RELIABILITY, ITEM DIFFICULTY AND DISCRIMINATION INDICES OF SCIENCE TEST ITEMS IN COMMERCIAL TEST PAPERS AND THEIR CORRELATION TO STUDENTS KCPE PERFORMANCE IN SCIENCE IN LIMURU SUBCOUNTY.

BONIFACE CHEGE NDUNG'U

DEPARTMENT OF PSYCHOLOGY

SCHOOL OF EDUCATION

UNIVERSITY OF NAIROBI

2015

A research project submitted to the University of Nairobi in partial fulfillment of the requirements for the Masters of Education degree (M.Ed.) in Measurement and Evaluation.

DECLARATION

This Research is my original work and has not been presented for approval or examination in any University.

SIGN……………………..   DATE…………………….

Boniface Chege Ndung'u

E58/63591/2013

This Research Project has been submitted for examination with my approval as University supervisor.

SIGN………………………………. DATE…………………………..

Dr. Luke O.Odiemo

Department of Psychology

University of Nairobi

DEDICATION

I dedicate this research project to my late mum Loise Nyambura who believed in hard and honest work. To my dear family for being there for me ,to my Wife Esther, daughters Loise and Veronicah who supported me morally when the going became too tough. No religion is higher than the truth.

ACKNOWLEDGEMENTS

First to the almighty God, for the gift of life, time, strength and resources to complete this work. To my supervisor Dr. Luke Odiemo, for his fatherly guidance, scholarly and constructive advice throughout the project may God bless you.I also wish to express my gratitude to all lecturers who taught us different units with commitment, again the tireless effort of the coordinator Dr. Karen Odhiambo.

ABSTRACT

The study was carried out to establish the factors influencing the reliability, item difficulty and discrimination indices of science test items in commercial test papers and their correlation to students KCPE performance in science in Limuru Sub-County, Kiambu County in Kenya. The study involved 200 class eight students selected using stratified random sampling from three public primary schools, 39 teachers teaching science in class eight in public primary schools in Limuru sub-county. A descriptive survey research design was applied. Data was collected through questionnaires and sample science commercial test papers administered to the students. Data collected via the sample commercial test papers was analysed first using Ms Excel, then together with the data from the questionnaires analysed using SPSS and the frequency distribution was then described. The findings of the study revealed that test duration, test quality, teacher's; training on test construction, professional qualification, teaching experience, workload ,number of test items objectives, intended purpose of the test, testwiseness have an effect on the reliability, difficulty and discrimination indices. Other factors such as teacher and student gender, parental assistance and educational level, family type and socio- economic status have a varying influence on the student performance in summative evaluation. The teachers training on test development was seen as the main factor that would determine the quality of commercial test papers as flawed items would be moderated to come up with an almost perfect test paper. It was concluded that the commercial science test paper used in formative assessment are generally of good quality. The study recommended that science teachers should attend in- service training on test development and analysis to enhance their competency in verifying the quality of commercial test papers.

TABLE OF CONTENTS

CHAPTER ONE

INTRODUCTION

CHAPTER TWO

LITERATURE REVIEW

CHAPTER THREE

RESEARCH METHODOLOGY

CHAPTER FOUR

RESEARCH FINDING AND DATA ANALYSIS

CHAPTER FIVE

DISCUSSION SUMMARY AND CONCLUSION

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

1.1 Background information

The main problem in this study is that commercial tests used for mock examinations are not predictive of the final KCPE score attained by the learners. The national mean score for 2013 KCPE Science subject was 61.82 while that for Limuru Sub-county was 57.2 (KNEC examinations report, 2013), indicating that candidates in Limuru sub-county performed slightly below the national mean score in the Kenya certificate of primary education,2013. A comparison between the KCPE Science Subject mean score and the corresponding science subject mean score for mock examination in Limuru Sub-county are shown in the table below:

Table 1.1: Limuru Sub-county mock and KCPE results.

YEAR	MOCKS SCIENCE MEAN SCORE	KCPE SCIENCE MEAN SCORE	Expected mean	Deviation
2010	53.99	46.95	100	-7.04
2011	53.85	45.87	100	-7.98
2012	53.74	47.34	100	-6.4
2013	48.77	40.67	100	-8.1

Source: Limuru Sub- county education office 2015.

From the results shown in the table it is evident that the mock examinations science score over the four years were not quite predictive of the final KCPE score indicating that, the commercial tests used in preparing the students were lacking in some qualities.

In many countries world over there exists limited opportunities for all the individuals in the education system to be absorbed into institutions of further education (Geaney&Kellaghan, 1995).This necessitates making of choices on who should be considered for further education. Testing therefore becomes the main method that provides evidence of the level of learning upon which such choices can be made (Payne, 1992).

In Kenya, national examinations have received much emphasis in that a good school is not only looked at in terms of policies, practices or programs, but on how well it performs relative to other schools (Ouko,2004).The increased emphasis on performance in national examination has seen the rise of partnership examinations between schools to mock the national examination set by subject teachers (Chepkurui,1999).However with increased teacher workload and student apathy to teacher made mock examinations saw the rise of commercial mock examinations (Mwangi,2006), compiled from past papers or set by these exam vendors(Kigotho,2004).

Survey carried out by Leonard (2014) reported that the quality of mock exam results is affected by type of examination, conditions under which the mock exams are done, extent of syllabus coverage, individual characteristics of the test takers, marking errors and the quality of the test items. The first five factors are beyond the scope of this study therefore the study will lay emphasis on the quality of these commercial examinations. Quality of a test is judged on the basis of its psychometric properties.

According to Kline (2000) Psychometrics entails those aspects of Psychology that are concerned with psychological testing. The psychometric properties that every measuring instrument should

possess are validity and reliability. Test validity is determined by how well it samples a range of knowledge, skills and abilities the students are supposed to acquire in the period covered by the test (Angelica, 2006).It is the degree to which common sense, or theory supports any interpretations or conclusions about a student based on his or her test performance (Hunter & Schmidt,1999). Validity therefore is the extent to which inferences made from tests are justified and accurate (Frisbie, 1988).

Reliability on the other hand depends upon the grading consistency and discrimination between students of differing performance levels. It refers to the consistency of the test scores obtained from one administration to another and from one item to another (Hotiu, 2006).Reliability is a measure of how stable, dependable, trustworthy and consistent a test is in measuring the same each time (Worthen et al, 1993).

Types of reliability include test-retest, parallel form reliability, split half reliability, inter-rater reliability and internal consistency. A study at Monroe central high school, Ohio showed that the length and number of test items affect reliability. However the study only investigated test-retest reliability of the test scores. Carefully written tests with an adequate number of items usually produce high reliability (Justin& Justin, 1996) since they are less distorted by chance factors (Lin&Gronlund, 1995).

A test's validity and reliability can be ascertained through item analysis (Nwaobia, 1990). Item analysis is a process which examines students' responses to individual test items in order to assess the quality of those items and quality of the test as a whole (Sushma, 2013). Item analysis

therefore is done to indicate the difficulty level of each test item and its ability to discriminate between high performing and low performing students (Ebel, 1986).

Item difficulty is the proportion of students who responded correctly to an item while item discrimination is a measure to differentiate between the performance of students in the high score group and those in the low score group (Instructional Assessment Resources, 2011). Research carried out by Suruchi et al (1991) in India on the relationship between difficulty index and discrimination index of test items in an achievement test in biology using Pearson product moment correlation found out that the discrimination power increased with an increase in difficulty index but decreased for very easy and very difficult test items .

The research further found out that item analysis results are tentative , and are influenced by the number and kind of students, instructional procedure applied, chance errors and the purpose of the test. As in a class test very difficult or very easy items may be retained despite having poor discrimination power because the purpose is to test content masterly and attainment of set objectives and not to discriminate the superior and inferior students (Rana, 1991).

Studies by Cronbach & Warrington, 1952, Ebel 1973 and Anastasi 1976 showed that the difficulty level of any item that constitute a test determines not only the mean difficulty level of the test, but also the spread of the total test scores whose difficulty index clusters around 0.5.These studies further showed that the reliability coefficient of a test is highest when the difficulty level of the items is concentrated around 0.5 and particularly low in tests having extreme difficulty levels. An acceptable range of these two parameters was set by Ebel & Frisbie (1986) to be $0.2 < p > 0.8$ for difficulty index and >0.2 for the discrimination index. This is

supported by Brown (1983) and Algina (1983) who reported that any Discrimination index of 0.2 or higher is acceptable for the test items to be able to differentiate between the weak and good students. On the other hand in a study by Mitra N.K et al,2009 to determine levels of difficulty and discrimination indices in pre-clinical multiple choice questions found out that a wide scatter of item discrimination values for questions with a similar level of difficulty may reflect some extent of guessing practices.

A good commercial test should have difficulty level and discrimination indices of its test items within an acceptable range making it both valid and reliable, however if these qualities are not met, then it becomes impossible and even deceptive to use it for evaluation purposes.

1.2 Context of the study

In Kenyan education, students spend 8 years in primary school before they sit for Kenya Certificate of Primary Education (KCPE) for progression into secondary school. Due to the limited slots in secondary school, teachers in primary schools prepare their class eight candidates by giving them several internal examinations and one major exam just before they sit for the final national examination to mock it. These internal examinations are sourced from commercial examination vendors therefore their quality has not been ascertained (Mwangi, 2006).

This study therefore seeks to establish the quality of these commercial tests by determining their reliability, item difficulty and discrimination indices and also correlate them to the KCPE results. The study will be carried out in Limuru Sub-county of Kiambu County in Kenya which has a total of 40 public primary schools.

1.3 Statement of the problem.

The manner in which the test questions are prepared and put together to form an examination, and the procedure for scoring, analyzing and reporting the results all have a bearing upon the conclusion drawn from the performance of the individuals and groups tested (Shakil,2014).

In Limuru Sub-county the tests used for internal evaluation are sourced from commercial test vendors, and their Item difficulty and Discrimination indices therefore their Validity and Reliability have not been established.

According to the Kenya National Examinations Council (KNEC) report on KCPE analysis (2013) the student's raw scores are standardized into standard score to allow for differences in difficulty among the candidates to be established. This is done with the help of set examination mean which is normally 50 for normal distribution and a standard deviation set at 15,however for internal assessment only the mean score is established for results analysis. The national mean raw score for 2013 KCPE science subject was 30.91 while that for Limuru Sub-County was 28.6 (KNEC examinations report, 2013), indicating a slight deviation from the national mean.

Research carried out by the Nigeria Educational Research and Development Council and Kenya National Examinations Council (2000) revealed that teachers have too many responsibilities and the resources are scarce making it difficult for teachers to carry out sound assessment. Another research on assessment by Kurebwa& Nyaruwata (2013) in Midland's province of Zimbambwe on surveying 14 rural schools observed that there was a casual approach to testing, measurement and evaluation in most schools which was attributed to the teachers addressing only the lowest

level of the Bloom's taxonomy objectives, while others duplicating past examination papers. Best practices in assessment emphasize on tests that are valid and reliable however at policy level very little consideration have been focused on these commercial tests in terms of the two constructs given the importance with which these tests are regarded with by teachers for use in assessment(Mwangi,2006).

Test items for examinations in schools are supposed to be prepared and put together by the teacher who has the knowledge of the content covered, student's characteristics and the expected objectives to be met at the end of the learning cycle. According to Stiggins (2001) sound assessment results only when there is a clear purpose for assessment and appropriate targets, proper methods and appropriate sample of the targets and elimination of bias and distortion in measurement. In Kenya this has led to the proliferation of commercial test vendors who have taken over the task of test preparation, and teachers use them widely as testing tools for within the school assessment (Kigotho, 2004).

Very few studies have been conducted to test the quality of commercial test papers by looking at their item difficulty and discrimination index which in essence is the basis for validity and reliability. It is from this perspective that this study will be undertaken in primary schools in Limuru sub-county with a bias for class eight science commercial test papers.

1.4 Significance of the study

The study intends to bring out the importance of using quality testing tools in the assessment of science subject at the primary cycle of education. The findings of the study will provide useful information to Commercial test publishers so as to embrace change in setting science test items that are of quality.

To the science teachers and schools the study provides a way of identifying quality science test papers from commercial test vendors that are suitable tools for predicting student's performance in the summative assessment. . All the above will in a great way improve the Formative and Summative Assessment performance of science subject at the primary level of education in Kenya.

1.5 Main objective

The main objective of the study is to determine the quality of science commercial test papers used in primary schools in Limuru Sub-county by looking at their reliability, difficulty level and discrimination indices

1.6 Objectives of the study

i) To find out:

a) The Reliability of Commercial Science test papers used in Public Primary Schools in Limuru Sub-county.

b).The Difficulty level and Discrimination indices of individual test items in Commercial Science test papers used in Public Primary schools in Limuru Sub-county.

c).The correlation between student performance in Commercial Science test papers and KCPE science performance in public primary schools in Limuru Sub-County.

1.7 Research Questions

i) What is the:

a) Reliability of Commercial Science test papers used in assessment of pupils in Public Primary Schools in Limuru Sub-county?

b).Difficulty level and Discrimination indices of individual test items in Commercial Science test papers used in Public Primary Schools in Limuru Sub-county?

c).Correlation between student performance in Commercial Science test papers and KCPE science performance in public primary schools in Limuru Sub-County.

1.8 Research Hypothesis

H_O .The commercial science test papers used for formative assessment in public primary schools in Limuru sub-county are of good quality.

1.9 Assumptions of the study

All schools in Limuru sub-county purchase commercial examination papers for their class eight students to be used for internal assessment. Secondly the study assumes that all the schools

sampled and used in the study purchased these papers from the same vendors. Finally the Science test papers selected had test items lifted from previous KCPE examination papers.

1.10 Limitations of the study

A representative sample was drawn from Public Primary schools within Limuru Sub-county for the research to be practical. Secondly in the schools selected only two streams of class Eight were chosen and used to capture the data.

1.11 Delimitation of the study

The study was delimited to Public primary schools in Limuru Sub-county, which had candidates for the last four years.

1.12 Organisation of the study

The study will be organized in five chapters:

Chapter one contains the Background of the study, Statement of the problem, Purpose of the study, Objectives of the study, Research questions , Research hypothesis, Significance of the study, Justification ,Limitations ,Delimitations, Assumptions and Definition of significant terms to be used in the study.

Chapter two of the study contains an Introduction Review of the related Literature derived from relevant studies carried out on Reliability, factors affecting reliability and how they are controlled, Anatomy of Multiple Choice Questions, Quality of test items Item Analysis, The degree of Item Difficulty and Discrimination indices, Mean Item Discrimination, and the

Relationship between Item Difficulty and Discrimination indices, Interaction between Reliability ,Item Difficulty and Discrimination indices, Bloom's taxonomy, Theoretical Framework and the Conceptual framework

Chapter three of the study has an Introduction, Research design to be employed in the study, Target population, Sample and Sampling techniques, Research Instruments, Piloting and adjusting the research instruments, Coding scheme, Instrument reliability, Instrument Validity, Data analysis and Ethical issues that will arise in course of conducting the study.

Chapter four will have an Introduction, Data analysis and Interpretation of the findings in form of Tables, Charts and figures.

Chapter five of the study has an Introduction, Aim of the study, Summary of the Findings, Policy Recommendations, Suggestions for further research and the conclusion.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

The purpose of reviewing the earlier researches is not only to economise the historical perspective of the present work, but also to take cognizance of the related studies which have employed one or more variables included in this study.

The literature review in this study is organized under the following topics; Factors affecting Reliability, Difficulty index, Discrimination index and Performance in commercial test papers.

2.2 Factors affecting Reliability.

The variables affecting reliability of a test include; test length, test duration and test quality and teacher characteristics.

2.2.1 Test length

According to Ehigie (2012) the more the items there is in a scale designed to measure a particular concept, the more reliable will be the measurement. This implies that the larger the number of items in a test, the higher the chance of obtaining a better and more reasonably higher reliability coefficient for such a test (Scattler, 2001). This view is supported by studies by Linn&Gronlund (1995); Justin&John (1996) and Stephen (2004) which found out that carefully written tests with an adequate number of items usually produce high reliability, since they provide a representative sample of the behavior being measured and the scores are apt to be less distorted by chance factors such as familiarity with the given test or misinterpreting the expected response from an item.

According to Spearman brown prophesy equation, as the number of test items increase, there is a direct increase in the reliability estimate (Oncu, 1994; Traub, 1994). However a study by Gregory (1992) shows that increasing the number of test items does not necessarily increase reliability though there was a noted decrease in the measurement error.

According to a study by Meshkeni& Hossein (2005) on the multivariate analysis of factors influencing reliability of a teacher made test in Israel revealed that increasing the length of the test while maintaining other factors constant improves reliability. Using descriptive statistics method, the study employed multiple choice questions based tests as the data collection instrument and the results were analysed by statistical package for social sciences (SPSS) using multivariate regression analysis for determining the effects of each factor and Kudder Richardson formulae (KR20) was used to calculate the reliability.

 The study also found out that keeping other variables fixed, the interaction of the length and difficulty indices resulted in an increase in reliability by eight times for items with moderate difficulty and thirteen times for items that were relatively easy. However the study concluded that without considering the quality of test items, increasing the number of questions in order to increase reliability is a big mistake. This view is supported by Tekin (1977) who averred that if the test is not reliable to increase the number of items does not make the test reliable; moreover huge number of items may cause tiredness.

