the sample of 289 students, representing 66.8% of the research sample said the time provided for completing the Performing Skills (Task B) was about right. Another 42 students representing 14.5% of the sample said time provided for completing the performing skills was too much. Two hundred and thirty five out of the (289), representing 81.3% of the respondents agreed that time provided for completing the task was sufficient. Only 43 students representing

14.9% of the sample said the time provided for completing Task B was not enough, while another 11 students representing 3.8% of the sample did not respond to the item.

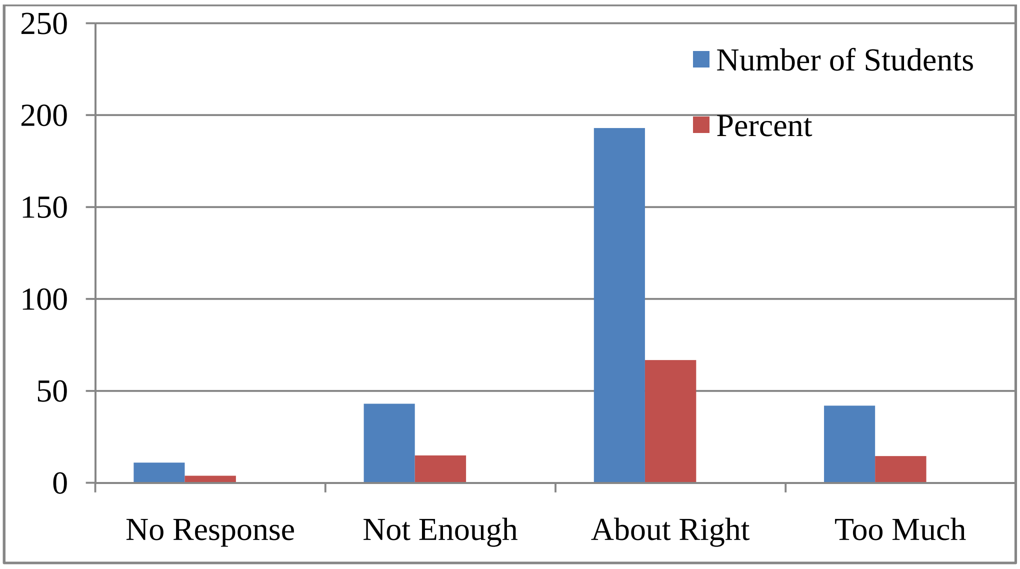

Figure 13: Distribution of Students' Responses on Time Provided for Performing Skills (Task B)

Table 33 and Figure 14 show responses of the students on difficulty of the Performing Skills (Task B).

Table 33

Distribution of Students' Responses on Difficulty Level of Performing Skills (Task B)

Response	Number of Students	Percentage (%)
No Response	13	4.5
Easy	122	42.2
About Right	148	51.2
Difficult	6	2.1

N = 289.

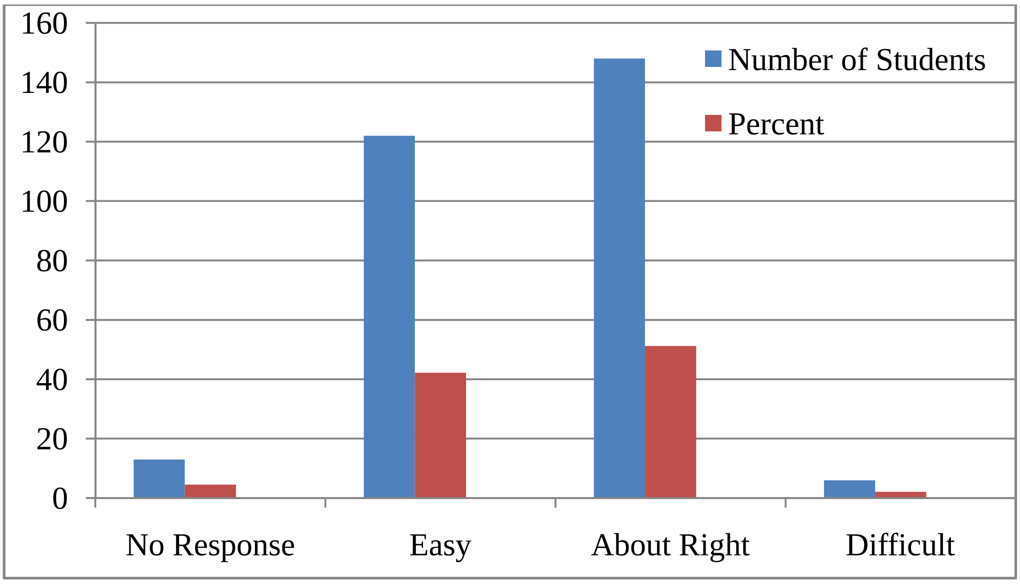

Figure 14: Distribution of Responses of Students on Difficulty of Performing Skills (Task B)

From Table 33 and Figure 14, majority of the respondents, 148 out of the 289 students, representing 51.2% of the research sample said the difficulty level of the Performing Skills was about right. One hundred and twenty two students representing 42.2% of the sample said Task B was easy. In total, 270 out of the 289 students, representing 93.4% of the sample agreed that the difficulty level of the task was within their reach. Only six students representing 2.1% of the sample said that Task B was difficult, while another 13 students representing 4.5% of the sample did not respond to the item.

Table 34 and Figure 15 show the distribution of the students' responses on the time provided for completing the reasoning task.

From Table 34 and Figure 16, 207 students out of the 289 research sample, representing 71.6% of the respondents said the time provided for completing Task C was about right. Thirty five students representing 12.1% of the respondents said the time provided for Task C was too much. Table 34 and Figure 15 show that a total of 242 students out of the sample of 289 students, representing 83.7% of the sample said the time provided for completing the reasoning skills was sufficient. On the other hand 36 respondents representing 12.5% of the sample said the time provided for completing Task C was not enough, while another 11 students representing 3.8% of the sample did not respond to the item.

Table 34

Distribution of Students' Responses on Time Provided for Reasoning Skills (Task C)

Response	Number of Students	Percentage (%)
No Response	11	3.8
Not Enough	36	12.5
About Right	207	71.6
Too Much	35	12.1

N = 289

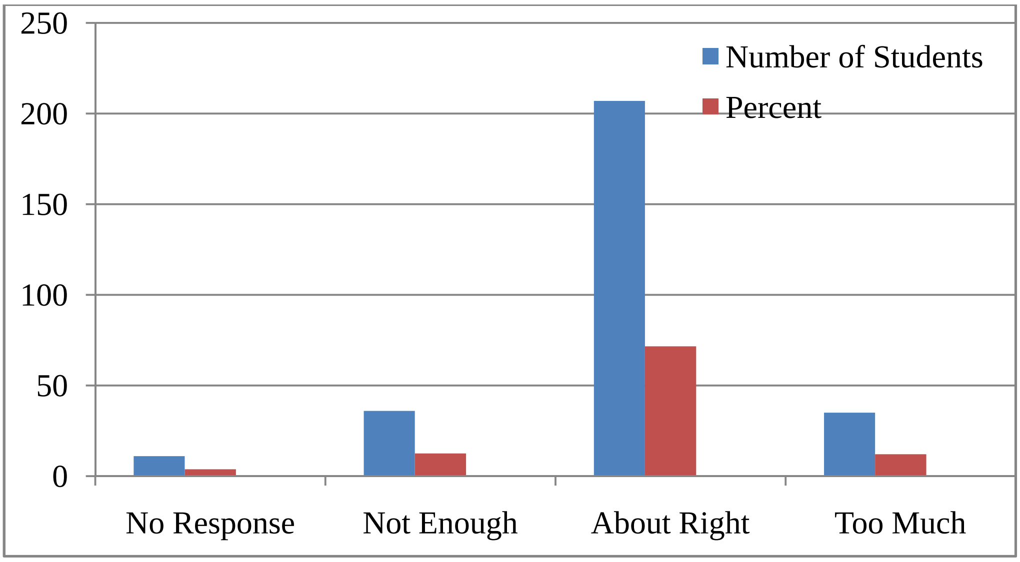

Figure 15: Distribution of Students' Responses on Time Provided for Reasoning Skills (Task C)

Table 35 and Figure 16 show distribution of the students' responses on the difficulty level of the reasoning skills (Task C).