2.2.2 Duration of the test

According to Mehrens&Lehmans (1991) and Concu (1994), if the test is prepared to measure a given attribute in a limited time, the insufficiency of time decreases the reliability of the scale. The time must be enough for respondents to answer all the test items. However if the time is insufficient it causes excitement and carelessness on part of the test taker resulting in reduced reliability of the test (Carey, 1988; Turgut, 1993; Traub, 1994)This view is supported by Linn& Gronlund(2005) who argued that in some instances low reliability of teacher made tests occur due to time constraints.

Test takers who work at similar rates have high degrees of variance correspondence artificially inflating the reliability of the test. However the situation becomes problematic when the construct that the test intends to measure has nothing to do with the speed competency (Crocker&Algina, 1986).Furthermore test duration is a factor of the test type with multiple choice questions test taking shorter compared to essay questions.

2.2.3 Test quality

This can be viewed in terms of item difficulty index, discrimination index and moderation of the test paper. According to James, (2003) test items that discriminate between students with different degrees of masterly based on the course content are desirable and will improve reliability. The researcher further observed that items with a positive point biserial correlation (r_{pbi}>0.20) will greatly improve reliability. On the other hand items with negative discrimination index will reduce reliability.

A study by Meshkeni et al(2005) on the multivariate analysis of factors influencing reliability in Tehran, revealed that keeping all other variables fixed, the interaction of length and difficulty index resulted in an increase in reliability. The study used descriptive statistics and the data was extracted from item analysis sheets of multiple choice questions and the results were analysed by statistical package for social sciences (SPSS) using multivariate regression analysis for determining the effect of each factor.

The results further showed that moderately difficult items increase reliability eight times more than the difficult items and thirteen times more than easy items(regression coefficient for moderately difficult at 6.4, 0.83 for difficult items and 0.43 for easy items). However the study was only conducted in one institution and covered only one subject therefore the findings may not be representative.

2.2.4 Teacher characteristics

These are looked at in terms of:-Teacher; training on test construction and analysis; educational level; experience and workload.

2.2.4.1 Teacher training on test construction and analysis.

Effective test item writers are trained and not born. Without specific training, most novice item writers tend to create poor quality flawed, low cognitive level test items that test unimportant or trivial content (Haladyna, 2004).

On training teachers are able to write tests with adequate number of items, therefore of high reliability since they provide a representative sample of the behavior being measured and the scores obtained by the test takers are apt to be less distorted by chance factors (Linn&Gronlund, 1995).However a survey carried out by Ngesu-Ngugi (2008) on teacher made tests concluded that the tests were of poor quality an indication that teachers did not have basic skills in test development. This view is supported by a research by Gall(2000) cited by Ogunniyi(2004) which established that ,some teachers generate substandard test items due to ineffective training, with some of the teachers unfortunately not cognizant of the appropriate procedures for test construction.

To the contrary Kubiszyn& Borich,(2007) pointed out that test construction competencies are assumed to have been well taught in training colleges; but the environment in which teachers construct these tests interferes with their quality and therefore their reliability.

A study by Woki Wachira (2012) on assessment practices in Kenyan primary schools found out that ten percent of the teachers never set the exams that they administer to their pupils, but obtain them from commercial test publishers. Using descriptive survey design, data was obtained from tests administered to 520 students in class seven in eleven schools, and questionnaires were administered to twenty two teachers in the same schools.

The survey also found out that fifty five percent of the teachers never had any in-service training in test development since they started teaching, despite the fact that most of the respondents were experienced teachers and had taught for more than ten years.

2.2.4.2 Teacher qualifications

One of the key factors in determining quality examinations results is the availability and the quality of teachers (UNESCO, 2000).This view is supported by Abagi& Odipo (1997) who contend that teacher quality depends on their qualification, experience and level of discipline which in turn determines their level of commitment.

According to Makau & Sommerset, (1990) academic and; professional qualifications of teachers are crucial factors in influencing performance. According to Kiragu& Okunya (2012) as the level of education rises also the reliability and validity gets better. The rise in reliability was attributed to the fact that the teachers become more efficient in test construction.

2.2.4.3 Teacher's experience

A study done in Malaysia by Hnushed as cited by Betts (1986) noted that the length of teachers experience is associated to learner's performance.

World Bank report (1987) on teachers contribution to education sector development, states that the teachers experience is the most consistently positive and significant contributor to learner's academic achievement.

2.2.4.4 Teacher's workload

A study by Ramsuran (2006) in South Africa found out that the major challenge faced by teachers in the effective implementation of assessment policies relate to their enormous workloads as evidenced by the large volumes of paper work required for the recording of assessment information.

Research carried out by the Nigeria Educational Research and Development Council and the Kenya National Examinations Council (2000) revealed that teachers have too many responsibilities and the resources are scarce making it difficult for teachers to carry out sound assessment.

2.3 Factors influencing Difficulty index of commercial test papers.

There are four factors that can affect the difficulty index of commercial test papers; the intended purpose of the results, number of objectives and teacher training on test development.

2.3.1 Number of objectives

Research carried out by Ofo (1994) noted that multiple choice tests with fewer options tend to have better difficulty than those with larger number of options. This view is supported by a study by Olatunji (2009) who investigated difficulty and discrimination indices of Senior Secondary Certificate Examinations (SSCE) multiple choice papers in economics among secondary schools in Nigeria. The repeated measures quasi-experimental design was employed for the study. Purposeful sampling technique was also used to select the study sample comprising of three hundred secondary school form three students, from six educational secondary schools in Kwara state Nigeria.

The instruments used were three, four and five option formats of the 2005 National Examination Council (NECO) and West African Examination Council (WAEC) multiple choice tests in economics. The findings showed that the three option format of NECO multiple choice tests had the best difficulty indices compared with four and five options formats. However submissions of

Costin (1970), Wuddah (1994) and WAEC (1995) which stated that the difficulty index of a test is independent of the format type based on the number of options contradicts the findings of this study.

This view is supported by Suen&McClellan (2003) who posited that, as the number of options in the multiple choice question tests increase the difficulty index decreases with the best difficulty index at 0.67 for three option items, however extreme difficulty index values indicate that the items are not very functional.

2.3.2 Student level

According to Brown (1976) the difficulty index is meaningful only in terms of a particular reference group as an item considered very difficult by one group might be found to be very easy by another. Research by Bharti (2013) found out that when ranking students based on the performance, items with a moderate difficulty index should be selected. The researcher further noted that teachers should aim at getting low difficulty indices as the aim of classroom teaching is not to distinguish between good and bad students but to ensure that all students have learnt the lesson correctly.

According to Abiri (2007), if the goal of measurement is to maximize differences between students, items with difficulties of 0.5 are the best, but a useful rule of thumb is to use items that 50% to 70% of the students can answer correctly. However there should be flexibility in selecting the items depending on the purpose that the test would serve.

2.3.3 Teacher training on test development.

Training of teachers on test development greatly improves the test quality. According to Lehmann (1991) poor quality and flawed test items introduce construct irrelevant variance to the assessment, therefore potentially decreasing student passing rate by increasing the mean item difficulty.

2.4 Factors influencing discrimination index of commercial test papers

There are three factors that influence discrimination indices of test items; student level, number of options and the number of students.

2.4.1 Number of objectives

Research carried out by Olatunji (2009) in Nigeria on discrimination and difficulty indices of senior secondary certificate examinations (SSCE) multiple choice papers in economics found out that the number of options significantly affected the discrimination index of the national examinations council (NECO).However the number of options did not significantly affect the discrimination indices of West African Examinations Council (WAEC) multiple choice test items in economics. This could be attributed to the small sample used in the study and the purposeful sampling technique used in selection of the study sample making the results not to be representative.

The study found out that test items with three option format had the best discrimination index therefore giving room for better content sampling and maximum item effectiveness.

2.4.2 Student level

According to Oyejide (1991) the discrimination power of a test item indicates the strength of each item to distinguish the higher achievers from the low achievers. This view is supported by Abiri (2007) who averred that item discrimination could be derived as an index of correlation between an individual item and the total test score. Therefore the index is used to differentiate between low and high scorers.

According to Nagaraja (2009) when the difficulty index is very small, it indicates that the questions are very difficult; the tested content was not taught well or is difficult for the students to grasp.

2.4.3 Teacher training on test development

Research carried out by Mehrens&Lehmann (1973) found out that the discrimination index reflects the degree to which an item and the test as whole are measuring a unitary ability or attribute. Values of the coefficient tend to be lower for tests measuring a wide range of content areas than for more homogeneous tests. The researchers further found out that test items with low discrimination indices are often ambiguously worded and should be examined, while a negative discrimination index may indicate that the item was miskeyed, so that students who knew the correct answer tended to choose an unkeyed but correct, response option.

Research by Nagaraja (2009) on the level of difficulty and discrimination indices in multiple choice questions in summative test in Malaysia found out that, items with poor discrimination index and too low or too high difficulty index should be reviewed by the respective content experts. The researcher further observed that this serves as an effective feedback to the respective departments in a school about the quality control of various tests.

2.4.4 Relationship between difficulty and discrimination indices of test items.

Research carried out by Si-Mui Sim& Raja I Rasiah (2006) on the relationship between the item difficulty and discrimination indices in true/false type multiple choice questions of a Para-clinical multidisciplinary paper in Malaysia, found out that multiple choice questions that demonstrated good discriminating potential tend to be moderately difficult items and the moderately to very difficult items are more likely to show negative discrimination.

The research further noted that a wide scatter of discrimination values for questions with similar level of difficulty may reflect that some extent of guessing practices occurred.

Studies by Surrender Singh (2014) on the relationship between difficulty and discrimination indices using Pearson correlation formulae found out that item discrimination increased with increase with difficulty level but decreased for very easy and very difficult items.

2.5 Factors influencing student performance in summative evaluation.

The variables influencing performance are grouped into four categories; Student characteristics, Teacher characteristics, and Home based factors

2.5.1 Student characteristics

This is looked at in terms of gender and testwiseness.

2.5.1.1 Student gender

The traditional female stereotypes emphasize dependence, personal relationships and feelings (Smith, 1992).The researcher further posits that these stereotypes encourage girls to adopt self conceptions and values that reduce the importance of interests and achievement in science and

mathematics unlike boys. The self conceptions among females lead to a pattern of internalized helplessness with respect to science with failure credited to lack of ability and success to luck.

A study by Adeyemi (2001) in Botswana found out that girls performed better than boys in English composition. Using pre test research design of Campbell and Stanley (1966), the study population comprised of 41 entrants purposefully chosen from year 2004 form one intake. The data collection instrument was five composition format tests administered and marked by the researcher and analysed by use of a t test adjusted separately for gender.

Although research on females performing better in English composition is scanty, the results agreed with those of ward (1952) who submitted that male students generally have lower average scores than females on English test. However the study population was too low for the results to be generalized furthermore the sample was purposively selected which may have influenced the results.

Another study by Magomotsi et al (2012) on the gender factor in the prediction of performance in senior secondary school in Botswana found out that gender did not have a significant influence on the prediction of BGCSE grades using coursework grades and forecast grades among secondary students in Botswana. Descriptive research design was employed with a study sample comprising of 2292 students (1432 male and 860 females) purposefully selected for the study. Secondary data were retrieved with permission from Botswana examinations council (BEC) academic records.

The data was then analysed by carrying out Pearson correlation and regression analysis using Statistical Package for Social Sciences (SPSS) version sixteen. The study also found out that for the combined data for all four years, a significant gender based differential prediction was observed with the female having over prediction and the males under prediction. However when the prediction was done for each of the four years the gender based significant differential prediction was not observed.

2.5.1.2 Student Testwiseness

Students should be able to read directions and questions, work independently, and most importantly write correct responses. The masterly of these skills is vital for doing well on a test (Anastassi, 1976; Sarnacki, 1979).

According to Mehrens&Lehmans (1973) most teachers are relatively naive of the test wiseness principles and even if they are familiar with it, their test writing skills are not as precise as the professional test constructor therefore item flaws tend to occur.

In testing the hypothesis that multiple choice tests reward the test wise students Alker, Carlson&Hermann (1969) found out that test wiseness was positively and significantly correlated with both multiple choice questions performance and ability to recognize item ambinguity.The same findings were observed by Rowley (1974) and Ebel (1972).

According to Thorndike (1951) and Stanley (1971) faulty or flawed test items used in tests provide cues which introduce variance in the results other than item content or random error. These cues may increase or depress the test takers test score and reliability.

2.5.2 Teacher characteristics

These include Teacher gender and Teachers Qualifications (level of education).

2.5.2.1 Teacher Gender

Despite students beginning kindergarten on equal footing, by the time male and female students are transitioning from middle school to high school, a clear academic disparity has developed in which female students typically outperform their male peers in subjects of reading and writing while male students outperform female students in areas of science and mathematics (McCrea, 2010).

A study by Awodun et al (2015) on the influence of teachers gender in secondary schools students performance in physics in Ekiti state Nigeria found out that there a significant difference in the performance mean scores of students taught by male and those taught by female physics teachers. Using expos-facto descriptive survey design, random sampling was done to collect the 200 year two secondary school physics students drawn from ten secondary schools to collect the primary data. ANOVAs statistical analysis was used to analyse three null hypotheses formulated for the study.

The study also found out that there is a significant difference in the performance mean score of boys taught by male teachers, and that of those taught by female teachers, and there was a significant difference in performance mean scores of girls taught by male teachers and that of those taught by female physics teachers. However the sample collected per school was too small for the results to be generalized.

The study results were supported by Okoro (2012) and Rodriguez (2002) whose studies revealed that teacher gender has a significant influence on the academic performance of the learner.

However studies by Holmlum&Sund's(2003);Kreig(2005);Marcus &Robert(2012) did not support the idea that same gender teacher had a positive causal impact on student outcome but noted that though a statistically significant relationship may be observed, the magnitude of the effect is too small.

2.5.2.2 Teacher qualification

One of the key factors in determining quality examination results is the availability and the quality of teachers (UNESCO, 2000). This view is supported by Abagi and Odipo, (1997) who contend that teacher quality depends on their qualification and level of experience.

A study by Ojerinde & Falayajo (1984) noted that most teacher education programmes in Nigeria do not provide adequate training for prospective teachers on assessment strategies. The study further proposed that further on effort to improve teacher skill in construction of assessment items, scoring, interpreting and communicating scores should be done through in-service and preservice courses.

According to Nderitu (1999) teacher attendance of in-service training and teacher academic qualification has a strong influence on performance.

Research by Fakeye (2012) in Oyo state, Nigeria on teacher qualification and subject masterly as predictors of student achievement in English language found out that teachers teaching qualifications has a significant relative contribution to students academic achievement. The study adopted a descriptive research design of survey type with the study sample comprising of twenty

senior secondary schools randomly sampled each with fifty senior secondary II students. All the English language teachers in the selected schools participated in the study.

Questionnaires were used as the data collection tools, and the data collected was analysed using frequency and multiple regression analysis. The findings were consistent with those of Komalafe (1989); Jones (1997); Strauss&Vogt (2001); Hariss &Sass (2008) and Oyo (2008) who claimed that teacher's qualifications and experience are strong determinants of student academic achievements.

However a study by Ruth Zuzovsky (2007) on the teachers qualifications and their impact on student achievement in Tel Aviv, Israel found out that many of the teacher variables regarded as qualifications such as advanced academic degrees, majoring in the field of teaching and years of experience, which were adopted as reform targets in teacher education programs and as criteria for remuneration had only marginal and statistically non significant positive effects on student achievement.

2.5.3 Home based factors

These include socio economic status, parental involvement and level of education, and family type.

2.5.3.1 Parental involvement and level of education

Parental involvement in a child's early education is consistently found to be positively associated with the child's academic performance (Hara&Burke, 1998; Hill & Craft, 2003). According to

Henderson(1994) the most accurate predictor of a student's achievement in school is not family income or social status but the extent to which the students family create a home environment that encourages learning, express high but realistic expectations for their children achievement and future careers and finally become involved in their children's education at school and community.

According to Philipson (2007) parental expectations that stem from their beliefs and practices have been a dominant predictor of achievement. Parental expectations and expressions of confidence in their children's ability and talent are paramount in promoting the potential that the children have in academic achievement (Chan, 2009).

Studies by Obanya& Ozewu (1988); Nabbumba (1988); Mugisha (1991); Nderitu (1999); Owen (1999) and Ayoo (2002) found out that educational attainment of parents is directly related to the educational of their children. These findings are supported by a study by Zappala (2002) in Australia who found out that, families where parents are educated foster a higher level of achievement in their children because of providing psychological support for their children. Another study positing this view by Ogoye (2007) in Mombasa Kenya showed that illiterate parents were unable to assist their children in homework; therefore they end up performing poorly in school.

However Cullen (1969) have a differing opinion and reported that students having illiterate parents performed comparatively better than those from literate homes due to their concerted effort to use education to escape from the poverty cycle.

2.5.3.2 Family type

Studies by Fadeiye (1985) and Uwafo (2008) found out that differences in academic performance of children exist in terms of those from single parent families and those from two parent families. They further observed that where both parents are present, they support each other to give the necessary tools for educational advancement, however for a single parent the support is lacking resulting in poor academic performance in part of the child.