Table 35

Distribution of Students' Responses on Difficulty Level of Reasoning Skills (Task C)

Response	Number of Students	Percentage (%)
No Response	7	2.4
Easy	113	39.1
About Right	161	55.7
Difficult	8	2.8

N = 289.

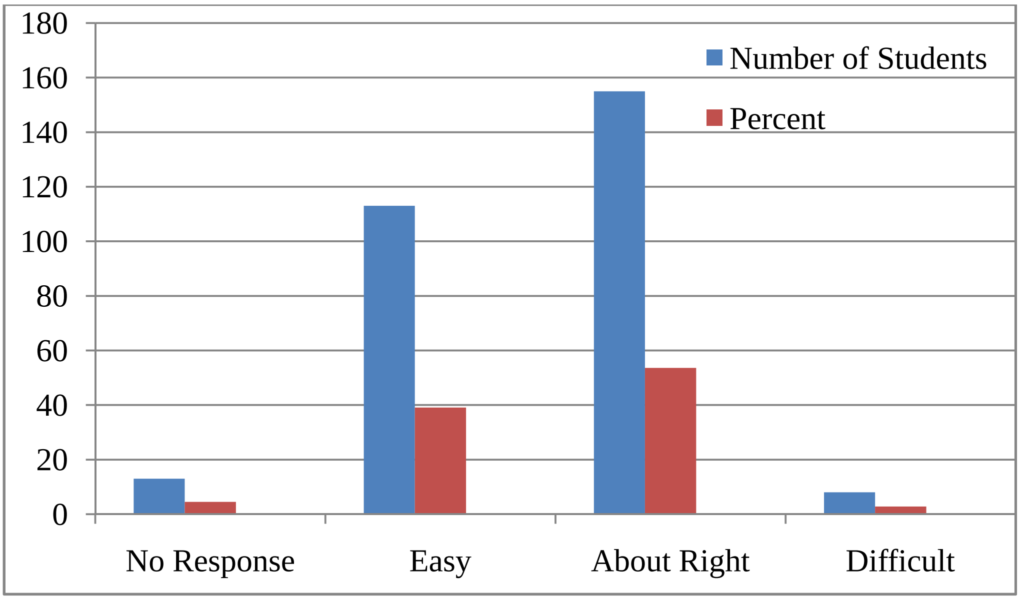

Figure 16: Distribution of Students' Responses on Difficulty of Reasoning Skills (Task C)

One hundred and sixty one out of the sample of 289 students, representing 55.7% of the respondents said the difficulty level of the reasoning skills was about right as shown in Table 35 and Figure 16. One hundred and thirteen students representing 39.1% of the sample said task C was easy. In total, 274 out of the research sample of 289 students, representing 94.8% of the respondents agreed that the difficulty of the reasoning skills was within their reach. Only eight students representing 2.8% of the sample said that Task C was difficult, while another seven students representing 2.4% of the sample did not provide any response to the item.

Summary of Findings

The summary of the findings are as follows:

1. The students demonstrated high levels of competency in laboratory planning skills in the selected topics in mechanics and optics.

2. The students demonstrated high competency levels in laboratory performing skills in the selected topics in mechanics and optics.

3. The students exhibited high proficiencies in laboratory reasoning skills in the selected topics in mechanics and optics.

4. The students demonstrated high levels of competency in all three laboratory skills of planning, performing and reasoning.

5. The students demonstrated the highest level of proficiency on the performing skills, followed by the reasoning skills; while levels of proficiency shown on the planning skills were the least among the three.

6. There was no significant difference in the levels of competency in laboratory planning skills demonstrated by the male and female physics students.

7. There was no significant difference in the levels of proficiency in laboratory performing skills demonstrated by the male and female physics students.

8. There was no significant difference in levels of competency in laboratory reasoning skills exhibited by the male and female physics students.

9. There was no significant difference in levels of proficiency shown in laboratory planning skills by students in boys and girls schools.

10. There was no significant difference in levels of proficiency shown in laboratory planning skills by students in boys and mixed schools.

11. There was no significant difference in laboratory planning skills exhibited by students in girls and mixed schools.

12. There was no significant difference in levels of competency in laboratory performing skills exhibited by students in boys and girls schools.

13. There was no significant difference in competencies demonstrated in laboratory performing skills by students in boys and mixed schools.

14. There was no significant difference in competencies in laboratory performing skills demonstrated by students in girls and mixed schools.

15. There was no significant difference in levels of proficiency in laboratory reasoning skills demonstrated by students in boys and girls schools.

16. There was no significant difference in levels of competency in laboratory reasoning skills shown by students in boys and mixed schools.

17. Students in girls school demonstrated higher levels of competency in laboratory reasoning skills than students in mixed schools.

18. The time provided for completing the Planning Skills (Task A) was sufficient.

19. The difficulty of the planning skills (Task A) was within the reach of the students.

20. Time provided for completing the Performing Skills (Task B) was sufficient.

21. The difficulty of the performing skills (Task B) was within the reach of the students.

22. Time provided for completing the Reasoning Skills (Task C) was sufficient.

23. The difficulty of the reasoning skills (Task C) was within the reach of the students.

CHAPTER FIVE

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

The previous four chapters of this study treated the background of the study, relevant literature reviewed, methodology, results and major findings of the research. This chapter states the conclusions from the results and major findings; and states recommendations based on these findings to educational practice. Suggestions for further studies have also been stated in this chapter.

Overview of the Problem

The research was geared towards assessing the laboratory skills of senior high school (SHS) physics students in selected topics in mechanics and optics. As a result of that, the study set out to find whether senior high school physics students engaged in laboratory activities demonstrate adequate laboratory planning, performing and reasoning skills. It was hoped that identification of at least minimum laboratory (performance) skills from the performances of the students, could inform and provide the basis for science teachers and educators to improve upon the methods of science instruction and assessment; as well as to parents and administrators to provide adequate resources for science education.

Overview of Research Methodology

The targeted population for the study was all SHS 3 physics students in Sekondi-Takoradi metropolis of the Western Region of Ghana. Due to the enormity of work involved, however, the accessible population for the research was students in seven senior high schools in Sekondi-Takoradi metropolis. The researcher obtained formal permission from the headmasters, headmistresses, heads of science department, and physics teachers of the schools before the study was conducted. A sample of 289 students were drawn from the seven senior schools within Sekondi-Takoradi metropolis for the main research; while a sample of 30 students were drawn from three schools in Cape Coast metropolis for the pilot study. Simple random and stratified random sampling methods were used to select the students for the study. The respondents were drawn from boys, girls and mixed schools that offer physics as a subject to students. The sample comprised of 131 female, and 158 male students.

In order to find out whether SHS 3 physics students could demonstrate adequate laboratory planning, performing and reasoning skills, four research questions were put forth as follows:

1. To what extent do senior high school physics students engaged in laboratory work exhibit adequate competencies in the skills of:

i. Planning?

ii. Performing?

iii. Reasoning?

2. To what extent do senior high school physics students engaged in laboratory work exhibit adequate competencies in:

i. One aspect of the skills of planning, performing and reasoning?

ii. All the three aspects of planning, performing and reasoning?

3. To what extent are male senior high school physics students more skilful than their female counterparts?

4. To what extent are the skills of planning, performing and reasoning exhibited by a student related to the type of school (boys, girls or mixed) the student attends?

In addition to the tasks, an opinionnaire was used to sought the respondents opinions, to find out whether their performances were affected by the time provided for completing the tasks; the difficulty level of the tasks; and the time they spent every two weeks doing physics laboratory work in their schools, as well any suggestions they could give to help modify the tasks.