According to a study by Bankston and Caldes(1998),single headed families are six times more likely to be poor compared to families having two parents, therefore students from such families will not only have to deal with absent parent figure but also financial constraints, and this eventually affects the school performance (Mulkey et al,1992). A study by Ushie et al (2012) on the influence of family structure on student's academic performance in public secondary schools in Agege, Nigeria found out that there was no significant difference in the academic performance of students from single parent families and those from two parent families.

However the stratified simple random sampling was done to collect the study sample which could have influenced the study results furthermore the study sample was too small for the findings to be generalized.

2.5.3.3 Social Economic factors

According to Eshiwani (1983) the environment in which students live, influences and shapes their aspirations, self esteem and motivation, therefore they can enhance or hinder their educational performance. This view is supported by Shittu (2004) whose study found out that

poor parental care with gross deprivation of social and economic needs of children usually yield to poor academic performance.

A study by Nzomo, Kariuki and Guantai (2001) established a positive correlation between the social economic status of standard six pupils and the level of their learning achievement in Kenya. From the results it was also observed that an improvement in social economic status of the pupil led to an increase in the mean scores in learning achievement due to these families having the ability to provide their children with the necessary facilities and materials pertinent in improving performance.

A study by Cecilia & Patrick (2006) in South Africa on the social background factors affecting the academic success of students found out that the family structure and living space affect the academic performance of the learner either positively or negatively.

According to Duncan (1989) in developed countries, there is a significant relationship between student academic performances and economic background of their parents. Students from deprived homes despite their mental potential tend to go to cheaper low performing schools whereas their counterparts whose ability could be average go to well performing schools and emerge successful while those from poor families perform dismally.

This premise is supported by a study by Dermis (2006) which attributed poor performance of Somali pupils in United Kingdom to overcrowded accommodation, making the students to

typically have no space at home to organize their learning materials and may experience excessive noise.

This view is supported by the Kenya Economic Survey (2008) which indicated that 46.8% of Kenyans live below poverty line, a clear indication that majority of the parents are not able to provide for school basic needs, therefore a conducive environment for the student to study is not availed, eventually affecting their performance.

On the other hand a study by Nderitu (1999) found no correlation between social economic background and performance of the learner, but agree that students from poor background are regularly sent home from school due to non payment of fees.

2.6 Theoretical Framework

The research adopted the Classical test theory (CTT) which was founded by Charles Spearman in 1904 and it comprises three components: the observed score, true score and the random error mathematically depicted in an equation as: $X = T + E$. Where X is the observed score of a student on any test,T is the expected value of the observed score received on several such tests of equal difficulty when run an infinite number of times and E is the difference between X and T and it is related to the standard error (Edward.C,1979).

An important concept of Classical test theory is that of standard error of measurement(SEM) which is the standard deviation of errors of measurement that are associated with test groups from a particular group of examinees. It is also thought as the determination of the amount of

variation or spread in the measurement error for a test (Norick, 1966). A person's true score is equal to the average of the observed scores. However it is difficult to know the measurement error therefore all standardized tests have an associated standard error of measurement which is expressed in standard deviation units, and is directly related to the reliability of the test.

From this it can be concluded that the smaller the standard error of measurement, the higher the reliability and better scores are obtained. For this to be possible the error in the Classical Test Theory is always assumed to be random and non-systematic which are attributed to both internal and external factors to the test taker. External factors include those attributed to the Test items that might be poorly created or inadequate testing conditions. Internal factors can result from lack of concentration, fatigue and stress, and may considerably contribute to the random error (Monar, 2014).

This theory also deals with both item and test level statistics. At the item level, it deals with item difficulty and discrimination. On the other hand at the test level, it deals with the reliability of a test (Monar Nasir, 2014).In conclusion for the estimation of the true score using the theory, four fundamental assumptions have been made first, and the observed score of a person is comprised of the true score and random error. Second, the expected value of any observed score is the person's true score. Third, the covariance of error components from two tests is zero in the population and finally errors in one test are uncorrelated with true scores in another (Monar, 2014).

2.7 Conceptual framework

Figure 2.1 Conceptual framework.
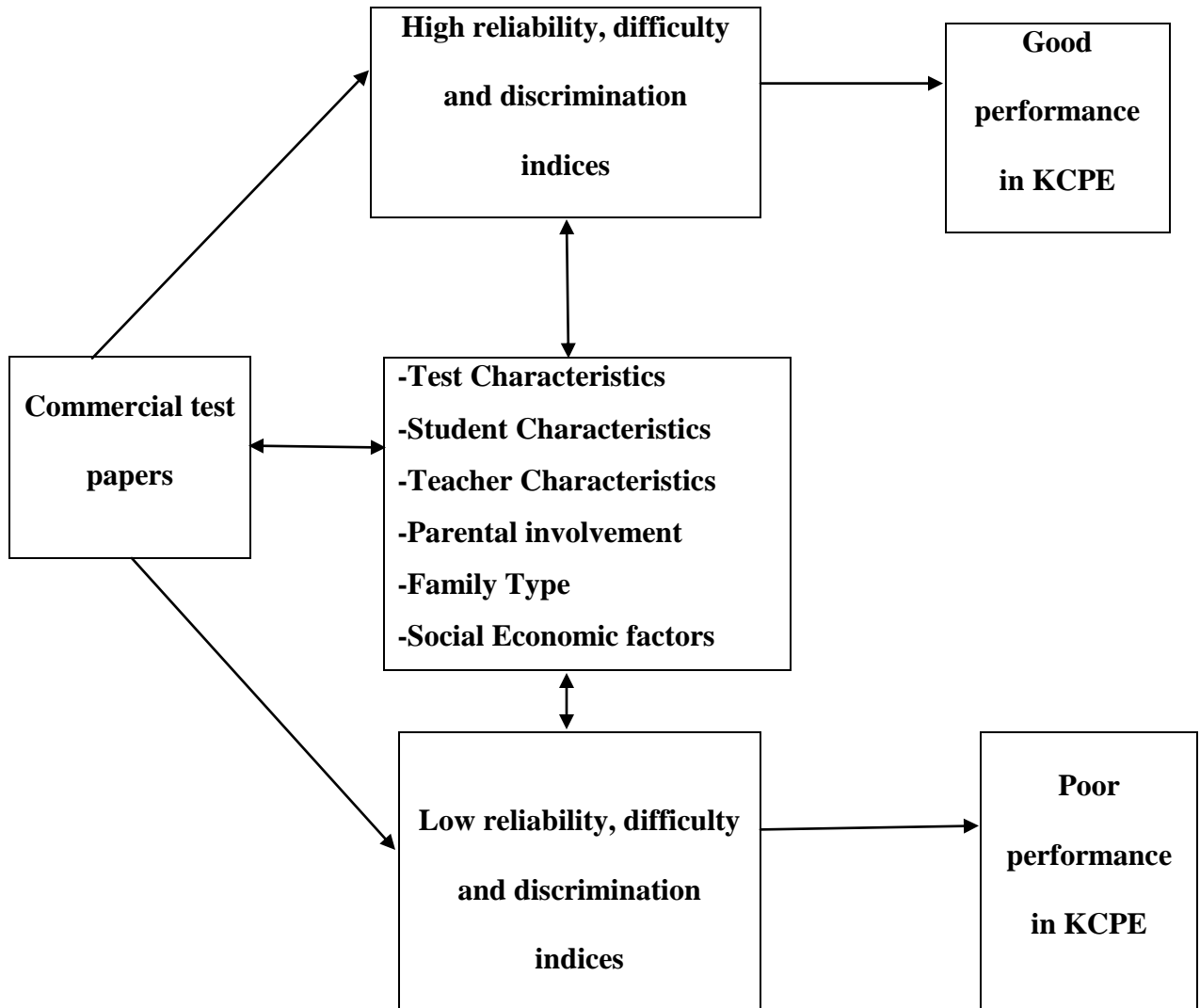
This conceptual framework illustrates the relationship between the variables addressed by this study. Research by Suruchi (1991), Si-Mui Sim (2006), Hotiu (2006), Olatuji (2009) and Chellamani (2013) suggest that test items with good positive discrimination power and moderate difficulty are ideal for a good test. However as in a class test, very difficult or very easy items may be retained despite of their poor discrimination power because the purpose is to test the

content masterly and attainment of the set objectives and not to discriminate the superior and inferior students (Suruchi et al, 1991).

CHAPTER THREE

RESEARCH METHODOLOGY

3.1 Introduction

The research methodology used in this study included the research design, target population, sample size, sampling techniques, research instruments, validity and reliability of instruments, data collection and analysis procedures.

3.2 Research Design

A descriptive survey research design was used to evaluate the quality of commercial science test papers in terms of reliability, difficulty index, discrimination index and their correlation to student performance in Kenya Certificate of Primary Education (K.C.P.E).

The study used statistical data on K.C.P.E performance drawn from the sampled schools records. Teachers sampled as key informants provided primary data used in the study, while commercial test papers in science subject were administered to the sampled students and then analysed to provide the secondary data.

3.3 Target Population

The study population was drawn from 3 public primary schools in Limuru Sub County comprising of 600 class eight candidates, 39 teachers teaching science subject in class eight in public primary schools in Limuru Sub County, who will served as key informants.

3.4 Sample size and sampling technique

A sample of 200 class eight students was selected using stratified random sampling on the basis of gender out of the population of 600 class eight students in three public primary schools in Limuru Sub County. 38 science teachers teaching science subject in class eight were purposefully sampled.

3.5 Research instruments

The study used questionnaires consisting of closed and open ended questions to collect the primary data. They were administered by the researcher through pick and drop method in order to avoid chances of data manipulation.

The researcher also randomly collected commercial science test papers from the 3 selected public primary schools and one paper which was common in all the schools was therefore chosen to be administered to 200 class eight students sampled in the three schools. Each commercial test paper consisted of 50 test items each with a stem and four objectives, the students were required to select the correct answer. The results of these tests were then be analysed for difficulty, discrimination indices and reliability coefficient.

3.5.1 Data collection procedure

The researcher personally visited the respective schools in which teachers and students were sampled. After getting the school heads permission to conduct the study, the researcher then distributed the questionnaires to the sampled teachers to complete. In the three schools selected, the student sat for the selected commercial test papers which were then scored by their teachers and the results tabulated.

3.5.2 Validity of Research Instrument

According to Mugenda and Mugenda (1990) Validity is the accuracy and meaningfulness of inferences, which are based on the research results. It is the degree to which results obtained from the analysis of the data actually represent the phenomenon under study.

According to Kerlinger (1993) to ensure validity of the instruments used in a study, expert judgment should be sort, consultation and a decision drawn on the instruments. In view of this, the researcher will sought expert opinion in assessing the content validity notably from the researcher's supervisor, to assess the concept the instrument was trying to measure and determine whether the set items accurately represent the concept under study.

3.5.3 Reliability of the Research Instrument.

According to Kothari, (2004) reliability is a measure of the degree to which a research instrument yields consistent results after repeated trials. To test the reliability of the instrument, split half method was employed whereby the answered questionnaires were scored manually and a comparison between the scores obtained made.Pearsons product moment formulae of test - retest was employed to compute the correlation coefficient in order to establish the extent to which contents of the questionnaire were consistent in eliciting the same responses(Orodho,2003).According to Mugenda and Mugenda (1990) a reliability coefficient of 0.80 or more indicates a high reliability of the data.

To enhance the reliability of the instrument, a pilot study was conducted in two schools which were not included in the main study. The aim of pre testing was to gauge the clarity and relevance of the instrument items so that those items found to be inadequate for measuring the

intended variables were either discarded or modified to improve the quality of the research instruments.

3.6 Data analysis techniques

Primary data was obtained from teacher's questionnaires that were administered to teachers in charge of science subject in class eight in the selected public primary schools and students' questionnaires that were administered to the sampled class eight students in three public primary schools. Secondary data was obtained from science commercial test papers bought from commercial test publishers and administered to the sampled students during the term.

After scoring the test items, test scores were arranged in descending order using MS-EXCEL. To carry out the item analysis, top 27% scorers and bottom 27% scorers of the 200 examinees were selected. Upper and lower 27% were used as "these values maximize the difference in normal distributions while providing enough cases for analysis" (Wiersman& Jurs, 1990). Middle 46% of the test scores were excluded as they behave in a similar pattern contributing insignificantly to discriminate the performance by students. Nunnally (1972) suggested 25% while SPSS (1999) uses the highest and lowest one third (33%).

Index of difficulty for each test item can be calculated as

$P = (Ru + Rl) / (Nu + Nl)$

P= item difficulty

Ru = the number of students in the upper 27% who responded correctly

Rl = the number of students in the lower 27% who responded correctly

Nu = the number of students in the upper group

Nl = the number of students in the lower group, (Ebel and Frisbie, 1986).

Discrimination power is estimated using the following formula:

Discrimination power (D) = RU-Rl/NU (or) Nl, (Ebel and Frisbie, 1986).

The quantitative data was then coded and run through the statistical program for social sciences (SPSS) version 17 renamed PASW, and presented using tables, frequencies, percentages statistical measures of relationship between the dependent and independent variables.

Qualitative data was analysed using descriptive statistics. The results obtained were then used to draw conclusions and make recommendations.

.

DATA ANALYSIS, RESULTS AND DISCUSSION

4.1 INTRODUCTION

This chapter deals with the data analysis and interpretation of the research findings. The results have been grouped under four sub headings. Factors influencing:-reliability of commercial test papers, Difficulty indices of commercial test papers, Discrimination indices of commercial test papers and Performance in summative evaluation.

4.2 Factors affecting reliability

4.2.1 Test duration

Table 4.2.1 test duration

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 hour	3	7.9	7.9	7.9
	1hr 40 min	35	92.1	92.1	100.0
	Total	38	100.0	100.0	

The table 4.2.1 shows the respondents view on the most appropriate duration of the test to enable learners to answer the entire questions.92.1% said that 1hr.40minutes was the most appropriate time, while 7.9% stated that 1hour was appropriate.

Figure 4.1 showing the test duration
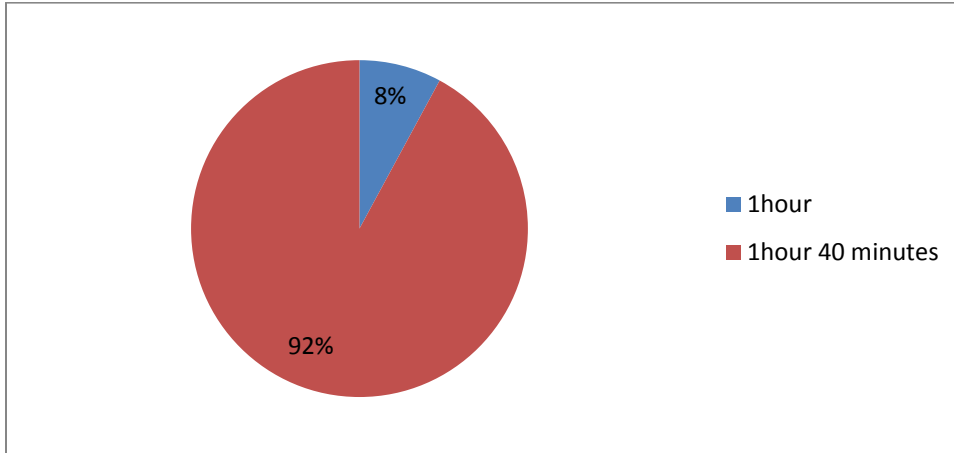
4.2.2 Test Quality

Table 4.2.2 test quality

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Good quality	3	92.1	92.1	92.1
	Poor quality	35	7.9	7.9	100.0
	Total	38	100.0	100.0	

Table 4.2.2 show that 92.1% of the respondents said that the test paper was of good quality while 7.9% of the respondents stated that the test paper was of poor quality.

Figure 4.2 showing test quality
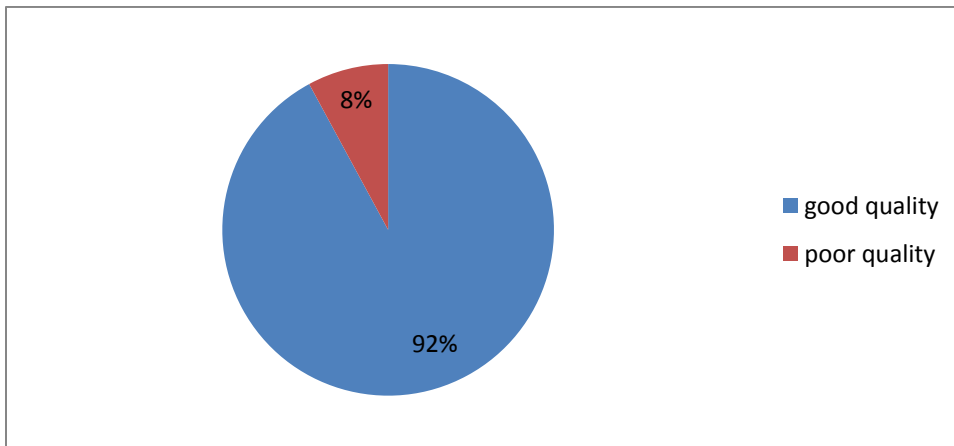
The reliability coefficient was also calculated using Kunder Richardson formula 21 which yielded $P_{KR21}=0.8708$. This indicated that the science commercial test paper was highly reliable.

4.2.3 Teacher characteristics

4.2.3.1 Training in test construction.

Table 4.2.3 Training in Test Development

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strongly disagree	26	68.4	68.4	68.4
	disagree	8	21.1	21.1	89.5
	neutral	2	5.3	5.3	94.7
	strongly agree	2	5.3	5.3	100.0
	Total	38	100.0	100.0	

Table 4.2.3 show that 89.5% of the respondents had not attended any in-service training on test development, 5.3% were not sure whether the training they underwent involved test development while 5.3% have attended training.

Figure 4.3 showing training in test development
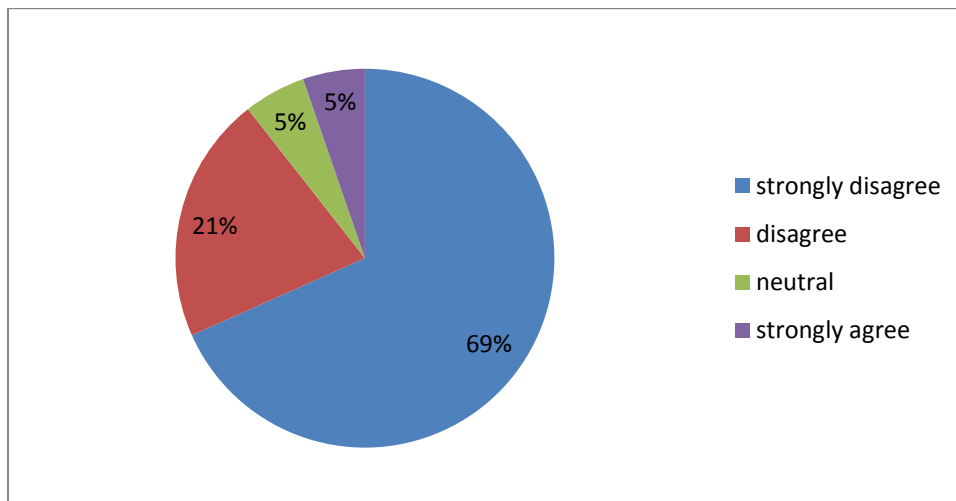
4.2.3.2 Educational level (Professional Qualification)

Table 4.2.4 Educational level

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	P1	14	36.8	36.8	36.8
	diploma	7	18.4	18.4	55.3
	B.Ed.	17	44.7	44.7	100.0
	Total	38	100.0	100.0	

 Table 4.2.4 show that 18.4% of the respondents teaching science subject in class eight were

diploma, 44.7% B.Ed holders while 36.8% were certificate holders

Figure 4.4 showing educational level
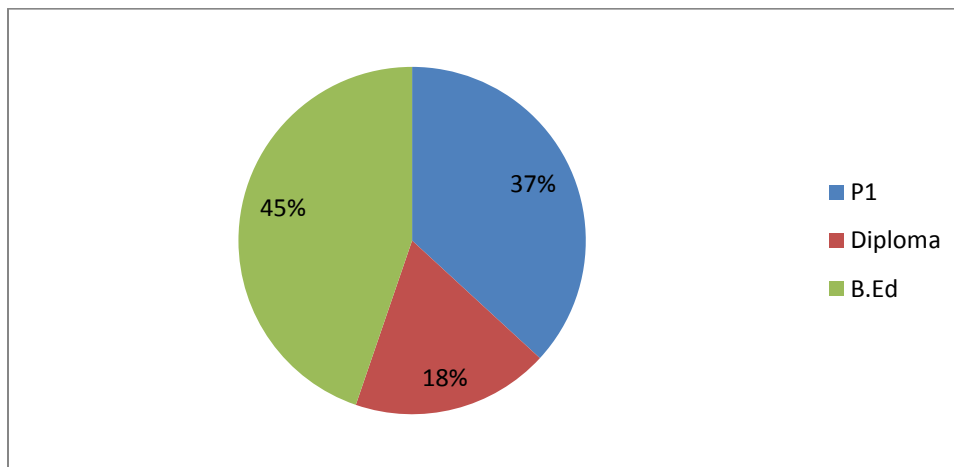
4.2.3.3 Teaching Experience

Table 4.2.5 Teaching Experience

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2-10 years	10	26.3	26.3	26.3
	above 10 years	28	73.7	73.7	100.0
	Total	38	100.0	100.0	

Table 4.2.5 shows that no respondent had an experience of less than one year, 26.3% had an experience of between 2-10 years while 73.7% had an experience of more than 10 years.

Figure 4.5 showing teaching experience
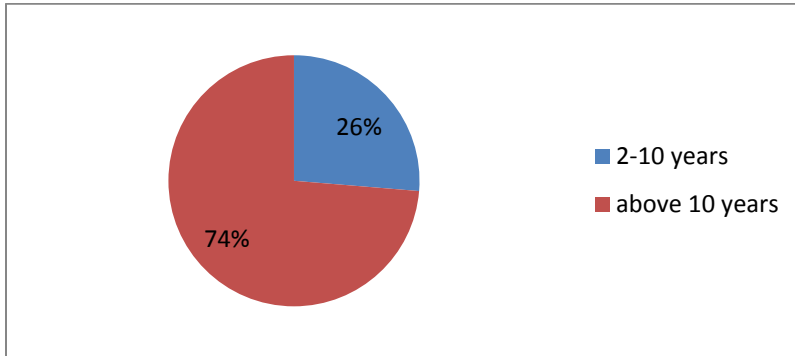
4.2.3.4 Teachers workload

Table 4.2.6 Lessons per week

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	20-29	10	26.3	26.3	26.3
	30-39	28	73.7	73.7	100.0
	Total	38	100.0	100.0	

Table 4.2.6 show that 26.3% of the respondents had between 20 -29 lessons in a week, while 73.7% of the respondents had between 30-39 lessons per week indicating that the workload was very high.

Figure 4.6 showing lessons per week
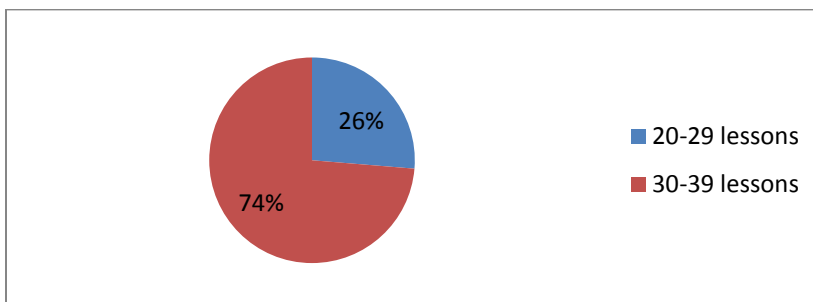
4.3 Difficulty index

Table 4.3.1 difficulty indices

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.13	1	2.0	2.0	2.0
	.14	1	2.0	2.0	4.0
	.17	1	2.0	2.0	6.0
	.23	1	2.0	2.0	8.0
	.24	1	2.0	2.0	10.0
	.26	2	4.0	4.0	14.0
	.27	1	2.0	2.0	16.0
	.30	1	2.0	2.0	18.0
	.31	1	2.0	2.0	20.0
	.32	1	2.0	2.0	22.0
	.33	1	2.0	2.0	24.0
	.36	1	2.0	2.0	26.0
	.37	1	2.0	2.0	28.0
	.44	3	6.0	6.0	34.0
	.45	2	4.0	4.0	38.0
	.46	1	2.0	2.0	40.0
	.48	2	4.0	4.0	44.0
	.50	2	4.0	4.0	48.0
	.51	2	4.0	4.0	52.0
	.52	1	2.0	2.0	54.0
	.52	1	2.0	2.0	56.0
	.53	2	4.0	4.0	60.0
	.56	2	4.0	4.0	64.0
	.57	2	4.0	4.0	68.0
	.60	1	2.0	2.0	70.0
	.61	2	4.0	4.0	74.0
	.63	1	2.0	2.0	76.0
	.64	1	2.0	2.0	78.0
	.66	1	2.0	2.0	80.0
	.67	1	2.0	2.0	82.0
	.68	2	4.0	4.0	86.0
	.69	2	4.0	4.0	90.0
	.70	1	2.0	2.0	92.0
	.71	1	2.0	2.0	94.0
	.77	1	2.0	2.0	96.0
	.79	1	2.0	2.0	98.0
	.86	1	2.0	2.0	100.0
	Total	50	100.0	100.0	

Table 4.3.1 show that the difficulty indices range from 0.13 to 0.86 with an average of 0.49.

Three test items had difficulty index values below 0.2 while one test item had a value above 0.8.

Table 4.3.2 difficulty indices distribution

Difficulty indices	Frequency	Percent	Cumulative percent	Quality
Below 0.2	3	6	6	Very difficult
0.2-0.5	21	42	48	Good
0.5-0.8	25	50	98	Best
Above 0.8	1	2	100	Very easy
Total	50	100		

Table 4.3.2 shows that 6% of the test items had a difficulty index value below 0.2,42% of the items had indices between 0.2 and 0.5,50% of the items had indices between 0.5 and 0.8 while 2% of the items had indices above 0.8.

Figure 4.7. Histogram showing item frequency for each difficulty index.
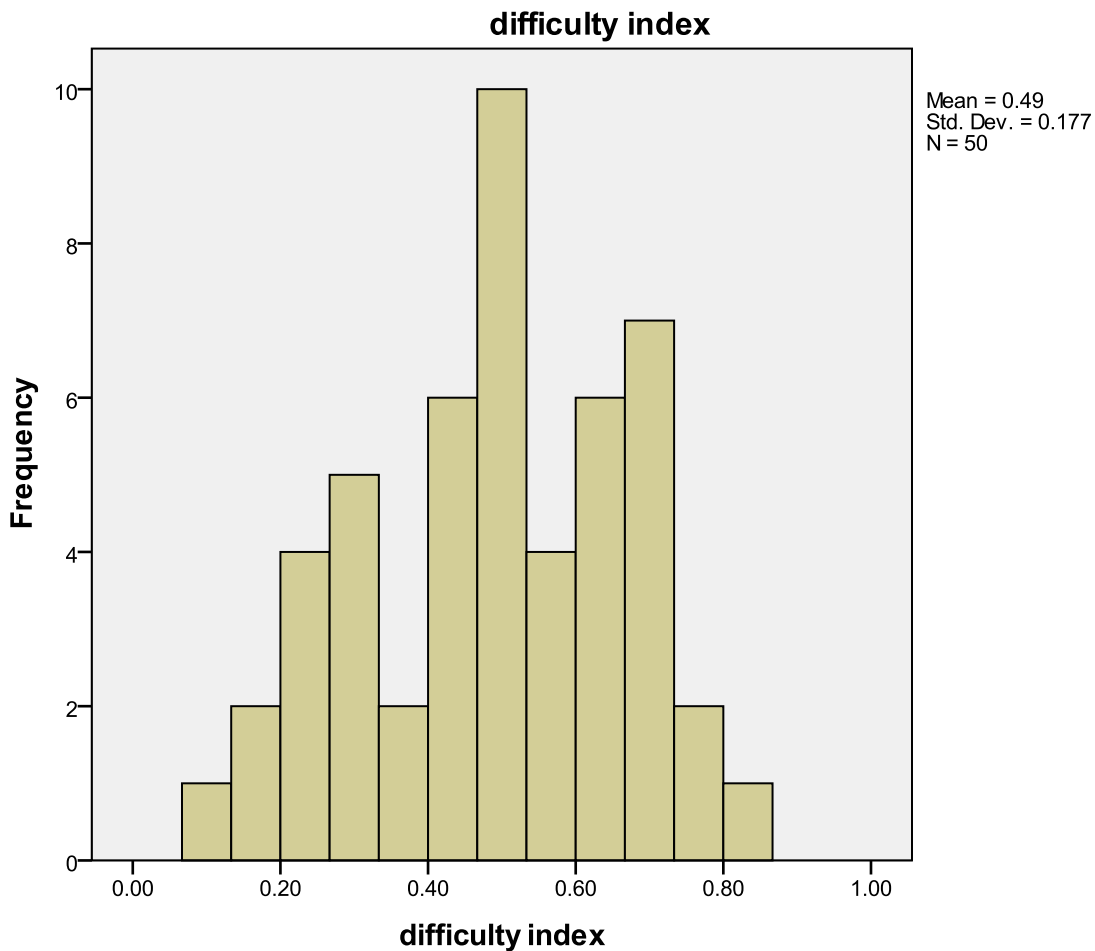
4.3.1 Factors influencing Difficulty index.

4.3.1.1 Number of objectives

Table 4.3.3 Number of multiple choice options

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2 options	0	0.0	0.0	0.0
	3 options	0	0.0	0.0	0.0
	4 options	38	100.0	100.0	100.0
	Total	38	100.0	100.0	

Table 4.3.3 show that none of the respondents use 2 and 3 options objective questions in their internal assessment while 100% of the respondents use 4-option objective questions.

Figure 4.8 showing the number of multiple choice options
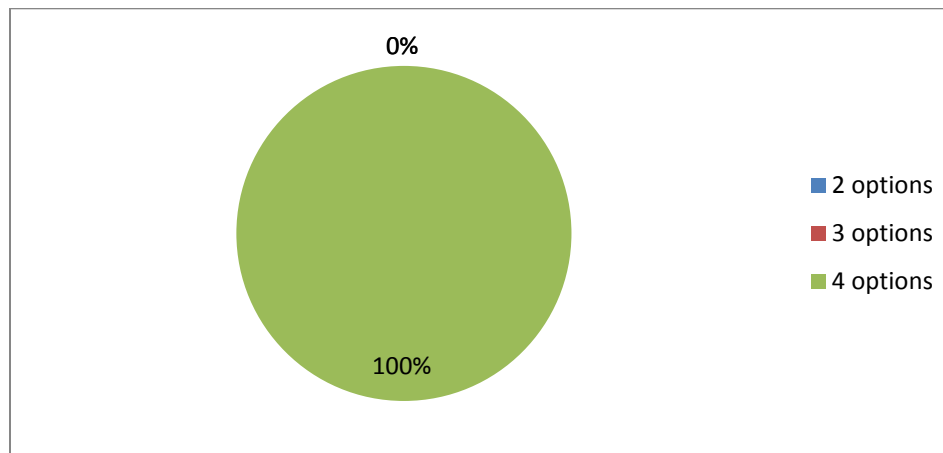
4.3.1.2 Intended purpose

Table 4.3.4 use of test results

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	rank the students	29	76.3	76.3	76.3
	reinforce teaching	9	23.7	23.7	100.0
	Total	38	100.0	100.0	

Table 4.3.4 show that 23.7% of the respondents use the test results to reinforce teaching while 76.3% of the respondents use the test results to rank the students.

Figure 4.9 showing various uses of test results.
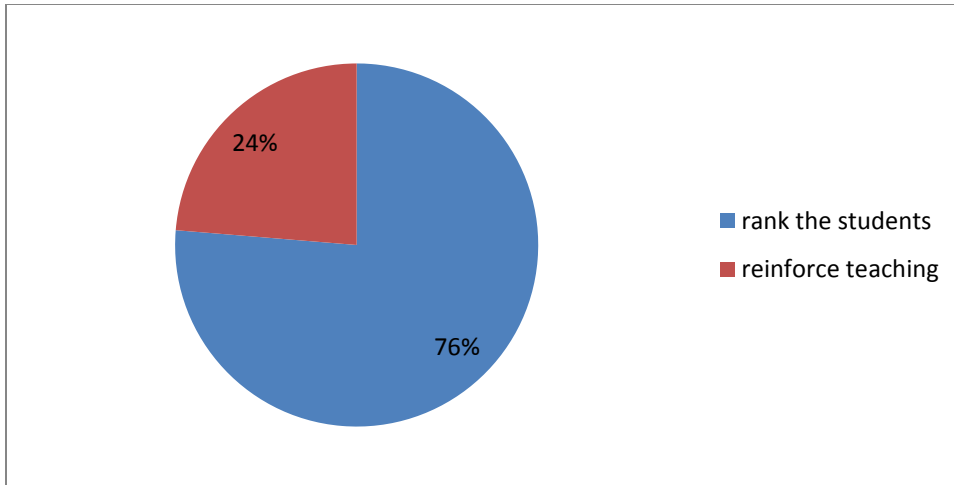
4.3.1.3 Teacher training on test development

Table 4.3.5 Teacher training on test development

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Never attended	34	89.5	89.5	89.5
	Not sure	2	5.2	5.2	94.7
	attended	2	5.2	5.2	100.0
	Total	38	100.0	100.0	

Table 4.3.5 show that 89.5% of the respondents had never attended an in-service training on test development, 5.2% had no idea on what training on test development entails while 5.2% had attended the in-service training.