Responses of the students to the tasks were hand-scored by the researcher. Another physics teacher was given 111 subsamples of the booklets of the students' responses to score. This helped the researcher to estimate the inter-rater reliabilities of the instruments (tasks). Statistical procedures were used to analyse the scores. The inter-rater-reliabilities for tasks were: 0.93 for the planning skills; 0.96 for the performing skills; 0.94 for the reasoning skills; and the total tasks had reliability of 0.93. Again, Cronbach alpha reliabilities of the tasks were: 0.83 for the planning skills; 0.90 for the performing skills; 0.88 for the reasoning skills and the total tasks had alpha reliability of 0.76. Moreover, inter tasks correlations were determined as follows: 0.073 between Task A and Task B; 0.015 between Task A and Task C; and 0.000 between Task B and Task C.

Means, standard deviations, frequencies, and percentages were used to find the levels of proficiencies of laboratory planning, performing and reasoning skills demonstrated by the respondents in the research. The data was also subjected to t-tests, ANOVAs and post-hoc analysis to find out whether there were significant differences in laboratory skills exhibited by the students due gender differences; and due to differences in type of schools (boys, girls or mixed) they attended.

Summary of Results

The planning skills (Task A) had a mean score of 5.8 and standard deviation of 2.36; the performing skills (Task B) had a mean score of 9.6 and a standard deviation of 2.55; the reasoning skills had mean score of 13.3 and a standard deviation of 3.50; while the total performance tasks had a mean score of 28.7 and a standard deviation of 5.78. The mean score of 5.8 for the planning skills represented an achievement of 72.5% out of the total marks of 8. The mean score of 9.6 for the performing skills represented an achievement of 87.3% out of the total credits of 11. The mean score of 13.3 for the reasoning skills represented an attainment of 83.1% out of the total credits of 16; while the mean score of 28.7 for the total performance tasks represented an attainment of 82.0% out of the total credits of 35. These results showed that majority of the respondents demonstrated high levels of laboratory planning, performing, reasoning skills and in the total tasks. The results further showed that the students demonstrated the highest level of proficiency in the performing skills (Task B); followed by the reasoning skills (Task C); while proficiency level in the planning skills (Task A) was the least. The results also showed that majority of the students exhibited high proficiencies in all the three aspects of laboratory planning, performing and reasoning skills.

The male students had a higher mean score of 5.88 and lower standard deviation of 2.270; as against a lower mean score of 5.74 and a higher standard deviation of 2.480 for the female students on the planning skills (Task A). These results apparently showed that boys performed better than girls on the planning skills. However, the p-value of 0.619 from the t-test for equality of means was greater than the 0.05 level of statistical significance. This indicated that the difference in the mean scores of the male and female students was not statistically significant. Thus both boys and girls demonstrated similar levels of proficiency in laboratory planning skills.

The female students had a higher mean score of 9.87 and a lower standard deviation of 2.066; as against a lower mean score of 9.33 and a higher standard deviation of 2.876 for the male students on the performing skills (Task B). These results appeared to show that the female students performed better than the male students on the performing skills. However, after the means were subjected to t-test analysis, the p-value of 0.064 was found to be greater than 0.05 level of statistical significance. Thus it came out that the difference in the mean scores was not statistically significant. Hence both boys and girls showed similar levels of competencies in the performing skills.

Again, the female students had a higher mean score of 13.56 and a lower standard deviation of 2.995; as against a lower mean score of 13.05 and higher standard deviation of 3.867 for the male students on the reasoning skills. These values apparently showed that girls performed better than boys on the reasoning skills. However, the p-value of 0.221 from the t-test was greater than the statistically significant level of 0.05. Therefore the difference in the mean scores was not statistically significant. Thus both male and female students demonstrated similar levels of competency on the reasoning skills. Moreover, girls had a higher mean score of 29.17 and a lower standard deviation of 5.309; as against a lower mean score of 28.27 and a higher standard deviation of 6.133 for the male students on the total performance tasks. This appeared to indicate that girls performed better than boys on the total performance tasks. However, after the mean scores were subjected to t-test analysis it was found that the p-value of 0.184 was greater than the 0.05 level of statistical significance. Thus the difference in the mean scores for boys and girls was not statistically significant. Hence, both the female and male students exhibited the same levels of proficiency on the total performance tasks.

Students from the girls school had the highest mean score of 5.88 on the planning skills (Task A); and they were followed by students from the boys schools with a mean score of 5.83; while students from the mixed schools had the lowest mean score of 5.80. These scores appeared to show that there were differences in the performances of the students as a result of the type of school they attended. However, when the mean scores were subjected to analysis of variance, the p-value of 0.983 was found to be greater than the statistically significant level of 0.05. This suggested that the differences in the mean scores of the students were not statistically significant. Therefore the level of proficiencies in laboratory planning skills demonstrated by the students in the research did not depend on the type of school they attended.

Students from the boys schools had the highest mean score of 9.83 on the performing skills (Task B); and they were followed by students from the girls school with a mean score of 9.68; while students from the mixed schools had the lowest mean score of 9.43 on the performing skills. These scores apparently suggested that there were differences in the proficiencies of performing skills demonstrated by the students as a result of differences in the types of schools they attended. However, the p-value of 0.508 from the analysis of variance indicated that the differences in the mean scores of the students were not statistically significant. Therefore the level of competencies in laboratory performing skills exhibited by the students in the research did not depend on the type of school they attend.

Students from the girls school had the highest mean score of 14.35 on the reasoning skills (Task C). They were followed by students from the boys schools with a mean score of 13.54; while students from the mixed schools had the lowest mean score of 12.91 on the reasoning skills. These scores showed that there were differences in laboratory reasoning skills exhibited by the students as a result of the type of school they attend. When the mean scores were subjected to analysis of variance, the p-value of 0.047 which was less than the statistically significant level of 0.05 did indicate that the differences in the mean scores were statistically significant. The mean scores were therefore subjected to post-hoc analysis to find the pair wise comparison of the means. It was found out that there was no statistically significant difference between the means scores of students from the boys and girls schools. Likewise there was no statistically significant difference between the mean scores of students from the boys and mixed schools. These results indicated that students from boys and girls schools demonstrated similar levels of competency in laboratory reasoning skills. Likewise students from boys and mixed schools exhibited similar levels of competency in reasoning skills. However, the post-hoc analysis did show that there was significant difference between the mean scores of students from the girls and mixed schools. Thus students from the girls school demonstrated higher levels of competency in laboratory reasoning skills than students from the mixed schools.

The students also showed some major inadequacies on the performance tasks. On the planning skills, many of the students had difficulty in choosing the relevant materials/equipment from the list provided to produce their plans. Again, many of them could not produce a workable plan that could lead to successful movement of the drum of engine oil from the garage onto the raised platform in the compound. Furthermore, many students could not draw appropriate diagrams and outlines for their plans; and yet other students could not state the second safety procedure correctly.

On the performing skills, the major inadequacy showed by the students was the inability of many of them to state the second safety procedure correctly.

On the reasoning skills, the major inadequacies showed by the students include the inability of many of them to describe the shape of the graph correctly; and the inability of many of them to deduce a correct relationship between the angles of incidence and deviation from their measurements.

Conclusions

From the results and findings of the study, it was concluded that the senior high school physics students used for the study did demonstrate high levels of proficiency in laboratory planning, performing and reasoning skills; as well as on the total performance tasks. These findings were supported by the findings of Addai (2001) in which the results on the whole indicated a high student achievement on all the non-traditional tasks (planning, performing and reasoning skills). Again these findings were supported by the findings of Ossei-Anto (1996) in which the students exhibited the highest level of proficiency on the performing skills, followed by the reasoning skills, and the lowest was on planning skills. Results and findings from this study has brought to the fore that if senior high school physics students are given opportunities to practice more laboratory planning, performing, and reasoning skills, they will exhibit high levels of competency in these skills.

It was also clear from the results and findings of the research that the physics students engaged in the study did demonstrated high competencies in all the three aspects of laboratory planning, performing and reasoning skills. This has also brought out to light that if physics students are given opportunities to engage in hands-on, performance, and laboratory activities it will enable them to develop all the three skills necessary for complete and holistic science education.