Figure 4.10 showing the attendance of teachers to test development training
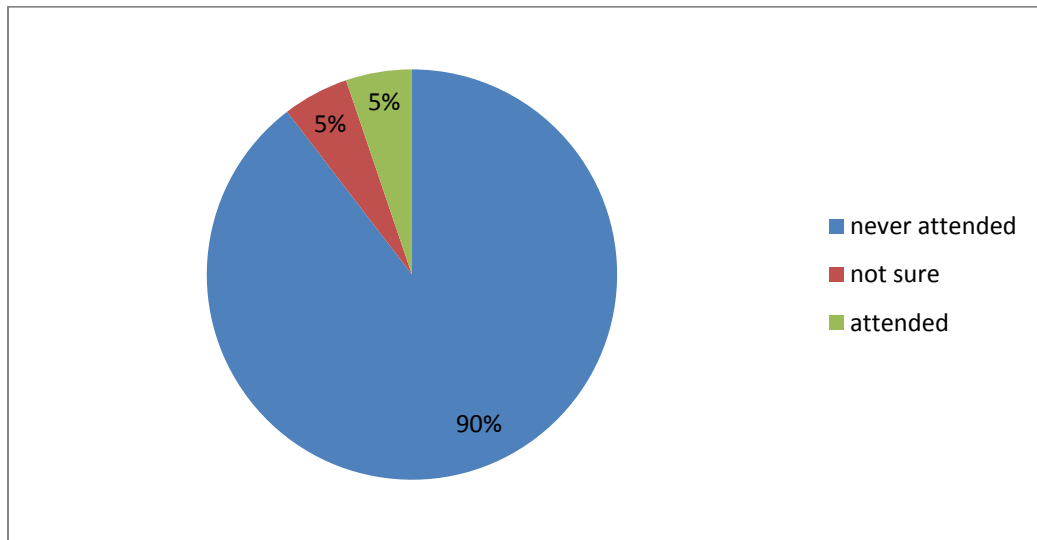
4.4 Discrimination index

Table 4.4.1 shows that discrimination indices range from0.0 to 0.76 with an average of 0.39.
70% of the items had an index of 0.30 and above while 10% of the tests items had a
discrimination index below 0.20.

Table 4.4.1 Discrimination indices

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	1	2.0	2.0	2.0
	.05	1	2.0	2.0	4.0
	.09	1	2.0	2.0	6.0
	.11	1	2.0	2.0	8.0
	.15	1	2.0	2.0	10.0
	.20	1	2.0	2.0	12.0
	.24	1	2.0	2.0	14.0
	.26	3	6.0	6.0	20.0
	.27	1	2.0	2.0	22.0
	.28	4	8.0	8.0	30.0
	.30	4	8.0	8.0	38.0
	.33	1	2.0	2.0	40.0
	.35	1	2.0	2.0	42.0
	.37	1	2.0	2.0	44.0
	.39	1	2.0	2.0	46.0
	.41	2	4.0	4.0	50.0
	.42	2	4.0	4.0	54.0
	.43	2	4.0	4.0	58.0
	.46	1	2.0	2.0	60.0
	.48	6	12.0	12.0	72.0
	.50	3	6.0	6.0	78.0
	.52	1	2.0	2.0	80.0
	.54	1	2.0	2.0	82.0
	.56	1	2.0	2.0	84.0
	.57	3	6.0	6.0	90.0
	.59	2	4.0	4.0	94.0
	.65	1	2.0	2.0	96.0
	.69	1	2.0	2.0	98.0
	.78	1	2.0	2.0	100.0
	Total	50	100.0	100.0	

Table 4.4.2 Discrimination indices distribution

		Frequency	Percent	valid Percent	Quality
Valid	Above 0.40	27	54	54	Excellent
	0.30-0.39	8	16	16	Good
	0.20-0.29	10	20	20	Average
	Below 0.20	5	10	10	Poor
		50	100.0	100.0	Very poor

Table 4.4.2 shows that 10% of the items had a discrimination index below 0.2 therefore should be adjusted or discarded, 20% had indices between 0.2 and 0.29, 16% had indices between 0.3 and 0.39 while 54% had indices above 0.40.

Figure 4.11. histogram showing item frequency for each discrimination index.
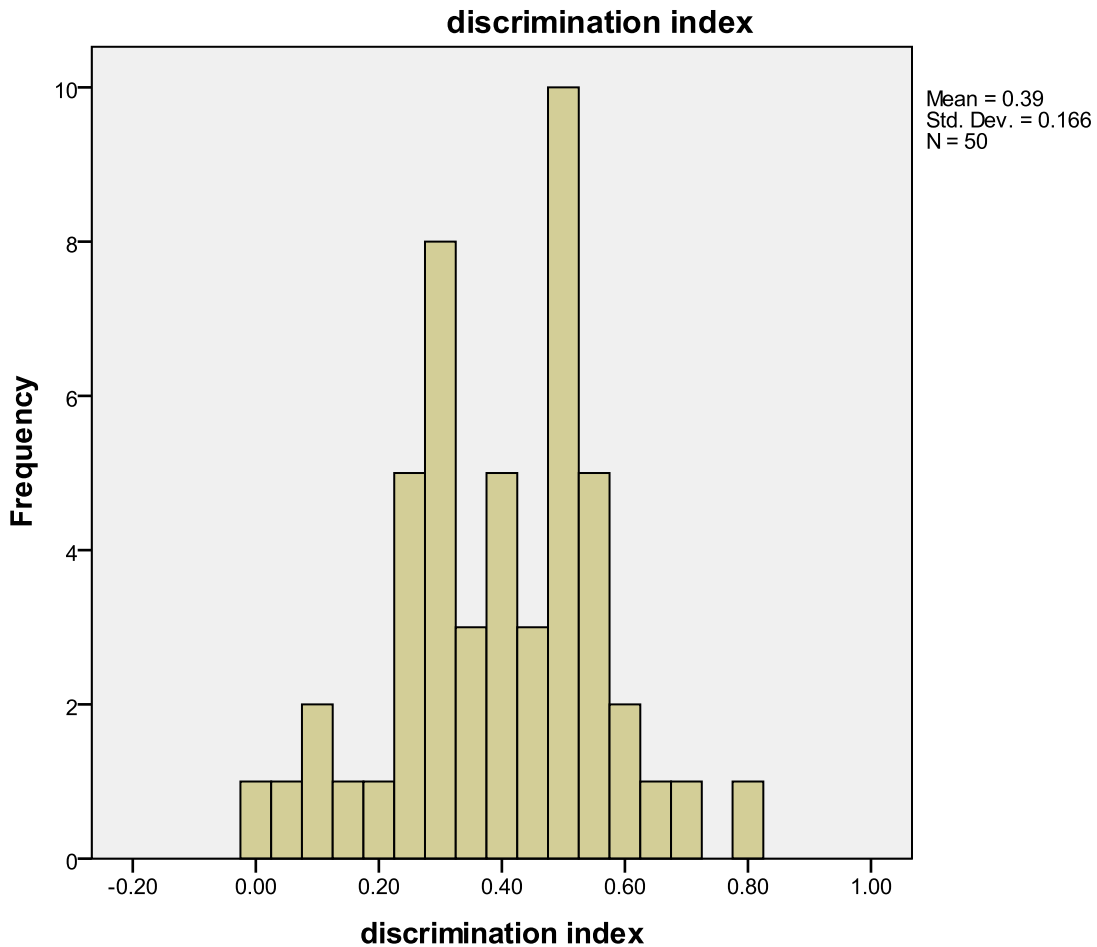
4.4.1 Factors influencing discrimination index

4.4.1.1 Number of objectives

Table 4.4.2 number of objectives

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2 options	0	0.0	0.0	0.0
	3 options	0	0.0	0.0	0.0
	4 options	38	100.0	100.0	100.0
	Total	38	100.0	100.0	

Table 4.4.2 show that none of the respondents use 2 and 3 options objective questions in their internal assessment while 100% of the respondents use 4-option objective questions.

Figure 4.12 showing responses to number of objectives.
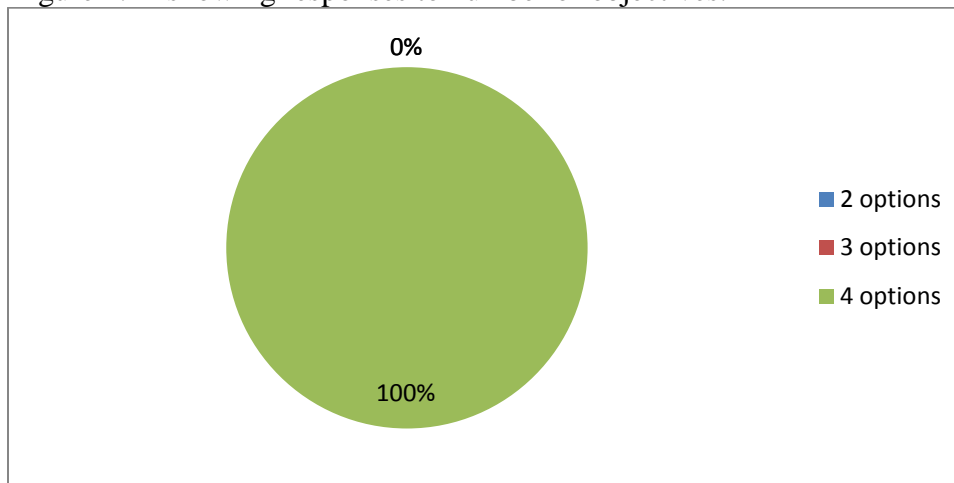
4.4.1.2 Student level

Table 4.4.3. Use of test results.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	rank the students	29	76.3	76.3	76.3
	reinforce teaching	9	23.7	23.7	100.0
	Total	38	100.0	100.0	

Table 4.4.3. Show that 76.3% of the respondents use the test results to rank the students while 23.7% use the test results to reinforce teaching depending on the student level.

Figure 4.13 showing use of test results
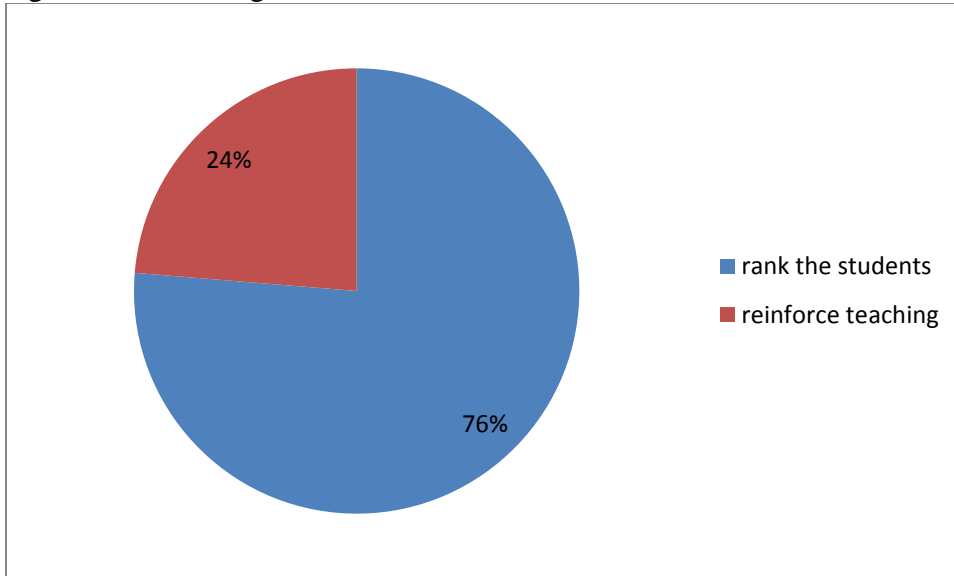
4.4.1.3 Teacher training in test development

Table 4.4.4. Teacher training on test development

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Never attended	34	89.5	89.5	89.5
	Not sure	2	5.2	5.2	94.7
	attended	2	5.2	5.2	100.0
	Total	38	100.0	100.0	

Table 4.4.4 show that 89.5% of the respondents had never attended an in-service training on test development, 5.2% had no idea what training on test development entails while 5.2% had attended the in-service training.

53

Figure 4.14 showing teachers training on test development.
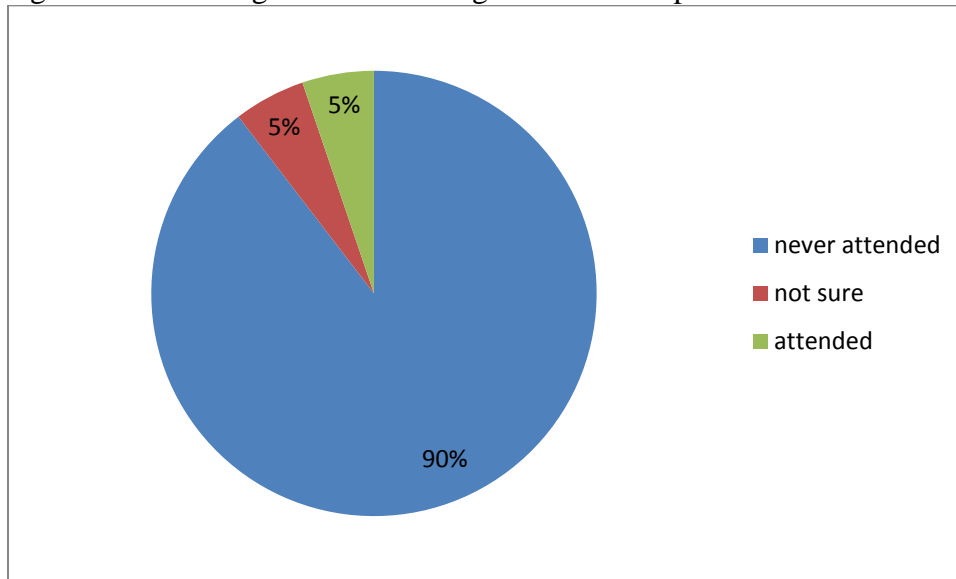
4.5. Relationship between degree of difficulty and discrimination index of test items

Table 4.5.1 mean difficulty and discrimination indices.

	Mean	Standard deviation
Difficulty index	0.4941	0.177
Discrimination index	0.3938	0.166

Table 4.5.1 show that the mean difficulty index for 50 test items in the commercial test paper is 0.4941 while the mean discrimination index is 0.3938.

Table 4.5.2 show that the correlation between difficulty index and discrimination index .The correlation coefficient r=0.443, p=0.001. The correlation is significant at the 0.01 level (2-tailed).This shows that there is a relationship between the difficulty index and the corresponding discrimination indices

Table 4.5.2: correlation between difficulty and discrimination indices.

		difficulty index	discrimination index
difficulty index	Pearson Correlation	1	.443[**]
	Sig. (2-tailed)		.001
	N	50	50
discrimination index	Pearson Correlation	.443[**]	1
	Sig. (2-tailed)	.001	
	N	50	50

**. Correlation is significant at the 0.01 level (2-tailed).

Figure 4.5 shows a scatter plot of difficulty indices and the corresponding discrimination indices.

55
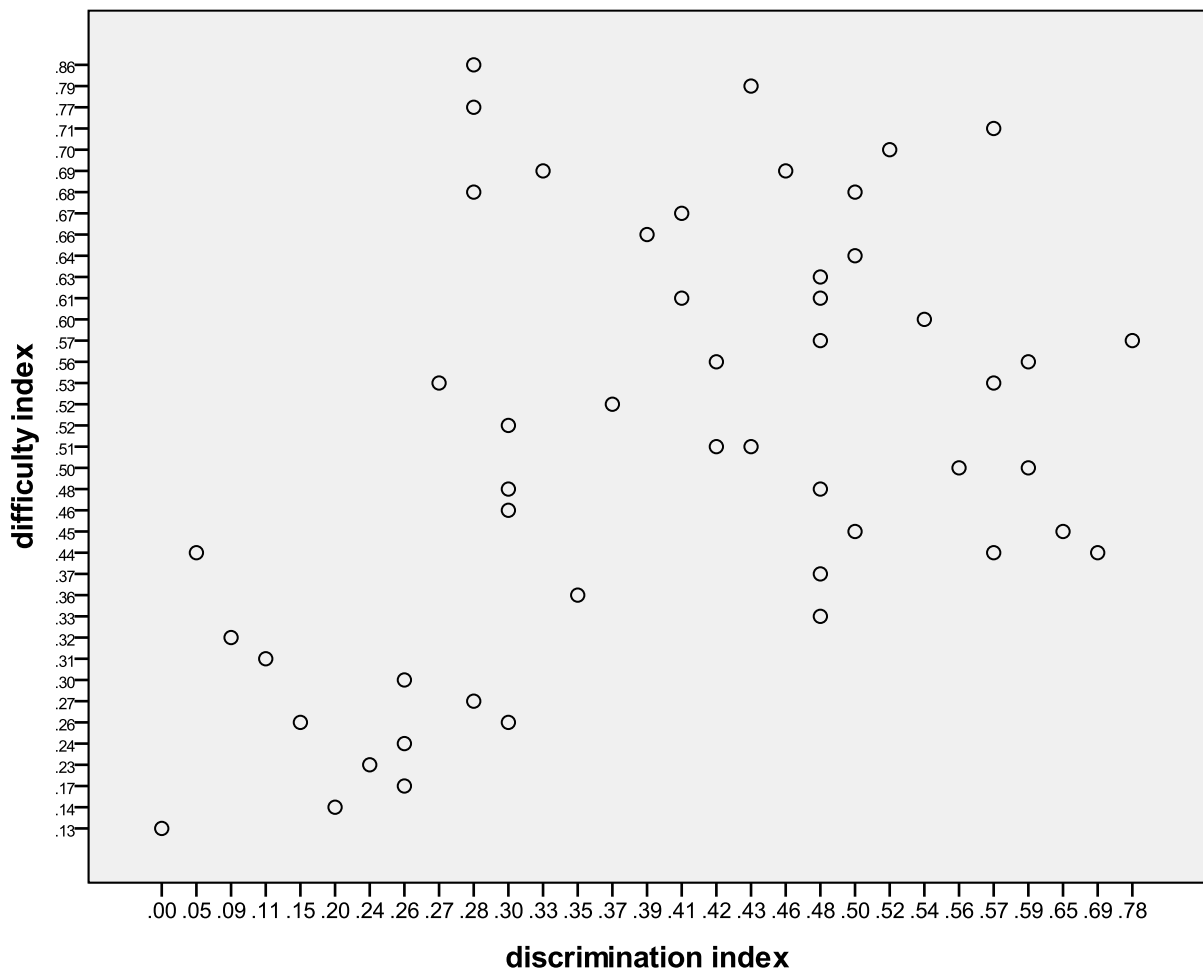
4.6 Factors influencing student performance in summative evaluation.