Again, results and findings from the study has shown that there were no significant differences in the laboratory planning, performing and reasoning skills shown by male and female physics students. In fact, the female students seemed to outperform the male students on performing and reasoning skills. These findings were also supported by the findings of Anthony-Krueger (2001) in which the sex of the students was independent of the performance of the students on the interpreting, inferring and predicting tasks. Again, Seshie

(2001) also found that both male and female chemistry students engaged in laboratory work exhibited similar levels of proficiency on the planning and performing skills. These results and findings therefore run counter to the notion that is held by many people that science, and physics for that matter is a preserve for male students alone.

Moreover, results and findings from the study did show that there were no significant differences in the laboratory planning and performing skills exhibited by the students as a result of differences in the type of school (boys, girls, or mixed) they attended. However, it was found out that there were significant differences in the laboratory reasoning skills shown by students from different types of schools; particularly the difference between students from the girls and mixed schools were prominent. This finding is supported by the findings of Johnson (2001) in which the type of school the student attended had a relationship with his/her performance; with students from girls schools having the highest mean score, followed by students from boys schools; while students from mixed schools had the lowest mean score.

From the results it was also clear that many of the students showed some major inadequacies on the planning, performing and reasoning skills.

To sum up, there have been numerous reports from many quarters that point to the fact that senior high school physics students and graduates do not show adequate laboratory skills of planning, performing and reasoning when engaged in non-traditional, laboratory-based work; However, results and findings from this study has shown clearly that if these students are given more opportunities to practice hands-on and performance activities, they are capable of demonstrating high levels of competency in laboratory planning, performing and reasoning skills.

Recommendations

Although this study was conducted using SHS 3 physics students within Sekondi-Takoradi metropolis of the Western Region of Ghana, and given the difficulty in generalizing results and findings from performance assessments, it is recommended that:

1. Science teachers and educators should give senior high school physics students more opportunities to engage in laboratory, hands-on and authentic activities that will enable them to develop laboratory skills of planning, performing and reasoning.
2. Science teachers and educators should modify their modes of assessments in school and outside the school to include hands-on and laboratory activities.
3. Assessment results from hands-on, laboratory-based performances should be given adequate weight in the grading and certification of senior high school physics students.
4. School and educational administrators should endeavour to provide adequate equipment, resources and motivation for the science teachers to engage physics students in laboratory work.

Suggestions for Further Studies

i. It is suggested that the study should be replicated with students in SHS 1 and 2 to see if the results and findings can hold true for them.

ii. It is also suggested that this study should be replicated in schools in other metropolises and rural towns of Ghana to check if the same results and findings hold true for these places.

iii. It is further suggested that the study should be conducted using other topics selected in physics to check if the results and findings can hold true.

iv. It is again suggested that a study be carried out to find why students from the mixed schools consistently perform lower than their counterparts from the girls and boys schools.

REFERENCES

Addai, J. K. (2001). *Evaluating Practical Skills of Senior Secondary School Students with Some Non-Traditional Tasks in Mechanics*. Unpublished Masters' Thesis: University of Cape Coast, Cape Coast.

Al-Sadaawi, A. S. (2007). *An Investigation of Performance-Based Assessment in Science in Saudi Primary Schools*. Unpublished PhD. Thesis Report: Victoria University, Melbourne. Retrieved on June 25, 2010, from, http://eprints.VU.ed.au/1497//

Amuah, I. R. (1996). Portfolio Assessment: New Trends in the Evaluation of Ghanaian Junior Secondary School Students' Musical Achievement [Electronic Version]. *The African Music Educator*. (8). Retrieved on August 18, 2009, from, http://www.scientific african.org/no8.Pub11

Anthony-Krueger, C. (2001). *Assessing Some Process Skills of Senior Secondary School Students in Elective Biology*. Unpublished Masters' Thesis: University of Cape Coast, Cape Coast.

Ayala, C. C., Shavelson, R. J. & Ayala, M. A. (2001). *On the Cognitive Interpretation of Performance Assessment Scores* [Electronic Version]. Los Angeles: Center for the Study of Evaluation. Retrieved on June 25, 2010, from, http://research.cse.ucla.edu/ Reports/TR546.pdf

Baxter, G. P., Shavelson, R. J., Goldman, S. R. & Pine, J. (1992). Evaluation of Procedure-Based Scoring for Hands-on Science Assessment. *Journal of Educational Measurement, 29* (1), 1 – 7.

Cangelosi, J. S. (1990). *Designing Tests for Evaluating Student Achievement.* London: Longman.

Clarke, J. (2009). *Studying the Potential of Virtual Performance Assessments for Measuring Student Achievement in Science*. Retrieved on June 25, 2010, from, http://Virtualassessment.org/Publications/ aera-2009-Clarke.pdf

Cunningham, G. K. (1986). *Educational and Psychological Measurement*. London: McMillan Publishing Company.

Daily Graphic (2007, April 12). *Another Round of Education Reform*. Retrieved on August 18, 2009, from, http://WWW.GhanaWeb.com /GhanaHomePage/NewsArchive/artikel.php? ID=122323

David, M. & Sutton, C. D. (2004). *Social Research: The Basics*. London: SAGE Publications Ltd.

De Ture, L. R., Fraser, B. J., Giddings, G. J. & Doran, R. L. (1995). Assessment and Investigation of Science Laboratory Skills Among Year 5 Students. *Research in Science Education, 25* (3), 253 – 266.

Downs, D. (2007). *Virtual Public Schools: Why National Achievement Tests Scores Engender Their Support, Part III*. Retrieved on June 28, 2010, http://www.americanchronicle.com/articles/view/23255

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

Ellis, P., Jablonski, E., Levy, A. & Mansfield, A. (2008). *High School Science Performance Assessments: An Examination of Instruments for Massachusetts* [Electronic Version]. Boston: Education Development Center. Retrieved on June 25, 2010, from, http://www7. Nationalacademies.org/base/tom_shiland_presentation_jun_04.pdf

Eubanks, I. D. (1997). *Assessment in Chemistry: New Strategies for New Times*. Retrieved on June 28, 2010, from, http://www.inform.umd. edu/EdRes/Topic/Chemistry/ChemConference/Chemcon97/assess.pdf

Eyiah, J. K. (2004). *Brief History of State-Organized Education in Ghana*. Retrieved on August 18, 2009, from, http//www.Ghanaweb.com GhanaHomePage/NewsArchive/artikel.php? ID=122323

Fraenkel, J. R. & Wallen, N. E. (2000). Experimental Research: *How to Design and Evaluate Research in Education* (4th ed.) London: McGraw-Hill Companies, Inc.

Gronlund, N. E. (1971). *Measurement and Evaluation in Teaching* (2nd ed.). New York: Macmillan Company.

Gronlund, N. E. (1998). *Assessment of Student Achievement* (6th ed.) London: Allyn & Brown.

Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S. & Gernsbacher, M. A. (2007). The Science of Sex: Differences in Science And Mathematics [Electronic Version]. *Psychological Science in the Public Interest, 8* (1). Retrieved on June 28, 2010, from, http://www.PsychologicalScience.org/Journals/pspi_1_article.pdf

Harvey, T.J. (1985). Science in Single-Sex and Mixed Teaching Groups. *Educational Research*, 27(3) pp 179-182.

Hofstein, A. & Lunetta, V. N. (2003). *The Laboratory in Science Education*: *Foundations for the Twenty-First Century*. Retrieved on June28, 2010, from, http://Citeserx.ist.Psu.edu/Veiwdoc/downloads/?doi = 10.1.1.126

International Association for the Evaluation of Educational Achievement [IEA] (1995). *Introduction: Performance Assessment*. Retrieved on June 20, 2010, from, http://timss.bc.edu/timss1995i/TIMSSPDF/ PAintro.pdf

Johnson, E. A. (2001). *Assessing Laboratory Skills of Physics Students in Selected Senior Secondary School Topics in Mechanics*. Unpublished Masters' Thesis: University of Cape Coast, Cape Coast.

Karmel, L. J. (1966). *Testing in Our Schools*. London: Collier-McMillan Limited.

Klein, S. P., Jovanovic, J., Stecher, B. M., McCafferey, D., Shavelson, R. J., Haertel, E., Solano-Flores, G. & Comfort, K. (1997). Gender and Racial/Ethnic Differences on performance Assessments in Science [Electronic Version]. *Educational Evaluation and Policy Analysis, 19* (2), Retrieved on June 28, 2010, from, http://www. Colorado.edu/education/faculty/Guillermosolanoflores/docs/Gender and Racial Ethnic Differences.pdf

Lindvall, C. M. (1967). *Measuring Pupil Achievement and Aptitude*. New York: Harcourt, Brace & World, Inc.

Lloyd-Jones, R. & Bray, E. (1986). *Assessment: From Principles To Action*. London: McMillan Educational Limited.

Marzano, R. J., Pickering, D. & Mctighe, J. (1993). *Assessing Students Outcomes: Performance Assessment Using the Dimensions of Learning Model*. Virginia: Association for Supervision and Curriculum Development. Retrieved on August 18, 2009, from http://www.eric.ed.gov/ERICdocs/data/ericdocs2aq1/contentstorage

01/0000019b/80/19/c8/33.pdf

Mcdaniel, E. (1994). *Understanding Educational Measurement*. Madison: Win. C. Brown Communications, Inc.

Micheels, W. J. & Karnes, M. R. (1950). *Measuring Educational Achievement*. London: McGraw-Hill Book Company, Inc.

National Center for Education Statistics [NCES] (2007). *Highlights From TIMSS 2007: Mathematics and Science Achievement of U.S. Fourth -and Eighth Grade Students in an International Context*. Retrieved on June 18, 2010, from, http://nces.ed.gov/pubs2007/ 2007049.pdf

Nitko, A. J. (2004). *Educational Assessment of Students* (4th ed.). Columbus: Merrill, PP 5 – 7.

Noll, V. H. (1965). *Introduction to Educational Measurement* (2nd ed.) Boston: Houghton Mifflin Company.

Nunnally, J. C. & Ator, N. A. (1972). *Educational Measurement and Evaluation* (2nd ed.). New York: McGraw-Hill Book Company.

Oloruntegbe, K. O. (1999). *Alternative Approaches to the Assessment of Science Process Skills*. Retrieved on June 25, 2010, from, http:// www.Imechanica.org/files/ALTERNATIVES.doc

Organization for Economic Cooperation and Development [OECD] (2007) *The Programme for International Student Achievement (PISA 2006)*. Retrieved on June 28, 2010, from, http://www.oecd.org/ Dataoecd115/13/39725224.pdf

Ossei-Anto, T. A. (1996). *Assessing Laboratory Skills of Students in Selected High School Topics in Optics*. Unpublished Doctoral

Thesis: University of New York at Buffalo.

Payne, D. A. & Mcmorris, R. F. (1967). *Educational and Psychological Measurement*. London: Blaisdell Publishing Company.

Ruiz-Primo, M. A. & Shavelson, R. J. (1996). Rhetoric and Reality in Science Performance Assessments: An Update. [Electronic Version] *Journal of Research in Science Teaching, 33* (10). Retrieved on June 25, 2010, from, http://www.Stanford.edu/dep/SUSE/SEAL/Reports -Papers/

Schwart, A., Tiedeman, S. C. & Wallace, D. G. (1962). *Evaluating Student Progress in the Secondary School*. New York: David McKay Company, Inc

Seshie, A. H. K. (2001). *Assessing Laboratory Skills of Students in Selected Senior Secondary School Elective Chemistry Topics in Titrimetric Analysis*. Unpublished Masters' Thesis: University of Cape Coast, Cape Coast.

Shavelson, R. J., Ruiz-Primo, M. A., Li, M. & Ayala, C. C. (2003). *Evaluating New Approaches to Assessing Learning* [Electronic Version]. Los Angeles: Center for the Study of Evaluation. Retrieved on June 25, 2010, from, http://www.cse.Ucla.edu/ Products/Reports/R604.pdf

Shaw, J. M. & Nagashima, S. O. (2009). The Achievement of Student Subgroups on Science Performance Assessments in Inquiry-Based Classrooms. *Electronic Journal of Science Education, 13* (2). Retrieved on June 28, 2010, from, http://ejseSouthwestern.edu/ Volumes/v13n2/articles/artl.pdf

Slater, T. F. (1997). Performance Assessment. Retrieved on June 25, 2010, from, http://www.flaguide.org/extra/download/Cat/Perfass/Perfass.pdf

Somekh, B. & Lewin, C. (Eds.). (2005). *Research Methods in the Social Sciences*. London: SAGE Publications Ltd.

Spelke, E. S. (2005). Sex Differences in Intrinsic Aptitude for Mathematics and Science? A Critical Review [Electronic Version]. *American Psychologist, 60* (9). Retrieved on June 28, 2010, from, http://www. Wjh.havard.edu/Ids/pdfs/Spelke2005.pdf

Ssempala, F. (2005). *Gender Differences in Performance of Chemistry Practical Skills among Senior Six Students in Kampala District* [Electronic Version]. Florida: Dissertation.com. Retrieved on June 28, 2010, from, http://www.bookpump.com/dps/pdf_b/942700/b.pdf

Stranahan, H. A., Borg, J.R. & Borg, M. O. (2002). School Grades Based on Standardized Test Scores: Are They Fair? [Electronic Version]. *Journal Of Academic and Business Ethics*. Retrieved on June 28, 2010, from, http://www.aabri.com/manuscripts/08006.pdf

Underwood, G. & McCafferey, M. (1990). Gender Differences in a Cooperative Computer-Based Language Task. *Educational Research, 32* (1), 44 – 49.

WAEC (2000). *Chief Examiner's Report on Senior Secondary School Certificate Examinations*. Accra: Wisdom Press.

WAEC (2001). *Chief Examiner's Report on Senior Secondary School Certificate Examinations*. Accra: Wisdom Press.

WAEC (2002). *Chief Examiner's Report on Senior Secondary School Certificate Examinations*. Accra: Wisdom Press.

WAEC (2003). *Chief Examiner's Report on Senior Secondary School Certificate Examinations*. Accra: Wisdom Press.

WAEC (2004). *Chief Examiner's Report on Senior Secondary School Certificate Examinations*. Accra: Wisdom Press.

WAEC (2005). *Chief Examiner's Report on Senior Secondary School Certificate Examinations*. Accra: Wisdom Press.

WAEC (2006). *Chief Examiner's Report on Senior Secondary School Certificate Examinations*. Accra: Wisdom Press.

Appendix A

DEPARTMENT OF SCIENCE AND MATHEMATICS EDUCATION

TELEGRAMS & CABLES:
UNIVERSITY, CAPE COAST
TELEPHONE:
042-32440/4 EXT. HEAD: 202
OFFICE: 042-34890

FACULTY OF EDUCATION
UNIVERSITY OF CAPE COAST
CAPE COAST, GHANA

Your Ref.: Our Ref.: SED/045/ Date: January 20, 2010.

TO WHOM IT MAY CONCERN

RESEARCH VISIT

I am introducing the bearer **MR. SALIFU MAIGARI MOHAMMED** who is an **M.PHIL (SCIENCE EDUCATION)** student of this Department with Registration Number **ED/SCP/08/0007**, embarking on a research which will require the participation of staff/students/clients in your organization/institutions/health facility.

I would be grateful if you could give him your usual co-operation.

Thank you.

Yours faithfully,

Prof. Theophilus A. Ossei-Anto
[Supervisor]

Appendix B

Task A: Planning

Instruction:

You have **5** minutes to read carefully through the task, and another **25** minutes to respond to the task. At the end of the period your work will be collected from you. During the period of reading you may ask questions for clarification of any point that seems unclear. No questions will be entertained when you start responding to the task.

Information:

Workers in Construction firms use simple machines to move heavy objects from one place to another. Heavy objects are moved by rolling them on rollers on the ground; pushing them on inclined planes onto platforms; and by tying the objects to ropes and pulling them along pulleys onto tops of buildings.