Table 4.6.1 Schools performance in the science commercial test paper.

	SCHOOL A	SCHOOL B	SCHOOL C
Number of students	80	60	60
Mean score	50.62	41.16	55.3

Table 4.6.1 show the students performance in the three selected schools in the science commercial test papers, with school A having a mean of 50.62,school B 41.16 and school C 55.30.

Table 4.6.2 KCPE mean scores for the three schools for the last three years.

Year	School A	School B	School C
2012	40.64	40.99	58.81
2013	33.66	38.28	56.68
2014	32.09	37.44	57.90
mean	35.46	38.90	57.80

Table 4.6.2 shows the KCPE mean scores for the selected three schools for the last three years. The average mean score for school A is 35.46, School B 38.9 and School C 57.80.

Table 4.6.3correration between commercial paper mean score and KCPE mean score

		Commercial test paper score	KCPE science mean score
Commercial test paper score	Pearson Correlation	1	.6525**
	Sig. (2-tailed)		.001
	N	3	3
KCPE science mean score	Pearson Correlation	.6525**	1
	Sig. (2-tailed)	.001	
	N	3	3

**. Correlation is significant at the 0.01 level (2-tailed).

Table 4.6.3 show the correlation coefficient between mean scores for the three selected school in the commercial test paper and their corresponding mean scores in the KCPE performance for the last three years is 0.6525.This indicates a strong positive relationship between student performance in commercial test papers and the K.C.P.E performance.

4.6.1 Student characteristics

4.6.1.1 Student gender.

Table 4.6.4 Student gender(Upper group).

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	male	29	53.7	53.7	53.7
	female	25	46.3	46.3	100.0
	Total	54	100.0	100.0	

 Table 4.6.4 show that 27% of the 200 students who scored highly in the commercial test paper comprised of 53.7% male students and 46.3% female students.

Figure 4.16 showing student gender(upper group)
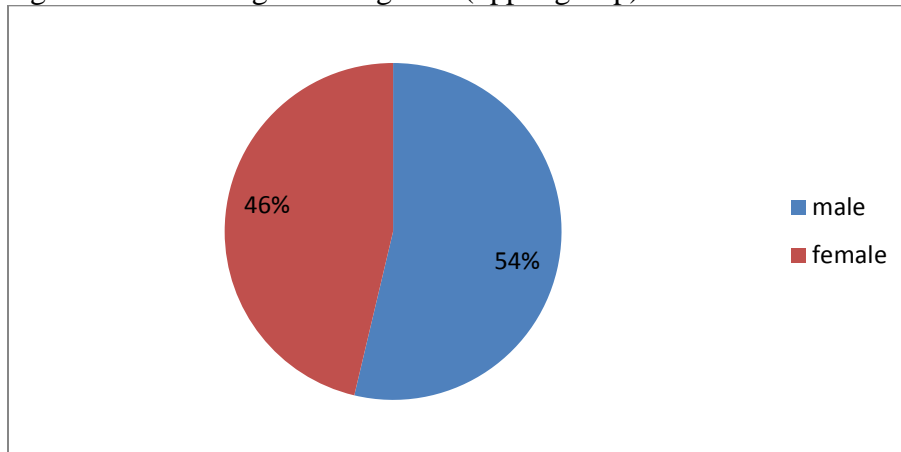
Table 4.6.5 Student gender (Lower group)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	male	23	42.6	42.6	42.6
	female	31	57.4	57.4	100.0
	Total	54	100.0	100.0	

Table 4.6.5 show that 27% of the 200 students who scored poorly in the commercial test paper comprised of 42.6% male students and 57.4 % female students.

Figure 4.17 showing student gender(lower group)
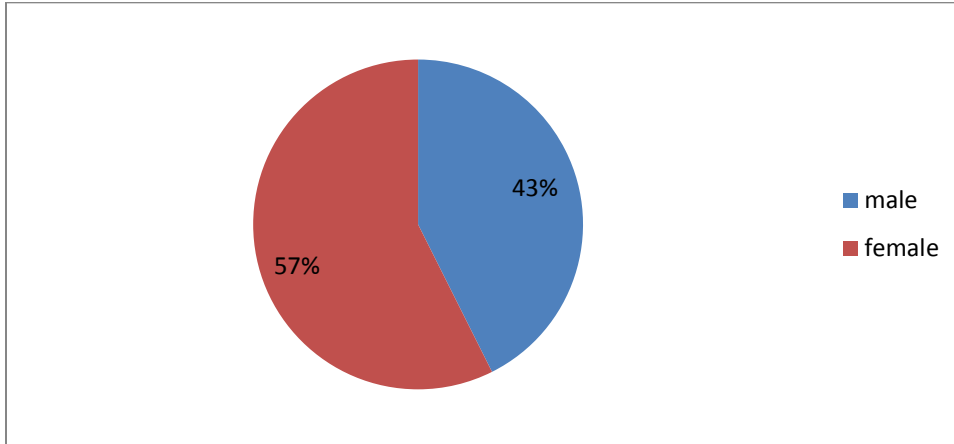
4.6.1.2 Testwiseness

Table 4.6.6 Use of testwiseness to answer test questions

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	yes	31	81.6	81.6	81.6
	no	7	18.4	18.4	100.0
	Total	38	100.0	100.0	

Table 4.6.6 show that 81.6% of the respondents train their students to use testwiseness in answering examination questions while 18.4% do not encourage their students to testwiseness.

Figure 4.18 showing use of testwiseness
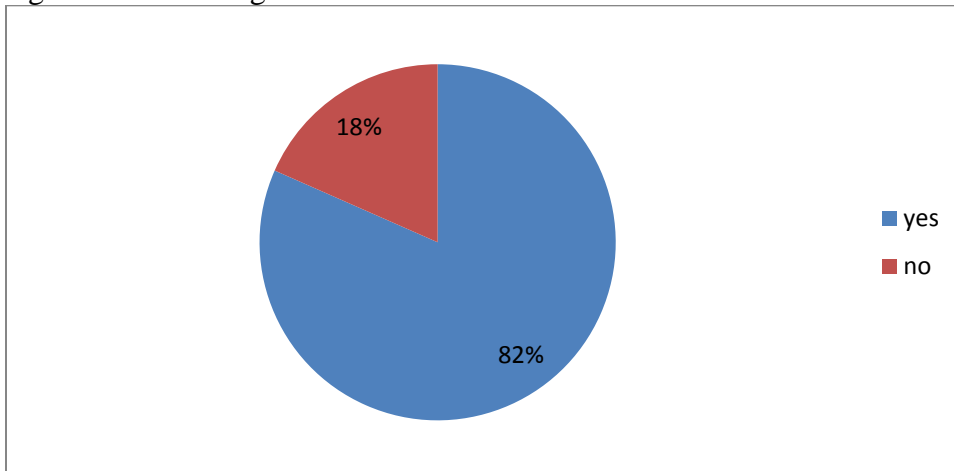
4.6.2 Teacher characteristics.

4.6.2.1 Teacher's gender

Table 4.6.7 Teacher's gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	male	26	68.4	68.4	68.4
	female	12	31.6	31.6	100.0
	Total	38	100.0	100.0	

Table 4.6.7 show that 68.4% of the respondents teaching science subject in class eight in Limuru sub-county are male while 31.6% are female.

Figure 4.19 showing teachers gender distribution
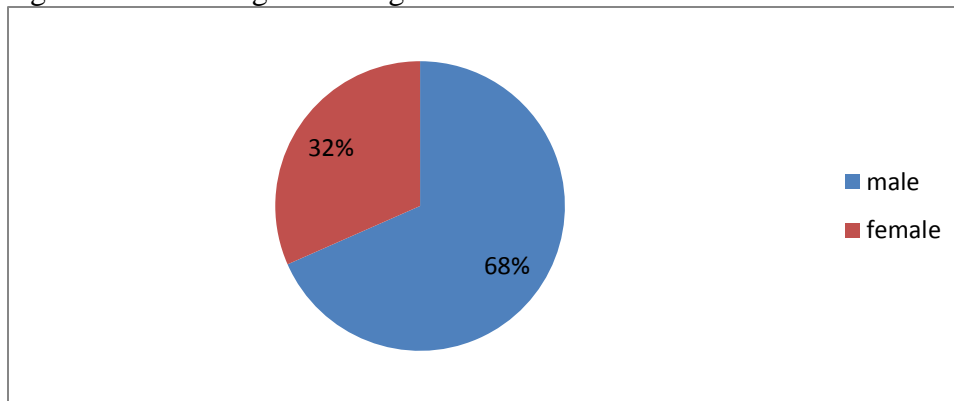
4.6.2.2 Teacher's professional Qualification.

Table 4.6.8 Teacher's professional Qualification.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	P1	14	36.8	36.8	36.8
	diploma	7	18.4	18.4	55.3
	B.Ed.	17	44.7	44.7	100.0
	Total	38	100.0	100.0	

Table 4.6.8 show that the respondents teaching science subject in class eight in Limuru sub-county, 36.8% are certificate holders, 18.4% are Diploma holders while 44.7% are B.Ed holders.

Figure 4.20 showing teachers professional qualification
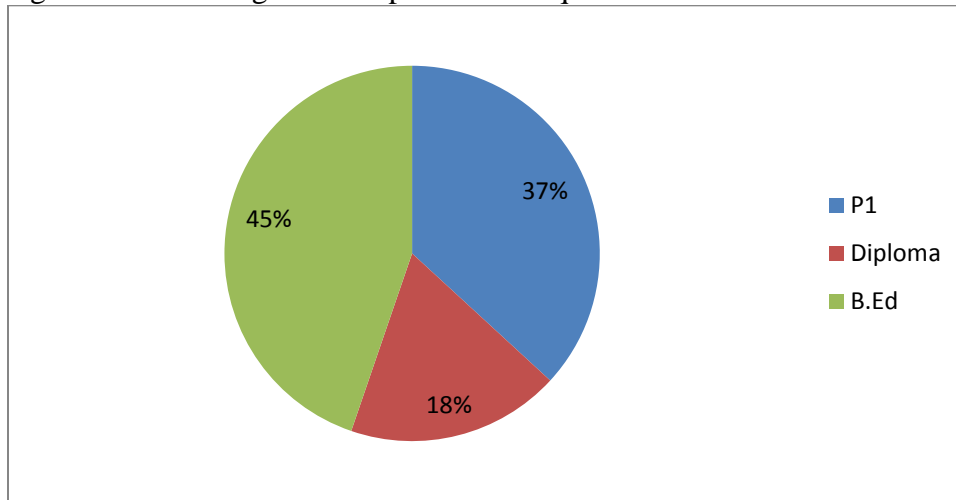
4.6.3 Home based factors

4.6.3.1 Parental involvement

Table 4.6.9 Parental assistance(Upper group)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes	42	77.8	77.8	77.8
	No	12	22.2	22.2	100.0
	Total	54	100.0	100.0	

Table 4.6.9 show that for the students in the upper group (good performance) 77.8% were assisted by their parents in school work while 22.2 were not assisted.

Figure 4.21 showing parental assistance (upper group)
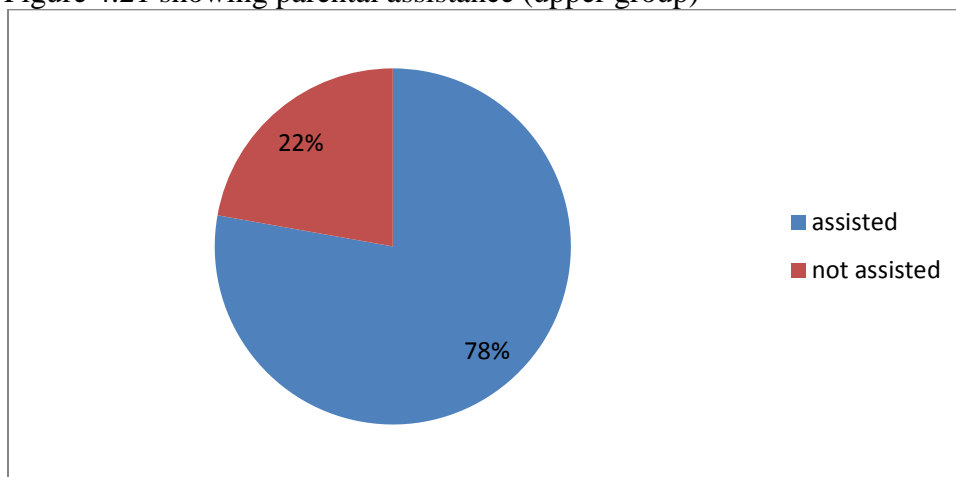
Table 4.6.10 Parental assistance (Lower group).

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	yes	34	63.0	63.0	63.0
	2	20	37.0	37.0	100.0
	Total	54	100.0	100.0	

Table 4.6.10 Show that for the students in the lower group (poor performance), 63.0% were assisted by their parents in school work while 37.0% were not assisted.

Figure 4.22 showing parental assistance (lower group)
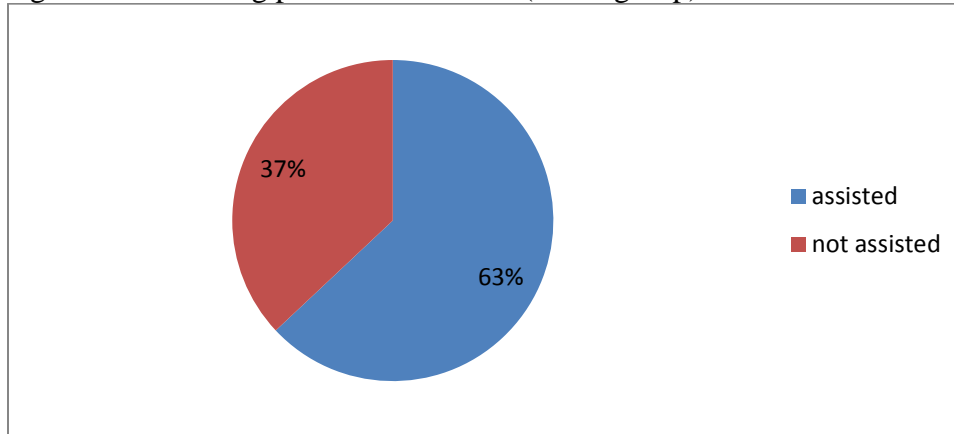
4.6.3.2 Parent's highest educational level reached.

Table 4.6.11 parents educational level(upper group)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	primary	6	11.1	11.1	11.1
	secondary	21	38.9	38.9	50.0
	tertiary	12	22.2	22.2	72.2
	don't know	15	27.8	27.8	100.0
	Total	54	100.0	100.0	

Table 4.6.11 show that for the students who performed well in the commercial test paper (upper group), 11.1% of their parents had primary education while 22.2% had tertiary education.

Figure 4.23 figure showing parents educational level(upper group)
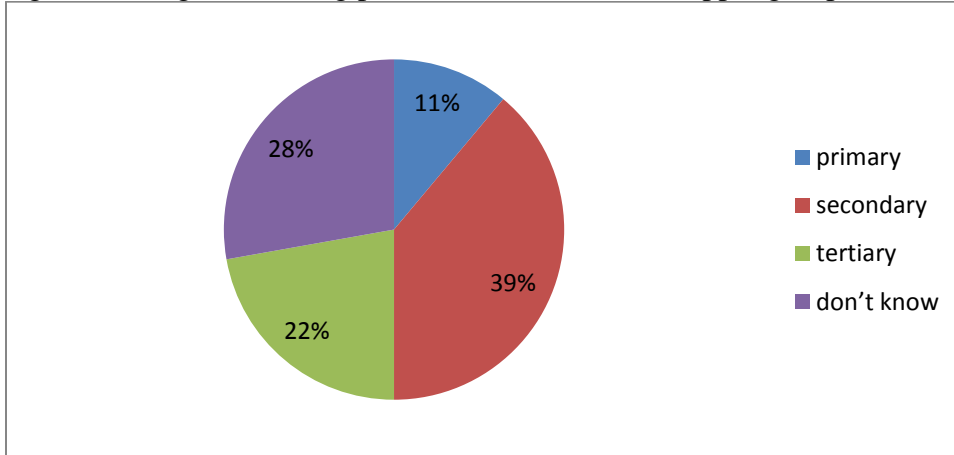
Table 4.6.12 Parents education level (Lower group).