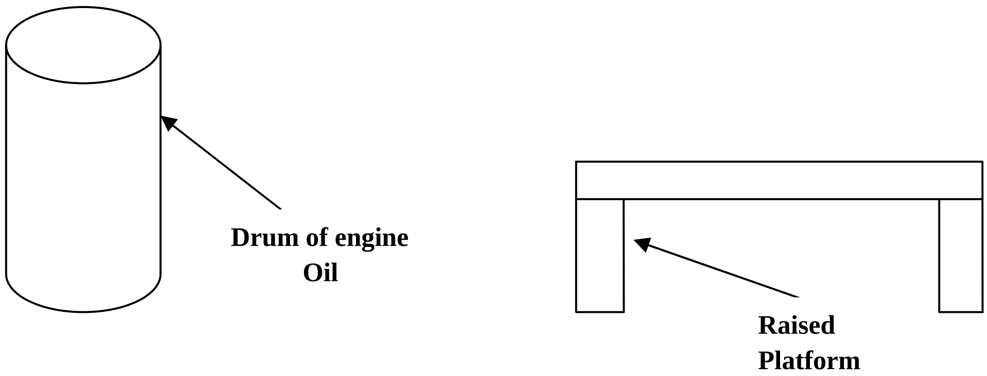

Figure 17: Diagram of Drum of Engine Oil and Raised Platform

Problem:

The task is that three of you have been employed in a construction firm and you are to move a heavy drum full of engine oil from the garage onto a raised platform in the compound. Indicate in the space provided for "PROCEDURE" the materials and steps you will use to complete the task. You may draw diagrams to show your plan.

Materials:

(a) 4 steel poles (b) car jack (c) rope (iv) wooden beam (d) trolley truck (e) ladder (f) Crowbar (g) Pulley

Procedure:

....

....

....

....

....

....

....

....

....

....

....

....

....

....

Note: State below any two (2) safety precautions you should have to take to complete the task.

Safety Precautions:

1.

2.

Appendix C

Task B: Performing

Instruction:

You have **5** minutes to read carefully through the task, and another **25** minutes to respond to the task. At the end of the period your work will be collected from you. During the reading period you may ask questions for clarification of any point that seem unclear. No questions will be entertained when you start responding to the task.

Information:

Density measurements are made to determine the densities of various substances and their suitability for use in various works.

Procedure:

1. Weigh the empty density bottle and record its mass ($\mathbf{m_1}$).
2. Half fill the density bottle with sand, weigh the mass and record it ($\mathbf{m_2}$).
3. Add water to the sand to fill the bottle, weigh the mass and record it ($\mathbf{m_3}$).
4. Pour the sand and water away and thoroughly clean the density bottle.
5. Fill the bottle with water only, weigh the mass and record it ($\mathbf{m_4}$).
6. Subtract the (mass of the empty bottle – $\mathbf{m_1}$) from the (mass of bottle and sand – $\mathbf{m_2}$) to find the mass of sand = $(\mathbf{m_2 - m_1})$.
7. Subtract the (mass of the bottle and sand – $\mathbf{m_2}$) from the (mass of bottle, sand and water – $\mathbf{m_3}$) to find the mass of water added to sand = $(\mathbf{m_3 - m_2})$.
8. Subtract the (mass of empty bottle – $\mathbf{m_1}$) from the (mass of bottle and water only – $\mathbf{m_4}$) to find the mass of water only = $(\mathbf{m_4 - m_1})$.
9. Subtract the answer you got in step **7** from the answer you got in step **8** to find the mass of water of equal volume as sand = $(\mathbf{m_4 - m_1}) - (\mathbf{m_3 - m_2})$.

10. Divide the answer you got in step **6** by the answer you got in step **9** to find

the relative density of sand = $\frac{\text{(mass of sand)}}{\text{(mass of equal volume of water)}}$

$$= \frac{(m_2 - m_1)}{(m_4 - m_1) - (m_3 - m_2)}$$

Record your results in the space provided below.

Materials:

You are provided with the following materials:

Density bottle, sand, water, beam balance/electronic balance, measuring cylinder.

Note: State below two (2) safety precautions you have to take to complete the task.

Safety Precautions:

1.

2.

Results:

Appendix D

Task C: Reasoning

Instruction:

You have **5** minutes to read carefully through the task, and another **25** minutes to respond to the task. At the end of the period your work will be collected from you. During the reading period you may ask questions for clarification of any point that seem unclear. No questions will be entertained when you start responding to the task.

Information:

The direction of a light ray changes as it moves from air into a glass prism. The light ray is deviated from its original path onto a new path as it enters and emerges from the glass prism. The angle between the incident ray and the normal is the incident angle (I^o), and the angle between the incident and the emergent rays is the angle of deviation (D^o). The line drawn perpendicular to the surface of the glass prism at the point where the incident light strikes the glass prism is called the normal (N).

Problem:

In an effort for a student to investigate the paths taken by light rays as they move from air into triangular glass prism, the student incident light rays at different angles of incidence on the glass prism and traced the corresponding deviations in the paths of the light as they emerge out of the prism. The angles of incidence (I^o) and the corresponding angles of deviations (D^o) are shown in diagrams **a**, **b**, **c**, **d**, **e**, and **f** below (Figures 18a and 18b). Measure the angles of incidence and the corresponding angles of deviations. Record your measurements in a tabular form as shown in the table below. An example of the measurements taken and the recordings are shown in Table 36 below.

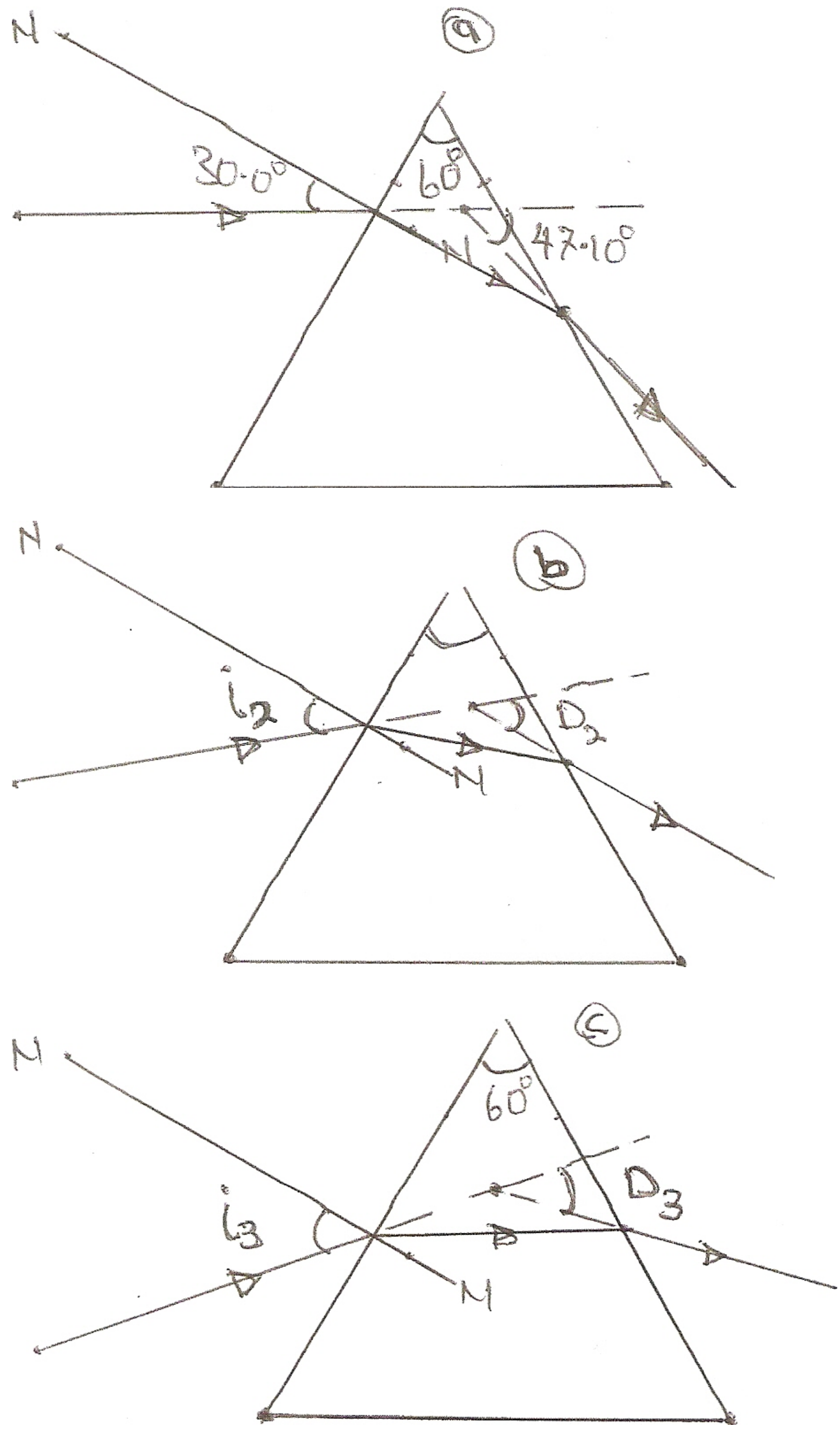

Figure 18a: Diagrams Showing Traces of Light Rays Through Triangular Glass Prism at Different Angles.

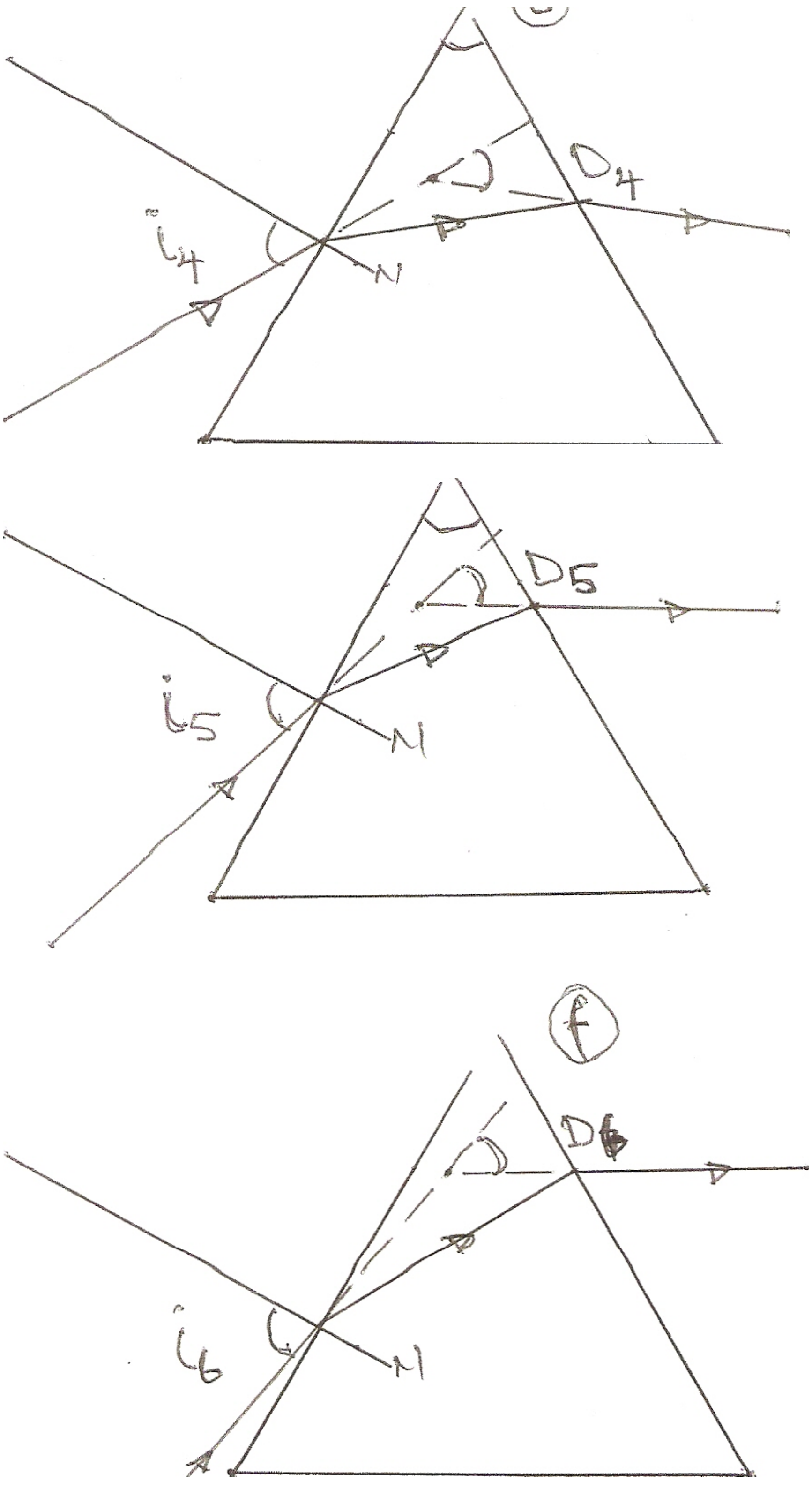

Figure 18b: Diagrams Showing Traces of Light Rays Through Triangular Glass Prism at Different Angles.

Table 36

Table Showing Angles of Incidence and Deviation

Angle of Incidence/i^o	Angle of Deviation/D^o
30.00	47.10

1. Choosing convenient scales draw vertical and horizontal axes.

2. Label the horizontal axis incident angle and the vertical axis angle of deviation.

3. Plot a graph of incident angle versus angle of deviation.

4. Describe the shape of the graph.

5. From the graph deduce a relationship between angle of incidence and angle of deviation.

Appendix E

Task D: Opinionnaire

This section is meant to solicit your opinion about the performance tasks you have just completed.

1. Task A

Planning

Time Provided: Not Enough ....

About Right: ....

Too Much: ....

Difficulty of Task: Easy...

About Right..

Difficult..

Any other comments:..

....

2. Task B

Performing

Time Provided: Not Enough...

About Right..

Too Much..

Difficulty of Task: Easy...

About Right...

Difficult..

Any other comments:...

....

....

3. Task C

Reasoning

Time Provided: Not Enough...

About Right...

Too Much..

Difficulty of Task: Easy..

About Right...

Difficult...

Any other comments:...

....

....

4. How would you commend revising these tasks to make them better?

....

....

....

5. How much time every two weeks do you spend doing physics practical in the school?

....

....

....

Appendix F

Scoring Format: Planning

General Strategy	(0	1)
Sequential Plan	(0	1)
Detailed Plan	(0	1)
Workable Plan	(0	1)
Materials	(0	1)
Diagrams/Outline	(0	1)
Safety Procedure S1	(0	1)
Safety Procedure S2	(0	1)

Scoring Details

General Strategy

Score 1 if a student demonstrates any sort of planning.

Sequential Planning

Score 1 if a student's plan is sequenced logically or scientifically.

Detailed Plan

Score 1 if a student's plan is detailed enough.

Workable Plan

Score 1 if a student's plan seems workable.

Materials

Score 1 if a student uses correct materials.

Diagrams/Outline

Score 1 if a student draws appropriate diagram of the plan.

Safety Procedure (S1)

Score 1 if a student states the 1st correct safety procedure.

Safety Procedure (S2)

Score 1 if a student states the 2nd correct safety procedure.

Total: 0 1 2 3 4 5 6 7 8

Appendix G

A Model Answer: Planning

The four steel rollers are arranged horizontally on the floor of the garage near the drum of oil. The three of us (construction workers) gently lower the drum of oil onto the steel rollers on the floor. The drum of oil is then pushed to roll it out of the garage into the compound till it gets to the foot of the raised platform. Rollers that are left behind as the drum is being rolled toward the platform are rearranged on the ground in front of the drum. At the foot of the raised platform, one end of the wooden board is placed on the ground while the other end is placed against the platform to form in inclined plane. The three of us then roll the drum of oil gently on the wooden incline plane until it gets to the top of the platform. The ladder is then placed against the platform. The three of us climb the ladder to the top of the platform. On the platform we gently raise the drum of oil to make it stand upright onto it's base.