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	primary	14	25.9	25.9	25.9
	secondary	24	44.4	44.4	70.4
	tertiary	5	9.3	9.3	79.6
	don't know	11	20.4	20.4	100.0
	Total	54	100.0	100.0	

Table 4.6.12 Show that for the students who performed poorly in the commercial test paper (Lower group), 25.9% had only primary education while 9.3% had tertiary education

Figure 4.24 showing parents educational level(lower group)
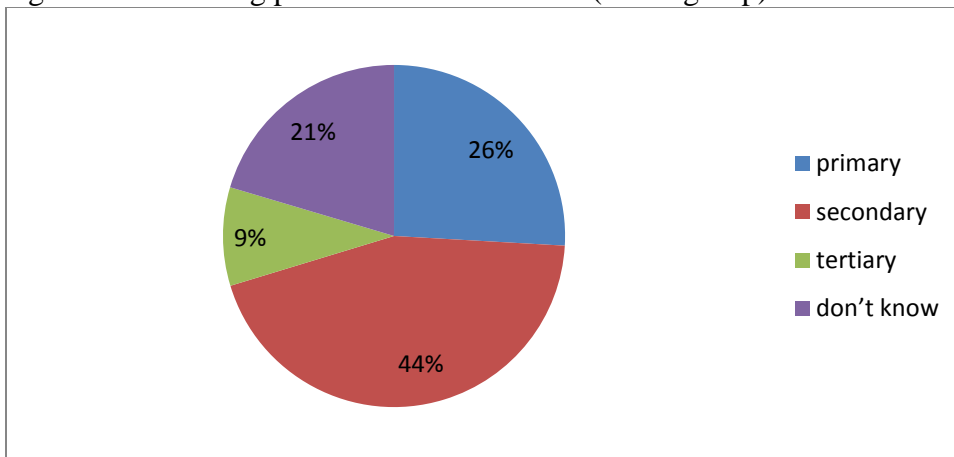
4.6.3.3 Family type

Table 4.6.13Family type (Upper group).

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	both parents	40	74.1	74.1	74.1
	one parent	12	22.2	22.2	96.3
	grand parents	2	3.7	3.7	100.0
	Total	54	100.0	100.0	

 Table 4.5.13 show that for the students in highest performing group (Upper group), 74.1% came from nuclear families (had both parents), while 25.9% came from single parent families.

Figure 4.25 showing family type distribution(upper group)
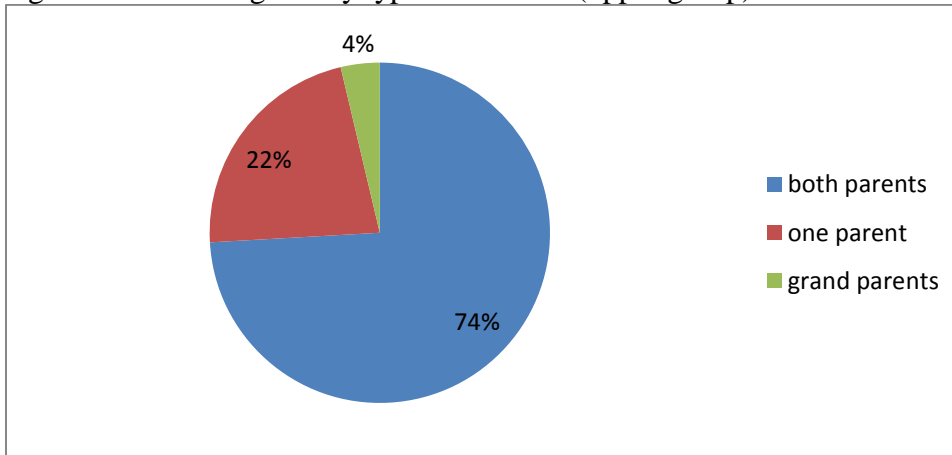
Table 4.6.14 family type (Lower group).

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	both parents	35	64.8	64.8	64.8
	one parent	19	35.2	35.2	100.0
	Total	54	100.0	100.0	

 Table 4.6.14 Show that for the students in the low performing group (Lower group), 64.8% came from nuclear families (had both parents) while 35.2 % came from single parent families.

Figure 4.26 showing family type distribution(lower group)
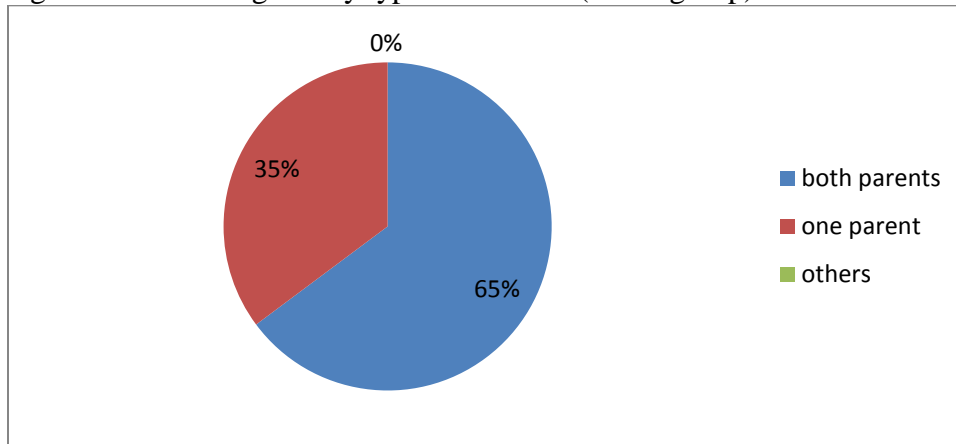
4.6.3.4 Social Economic status.

Table 4.6.15 payment of school levies.(Upper group)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Sent	41	75.9	75.9	75.9
	Not sent	13	24.1	24.1	100.0
	Total	54	100.0	100.0	

 Table 4.6.15 show that for the students in the Upper group (Good performers) 75.9% were sent home for non-payment of school levies, while 24.1% were not sent home.

Figure 4.27 showing parents ability to pay school levies (upper group)
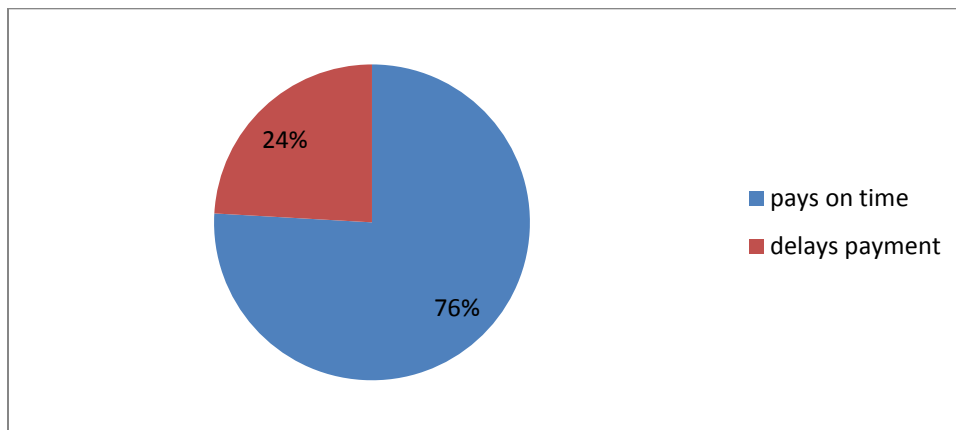
Table 4.6.16 payment of school levies.(Lower group)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	sent	46	85.2	85.2	85.2
	Not sent	8	14.8	14.8	100.0
	Total	54	100.0	100.0	

Table 4.6.16 Show that for the students in the Lower group (Poor performers) 85.2% were sent home for non-payment of school levies, while 14.8% were not sent home.

Figure 4.28 showing parents ability to pay school levies(lower group)
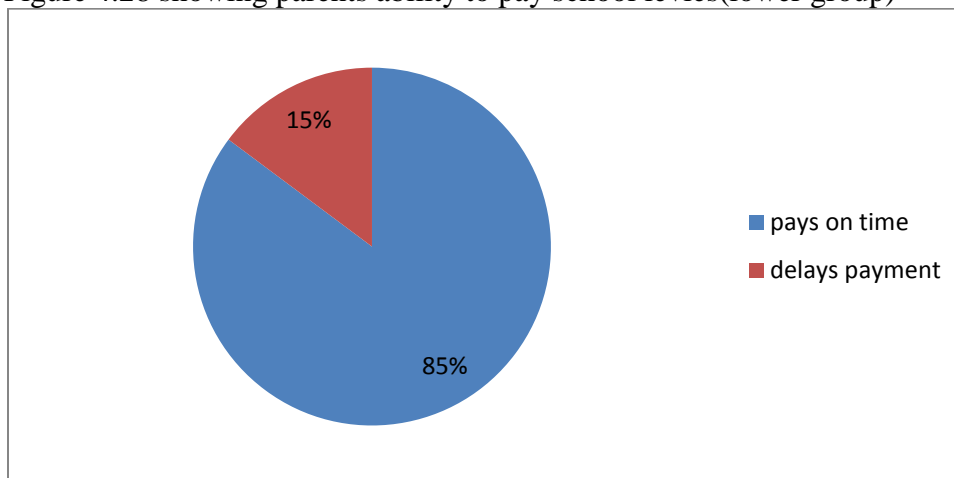
CHAPTER FIVE

DISCUSSION, SUMMARY AND CONCLUSION

5.1 Introduction

The main objective of the study was to determine the quality of commercial science test papers used in primary schools in Limuru Sub-county by looking at their Reliability, Difficulty and Discrimination indices. A descriptive research design was used to evaluate the quality of these commercial test papers and data was for the study was obtained through the use of teacher's questionnaires, students questionnaires and sample commercial test paper administered to the selected students and analysed. The study sample comprised of 200 class eight students sampled from 3 public primary schools in Limuru sub-county and 38 teachers teaching science in class eight in public primary schools. The data collected was then analysed using SPSS for quantitative data and descriptive statistics for qualitative data.

5.2 Generalisation

5.2.1. Internal Validity

The students test scores were used to calculate the reliability of the test papers whereby r= 0.870.Data obtained from the teachers questionnaires was then analysed in frequency tables and graphs expounding on the factors influencing reliability whereby 89.5% of the respondents had not attended in service training on test development. This indicates that the teachers teaching science in class eight in public primary schools in Limuru sub-county were not able to verify commercial test papers for quality. The proportion of these teachers who had high educational qualifications was 63.1%, while 73.7% had a workload of between 30 and 39 lessons per week therefore a high workload.

The difficulty and discrimination indices were calculated using the data from students' results analysis (Appendix 1) with the average difficulty index at P= 0.3938 and discrimination at P=0.4941. A correlation coefficient was then calculated between the two variables at r=0.443 showing a weak relationship.

On the factors influencing performance the study relied on data obtained from the students questionnaire which indicated that for the 27% of the 200 students who scored highly in the commercial science test paper, 53.35% were male students while 46.3% were female. On the other hand for the 27% of the 200 students who performed poorly 42.6% were male while 57.4% were female students. This data was therefore able to answer the research questions indicating that the students' gender was able to influence their performance.

However the question on the parent's educational level influence on student's performance was not well answered with 27.8% of the respondents indicating that they had no idea on the response. Future researchers on the area should include the parent's questionnaires or interview guide in their study.

5.2.2 External validity

According to Frisbie (1988) experts in educational measurement agreed informally that the reliability coefficient should be at least(r= 0.85) if the scores are to be used to make decisions about individuals and at least r= 0.65 if the decision is about a group of individuals. The current study has a reliability coefficient of 0.8708 which is within the set standard. A correlation

between the commercial science test paper score and the KCPE performance in science gave r=0.6525 indicating a strong positive relationship. With the KCPE having a national outlook the results of the current study can therefore be generalized widely despite it being delimited to Limuru sub-county.

According to Ebel and Frisbie (1986), Brown (1983) the acceptable range of the difficulty is between 0.2 and 0.8 while the satisfactory discrimination indices should be above 0.2(Aggarwal, 1986).In the current study 92% of the test items had acceptable difficulty index (between 0.2-0.8) while 90% of the items had acceptable discrimination indices (above 0.2) therefore the results are acceptable. However the current study was delimited to public primary schools in Limuru Sub-county and the factors influencing reliability, difficulty indices, discrimination indices and performance may vary in other areas.

According to the Kenya National Examinations Council (KNEC) examination report (2013), male students performed better (mean raw score of 33.67(2012) and 32.14(2013)) than the female students (mean raw score30.30 (2012) and 29.63(2013).This is also reflected by the current study whereby 53.7% of male students performed better than 46.3% female students in the 27% of the 200 students who scored highly in the commercial science test paper (tables 4.6.3 and 4.6.4) making the results to be acceptable within the scope of the study.

The research hypothesis: The commercial science test papers used for evaluation in public primary schools in Limuru Sub-County are of good Quality, was proven to be correct therefore accepted.

5.3 Discussion

5.3.1 Objective 1. Reliability of commercial science test papers used for assessment in public primary schools in Limuru Sub –county.

According to Scattler (2001) the higher the number of test items the higher the reliability coefficient. This is supported by this study whereby 50 test items yielded a reliability coefficient of 0.8708 which was high. On test duration when the test was given within 1hour 40 minutes 92.1% of the respondents were able to answer all the test questions positively influencing the reliability.

In the current study the 73.7% of the respondents had an experience of above 10 years while 63.1% had educational qualification of above diploma level positively influencing the reliability. This corroborates with studies by Kiragu &Okunya (2012) who posited that as the level of education raises the reliability also gets better. Therefore the commercial science test papers used for evaluation in public primary schools in Limuru Sub-county are highly reliable.

5.3.2 Objective 2: Difficulty level and discrimination indices of individual test items in commercial science test papers used in public primary schools in Limuru Sub-County.

Any discrimination index of 0.2 and above is acceptable and the test items would be able to differentiate between weak and good students (Bharti et al, 2013). In the present study 90% of the test items had a discrimination index of above 2 indicating that they were acceptable. Five out of the fifty test items had a discrimination index below 0.2 and therefore they ought to be modified or rejected outright. Despite this the test was able to differentiate between the poor and the good performers.

Research by Sim Rasiah (2006) found out that maximum discrimination occurred with difficulty index between 0.4 and0.74.In the present study 78% of the test items with a difficulty index above 0.28 had excellent discrimination index, while the correlation coefficient of 0.443 indicated a relationship between the difficulty index and the corresponding discrimination indices. The test items found in commercial science test papers used for evaluation in public primary schools in Limuru Sub-county were found to have difficulty and discrimination indices within acceptable range.

5.3.3 Objective 3: Correlation between student performance in science commercial test papers and their K.C.P.E performance in science.

The current study found out that the mean of the KCPE results in science improved and was higher than that of the science commercial test papers. Further results on the Pearson correlation coefficient between commercial test papers results and the KCPE science results at 0.01level (2 tailed) was highly significant at 0.652 indicating a strong positive correlation. The study found out that 53.7% of the male students performed well as compared to 46.3% female students therefore supporting findings by smith (1992) who posited that girls perform poorly in science as compared to boys due to self conceptions leading to a pattern of helplessness.

The study further found out that teachers employ testwiseness principles on the learners unconsciously leading to better performance in line with a study by Mehrens et al (1973).Finally the current study supported findings by Fadeiye (2008) who posited that differences in academic performance of children exists in terms of the family they come from with those from two parent

families performing better. The commercial science test papers used for evaluation in public primary schools are therefore highly predictive of the KCPE science performance.

5.4. Contributions to the general theoretical framework.

The study employs the Classical test theory (CTT) which assumes that the raw score(X) obtained by an individual is made up of the true component (T) and the random error (E) component. The current study expounds on the principle of standard error of measurement, whereby when this error is greatly reduced the reliability of the test is increased highly.

The unsystematic errors (e.g. student's wellbeing and structural surrounding) were beyond the scope of this study while the systematic errors (related to test analysis) were the main focus of this study. When the systematic errors were reduced as much as possible then the X=T. In the current study when the test papers were analysed for reliability r=0.8708. This indicates a high reliability resulting in a great reduction of the systematic error making X = T.

This reliability was influenced positively in the study by the test duration being maintained at 1hour, 40 minutes by 92.1% of the respondents. The teacher characteristics notably qualification was pegged at above diploma level (63.1%) and experience pegged above 10 years by 73.7% of the respondents. However this reliability could be compromised by the majority of the teachers (94.7%) having no training in test development thereby increasing the measurement error.

The difficulty and discrimination indices analysis showed that 90% of the test items had a discrimination index of above 0.2 indicating that the test items were of good quality, while 92%

of the of the test items had a difficulty index between 0.2 and 0.8 which is highly acceptable. This item level analysis results indicated positive results greatly reducing the systematic error making the student raw score obtained to be almost equal to the true score of the test.