Safety Procedures:

1. The drum of is gently lowered onto the rollers.

2. The drum of oil is tightly covered to prevent oil spillage.

3. The wooden board is inclined at a low angle (not steep) to the platform.

4. It is ensured that the wooden board is strong enough to support the weight of the drum of oil.

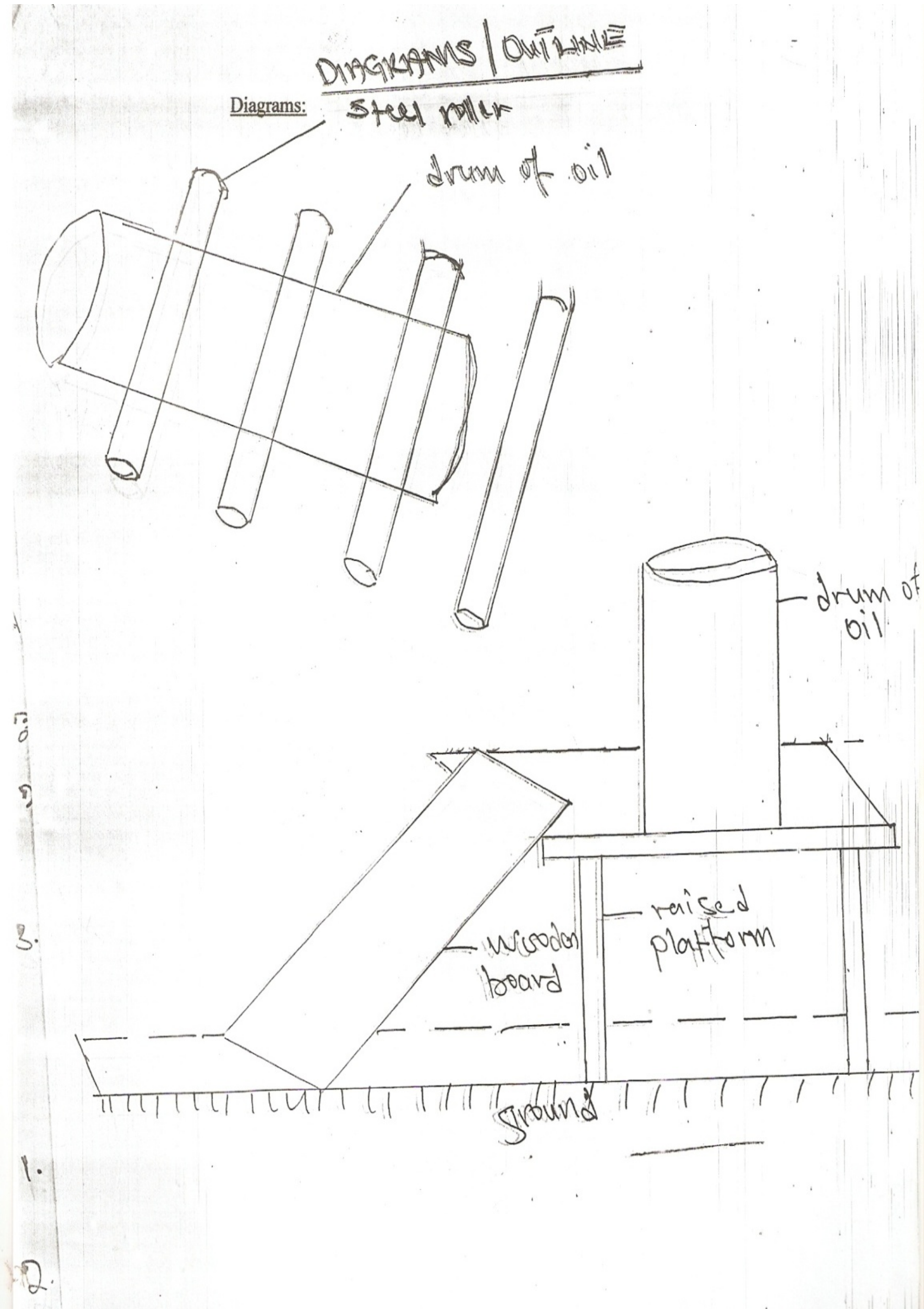

Figure 19: Diagrams Showing Drum of Engine Oil on Rollers and on Raised Platform.

Appendix H

Scoring Format: Performing

Mass of empty density bottle correctly measured	(0	1)
Mass of bottle and sand correctly measured	(0	1)
Mass of bottle, sand and water correctly measured	(0	1)
Mass of bottle and water only correctly measured	(0	1)
Mass of sand correctly determined	(0	1)
Mass of water added to sand correctly determined	(0	1)
Mass of water only correctly determined	(0	1)
Mass of equal volume of water as sand correctly determined	(0	1)
Relative Density of sand correctly determined	(0	1)
Safety Procedure (S1)	(0	1)
Safety Procedure (S2)	(0	1)

Total: 0 1 2 3 4 5 6 7 8 9 10 11

Appendix I

A Model Answer: Performing

Mass of empty density bottle (m_1) = 23.1g

Mass of bottle + Sand (m_2) = 48.5g

Mass of bottle + Sand + Water (m_3) = 62.4g

Mass of bottle + Water (m_4) = 48.0g

Mass of Sand ($m_2 - m_1$) = 25.4g

Mass of Water added to Sand ($m_3 - m_2$) = 13.9g

Mass of Water only ($m_4 - m_1$) = 24.9g

Mass of equal volume of water as sand ($m_4 - m_1$) – ($m_3 - m_2$) = 11.0g

$$\text{Relative Density of Sand} = \frac{(\text{mass of sand})}{(\text{mass of equal volume of water})}$$

$$= \frac{(m_2 - m_1)}{(m_4 - m_1) - (m_3 - m_2)}$$

$$= 25.4\text{g}/11.0\text{g}$$

$$= 2.31$$

<u>Safety Procedures</u>:

1. The density bottle should be thoroughly cleaned to remove any particle in it after filling it with sand.

2. The density bottle should not be held in warm hands to prevent evaporation of water.

3. Avoid parallax error when reading the beam balance.

4. Zero error of the balance must be accounted for.

Appendix J

Scoring Format: Reasoning

1^{st} incident angle and corresponding angle of deviation correct	(0	1)
2^{nd} incident angle and corresponding angle of deviation correct	(0	1)
3^{rd} incident angle and corresponding angle of deviation correct	(0	1)
4^{th} incident angle and corresponding angle of deviation correct	(0	1)
5^{th} incident angle and corresponding angle of deviation correct	(0	1)
Convenient Scale correctly chosen	(0	1)
Horizontal axis correctly labelled	(0	1)
Vertical axis correctly labelled	(0	1)
1^{st} point correctly plotted	(0	1)
2^{nd} point correctly plotted	(0	1)
3^{rd} point correctly plotted	(0	1)
4^{th} point correctly plotted	(0	1)
5^{th} point correctly plotted	(0	1)
6^{th} point correctly plotted	(0	1)
Correct description of curve	(0	1)
Correct Relationship Deduced from graph	(0	1)

Total: 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

Appendix K

A Model Answer: Reasoning

Measurements

Incident Angle	Angle of Deviation
30.0	47.1
40.0 ± 1	38.5 ± 0.5
50.0 ± 1	36.3 ± 0.5
60.0 ± 1	38.9 ± 0.5
70.0 ± 1	42.9 ± 0.5
80.0 ± 1	49.2 ± 0.5

Description of Graph

1. The graph is parabolic in shape.

2. The graph is U – shaped.

3. Minimum curve graph.

Description of Relationship between Incident angle and angles of Deviation

Angle of deviation decreases to a minimum point and then increases while the incident angle increases continuously.