5.5. Recommendations

The significance of the study was to provide useful information on the quality of commercial science test papers to the education sector stakeholders more so the ministry of education and the teachers who use them. In view of this the study found out that the commercial science test papers were of good quality but were lacking in some aspects. Therefore the researcher recommends the following:

i) From the research findings it is evident that 89.5% of the teachers teaching science in class eight in public primary schools lack training in test development. The ministry of education should include test development as a training item during the in service training of teachers and this training should be done often.

ii) The research findings indicate that science commercial test papers are reliable and correlate highly to the final KCPE performance in science. The ministry of education should recognize and set guidelines of the content to be included and the preparation procedure for the commercial test publishers.

iii) Commercial test publishers should liaise with the teachers trained on test development in their test setting order to come up with quality tests.

5.6 Further research

This research was delimited to commercial science test papers used in public primary schools in Limuru Sub-County therefore further research can be done in other parts of the country and also focus in other subjects. Due to resource limitations the study focused on a study population of 600 students in 3 public primary schools and 38 teachers teaching science in class eight. Further research therefore can be undertaken using a larger population and possibly a different research design.

A similar study can be done on reliability, item difficulty and discrimination indices of test items in commercial test papers and their correlation to KCSE performance in various subjects in secondary schools in Kenya. This is because the current study only focused on multiple choice types of questions which are not used in secondary schools in Kenya.

5.7 Conclusion

The general performance of students in science commercial test papers is higher than that of the KCPE in average and low performing schools but lower than that of the KCPE in high performing schools. This indicates that the tests are of good quality but teacher, student and the environmental characteristics influence the results obtained.

From the research findings the schools source the commercial test papers from the same test vendors however teachers' skills in test development are limited. This makes them not to be able to moderate the test items therefore improve on the test quality.

Finally the commercial test papers in science are of good quality and with minor adjustments they would make excellent testing tools.

REFERENCES

Aggarwal. Y.P.(1986). Statistical Methods, Concepts, Applications and Computations. New Delhi: Sterling Publication.

Benson, J. (1998) Developing a Strong Program of Construct Validation: A Test Anxiety Example, Educational Measurement: Issues and Practice, **17:** 10-17.

Crocker, L., and Algina, J. (1986) Introduction to Classical and Modern Test Theory, Harcourt Brace Jovanovich College Publishers: Philadelphia

Cronbach, L.J., & Meehl, P.E. (1955) Construct Validity in Psychological Tests, Psychological Bulletin, **52**: 281-302.

Devellis, R.F. (1991) Scale Development: Theory and Applications, Applied Social Research Methods Series 26, Sage: Newbury Park.

Ebel, R. L. & Frisbie, D. A. (1986). Essentials of education measurement. Englewood Cliffs, NJ: Prentice Hall

Gregory, R.J. (1992) Psychological Testing: History, Principles and Applications, Allyn and Bacon: Boston

Henson, R.K. (2001) Understanding Internal Consistency Reliability Estimates: A Conceptual Primer on Coefficient Alpha, Measurement and Evaluation in Counseling and Development, **34:**177-188.

Kenya national examinations council(Knec) 2013 KPCE examinations report

Kellaghan ,T & Greaney,V.(2001) Using assessment to improve the quality of education.Paris:UNESCO.

Kigotho, W; (2004)"Fake School Examinations, A special report". The East African Standard 1st July, 2004.

Kinyua, K & Okunya L.O. (2014) Validity and Reliability of teacher made tests: Case study of year II physics in Nyahururu district of Kenya. African educational research journal vol2 (2).

Kothari. C.R.(2010). Research Methodology: Methods and Techniques. New Delhi: New Age International Pvt.Ltd.

Mehrens, W.A., and Lehman, I.J. (1991) Measurement and Evaluation in Education and Psychology, (4$_{th}$ edn) Holt, Rinehart and Winston Inc: Orlando, Fl.

Messick, S. (1995) Validity of Psychological Assessment: Validation of Inferences from Persons Responses and Performances as Scientific Inquiry into Score Meaning, American Psychologist, **50** (9):741-749.

Mitra NK, Nagaraja HS, Ponndurai G et al. The levels of difficulty and discrimination indices in

type A multiple choice questions of pre-clinical semester 1 multidisciplinary summative tests.

IeJSME 2009;

.
Ministry of Education in Kenya. Education in Kenya. Retrieved on 10/5/2013 from
www.science and technology .go.ke.

Mugenda, O.M &Mugenda, A.G (2003) Research methods: Qualitative and Quantitative
Approaches. Nairobi. African Centre for Technology Studies Press.

Mwangi G (2006).'Commercial Mock Examinations' in The Daily Nation 4 March
2006.Nairobi.

Mwangi G and Ouko R (2004). 'National Examinations and Internal Examinations' The
Daily Nation 27 July 2004.

 Ndalichako, J.L (2004). Towards an Understanding of Assessment Practices of Primary School
Teachers in Tanzania, Zimbabwe journal for Educational research. Vol 16, No (3)

Nunnally, J.C., and Bernstein, I.H. (1994) Psychometric Theory, (3$_{rd}$ edn), McGraw-Hill: New
     York

Ogunniyi, M. B. (2004) Educational Measurement and Evaluation. Lagos: Longman Nigeria
Limited.

Oyejide, A. P. (1991). Effects of Confidence Scoring Procedure on the Psychometric Properties
of three Multiple-choice Test formats.

Sigmund, W.G. (2003) Business Research Methods, (7th edn), Thompson South-Western: Ohio

Sim SM, Rasiah RI. Relationship between item difficulty and discrimination indices in

true/false-type multiple choice question of a Para- clinical multidisciplinary paper. Ann Accad

med Singapore 2006; 35: 67-71.

Stiggins, R.J., (1991). Relevant Classroom Assessment Training for Teachers. Educational
Measurement:Issues and Practice.

Stiggins, R.J. and Conklin, N. (1992). In Teachers' Hands, Investigating the Practice of the Classroom Assessment. Albany. NY. SUNY. Press

WAEC (1995). An Empirical Check on the Effect of the Number of options of Multiple Choice Items on the Difficulty and Reliability of Ordinary Level Question Papers. Completed Research Project Report II, 1,63-72.

Wiersma William & Jurs Stephen G. (2009). Research Methods in Education: An Introduction. New Delhi: Pearson

Wollack, J. A (2004) Helpful Tips for creating reliable and valid Tests. London: Madson.

Wuddah, A. A (l994). Difficulty and Reliability of Basic Education Certificate Examination (BECE) Question Papers as Functions of the Number of Option in their Multiple Choice Test;WAEC RPAI/95.
.

Appendix I

UNIVERSITY OF NAIROBI

FACULTY OF ARTS
DEPARTMENT OF PSYCHOLOGY

Telegrams: Varsity Nairobi
Telephone: 3318262 ext.28439
Telex: 22095

P.O. BOX 30197
NAIROBI
KENYA

September 02, 2015

TO WHOM IT MAY CONCERN

RE: BONIFACE CHEGE NDUNG'U – E58/63591/2013

The above named is a student in the Department of Psychology studying Masters Psychology in M.Ed of Arts in counseling Psychology programme at the University of Nairobi. He is doing a research on *"Reliability item difficulty and discrimination indices of science test Items in commercial test papers and their correlation to students KCPE Performance in science in Limuru Sub County"*. The requirement of this course is that the student conducts research and collects data in the field on the topic area.

In order to fulfill this requirement, I would like to re-affirm that the said student is a registered student and is intending to go and carry out field work. Any assistance accorded to him will be highly appreciated.

Should there be any queries do not hesitate to contact the Chair of the Department of Psychology, University of Nairobi.
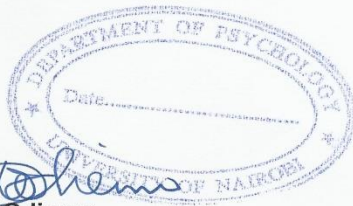
Dr. Luke Odiemo
Chairman,
Department of Psychology

Appendices II

Table 4.5.1 difficulty and discrimination indices

Number	Group					Total	P	D
1		A*	B	C	D			
	U	42	4	0	8	54	0.435	0.685
	L	5	16	19	14	54		
2		A*	B	C	D			
	U	42	8	1	3	54	0.454	0.65
	L	7	15	3	29	54		
3		A	B*	C	D			
	U	2	43	7	2	54	0.5	0.59
	L	10	11	23	10	54		
4		A*	B	C	D			
	U	29	5	15	5	54	0.36	0.35
	L	10	15	12	17	54		
5		A*	B	C	D			
	U	50	0	2	2	54	0.68	0.5
	L	23	12	9	10	54		
6		A	B	C*	D			
	U	10	2	31	11	54	0.33	0.48
	L	32	9	5	8	54		
7		A	B*	C	D			
	U	2	50	2	0	54	0.69	0.46
	L	11	25	8	10	54		
8		A	B	C	D*			
	U	2	0	0	52	54	0.70	0.52
	L	14	3	3	24	54		
9		A*	B	C	D			
	U	42	2	2	8	54	0.56	0.42
	L	19	10	8	17	54		
10		A	B	C	D*			
	U	8	6	2	38	54	0.45	0.5
	L	20	11	12	11	54		
11		A	B*	C	D			
	U	0	54	0	0	54	0.71	0.57
	L	12	23	8	14	54		
12		A*	B	C	D			
	U	39	2	8	5	54	0.51	0.42
	L	16	7	18	13	54		
13		A	B	C*	D			
	U	2	2	47	3	54	0.63	0.48
	L	8	11	21	14	54		
14		A	B	C	D*			

		A	B	C	D			
	U	3	2	10	39	54	0.44	0.57
	L	14	15	17	8	54		
15		A*	B	C	D			
	U	36	14	4	0	54	0.53	0.27
	L	21	15	11	7	54		
16		A*	B	C	D			
	U	19	33	2	0	54	0.23	0.24
	L	6	42	3	3	54		
17		A	B*	C	D			
	U	21	20	8	5	54	0.24	0.26
	L	35	6	10	3	54		
18		A	B	C*	D			
	U	0	5	44	5	54	0.68	0.28
	L	7	11	29	7	54		
19		A	B*	C	D			
	U	11	34	2	7	54	0.48	0.30
	L	17	18	7	11	54		
20		A	B	C	D*			
	U	0	5	0	49	54	0.77	0.28
	L	4	16	0	34	54		
21		A	B	C	D*			
	U	2	10	6	36	54	0.52	0.30
	L	12	11	11	20	54		
22		A	B	C*	D			
	U	10	5	39	0	54	0.51	0.43
	L	19	8	16	11	54		
23		A	B*	C	D			
	U	11	33	2	8	54	0.46	0.30
	L	8	17	16	13	54		
24		A*	B	C	D			
	U	46	2	0	6	54	0.61	0.48
	L	20	11	16	7	54		
25		A	B*	C	D			
	U	0	52	2	0	54	0.57	0.78
	L	10	10	10	24	54		
26		A	B	C*	D			
	U	2	2	44	6	54	0.57	0.48
	L	13	17	18	6	54		
27		A	B	C	D*			
	U	2	2	6	44	54	0.53	0.57
	L	8	8	25	13	54		
28		A*	B	C	D			
	U	42	4	4	4	54	0.50	0.56
	L	12	12	6	24	54		

						TOTAL	P	D
29		A	B	C	D*			
	U	0	0	0	54	54	0.79	0.43
	L	9	6	8	31	54		
30		A	B	C*	D			
	U	11	13	23	7	54	0.30	0.26
	L	24	9	9	12	54		
31		A	B	C	D*			
	U	16	0	25	13	54	0.14	0.20
	L	15	0	37	2	54		
32		A	B*	C	D			
	U	14	38	0	2	54	0.52	0.37
	L	16	18	12	8	54		
33		A	B*	C	D			
	U	2	47	5	0	54	0.60	0.54
	L	9	18	20	7	54		
34		A	B*	C	D			
	U	10	25	14	5	54	0.44	0.05
	L	7	22	18	7	54		
35		A	B	C	D*			
	U	5	7	3	39	54	0.48	0.48
	L	6	15	20	13	54		
36		A	B	C	D*			
	U	8	0	0	46	54	0.66	0.39
	L	15	10	4	25	54		
37		A	B	C*	D			
	U	2	4	7	41	54	0.13	0
	L	4	18	7	25	54		
38		A	B*	C	D			
	U	7	33	4	10	54	0.37	0.48
	L	22	7	10	15	54		
39		A	B*	C	D			
	U	21	20	5	8	54	0.32	0.09
	L	7	15	16	16	54		
40		A*	B	C	D			
	U	22	0	28	4	54	0.27	0.28
	L	7	10	27	10	54		
41		A	B	C	D*			
	U	11	4	23	16	54	0.17	0.26
	L	4	7	41	2	54		
42		A	B*	C	D			
	U	0	46	4	4	54	0.69	0.33
	L	5	28	10	11	54		
43		A	B	C*	D			
	U	2	2	46	4	54	0.56	0.59

		A	B	C	D			
	L	14	9	14	17	54		
44		A	B	C*	D			
	U	14	4	20	16	54	0.31	0.11
	L	12	14	14	14	54		
45		A	B	C	D*			
	U	2	2	2	48	54	0.64	0.5
	L	3	13	17	21	54		
46		A	B	C	D*			
	U	0	0	0	54	54	0.86	0.28
	L	2	6	7	39	54		
47		A	B	C*	D			
	U	26	5	18	5	54	0.26	0.15
	L	10	23	10	11	54		
48		A	B	C*	D			
	U	10	18	22	4	54	0.26	0.30
	L	14	14	6	20	54		
49		A	B*	C	D			
	U	0	44	5	5	54	0.61	0.41
	L	8	22	17	7	54		
50		A	B*	C	D			
	U	3	47	2	2	54	0.67	0.41
	L	8	25	18	5	54		
					AVERAGE		0.4941	0.3938

*correct response.

TEACHERS QUESTIONNAIRE

This questionnaire is designed to gather information on Reliability, Item difficulty and Discrimination indices of science test items in commercial test papers and their correlation to student KCPE performance in science in Limuru Sub-County. You are kindly requested to tick (√) the appropriate response or respond as indicated. Do not put your name or any other form of identification. The information you give will be confidential and will only be used for the purpose of this study.

Please respond to all items.

SECTION A: DEMOGRAPHIC DATA

1) What is your Gender?

i) Male   (   )

ii) Female   (   )

2) Which is your age bracket?

i) 20-30years   (   )

ii) 31-40 years   (   )

iii) Above 40 years   (   )

3) What is your marital status?

i) Single   (   )

ii) Married   (   )

iii) Divorced   (   )

4) What is your highest professional Qualification?

i)   Certificate   (   )

ii)   Diploma   (   )

iii)   B.Ed   (   )

Any other (specify)_____

5) For how long have you been in the teaching profession?

    i)   1 year and below.   (   )

    ii)   2 to 10 years   (   )

    iii)   Above 10 years   (   )

6) Which other subject do you teach other than science?

    i)   Mathematics   (   )

    ii)   Languages   (   )

    iii)   Humanities   (   )

7) How are you housed?

    i)   School house   (   )

    ii)   Own house   (   )

    iii)   Renting   (   )

8) How would you classify your social economic status?

    i)   Below 30,000 ksh   (   )

    ii)   30,000-50,000 ksh   (   )

    iii)   Above 50,000 ksh   (   )

SECTION B: FACTORS INFLUENCING RELIABILITY OF COMMERCIAL

    TEST PAPERS.

9) I have attended in-service training on test development.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

10) The number of lessons that I teach per week are:-

i)   20-29   (   )

ii)   30-39   (   )

iii)   40-49   (   )

Others (Specify) _____

11) My workload is manageable.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

12) When the test is allocated 1HOUR the students answer all the questions.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

13) When the test is allocated 1 HOUR 40 MINUTES the students answer all the questions.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

14) When the number of Questions is 30 the students perform better in the exam.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

15) When the number of Questions is 40 the students perform better in the exam.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

16) When the number of Questions is 50 the students perform better in the exam.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

17) Internal examinations test papers are set by individual teachers.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

18) Internal examinations test papers are set by a subject panel.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

19) Internal examinations test papers are purchased from commercial test vendors.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

SECTION C: FACTORS INFLUENCING DIFFICULTY AND DISCRIMINATION INDICES

20) Test results are used to rank the students.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

21) Test results are used to reinforce teaching.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

22) 2-multiple choice options are used in the internal exams.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

23) 3- Multiple choice options are used in the internal exams.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

24) The students' results are analysed by use of the mean.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

25) The students' results are analysed by use of the standard deviation.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

26) The academic level of students influences the type of exams that is given.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

SECTION D: FACTORS INFLUENCING STUDENT PERFORMANCE IN

SUMMATIVE EVALUATION.

27) Please provide the schools KCPE mean score in science for the last three years.

YEAR	KCPE SCIENCE MEAN SCORE	MALE MEAN SCORE	FEMALE MEAN SCORE
2012			
2013			
2014			

28).I have heard about the word **test wiseness**.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

29) In a multiple choice exam I advice students to Always guess if right answers only are

scored)

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

30) In a multiple choice exam I advice students to Choose neither or both of two option

which imply the correctness of each other.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

31) In a multiple choice exam I advice students to Eliminate options which are known to be

incorrect and choose from among the remaining options.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

32)   In a multiple choice exam I advice students to Read all the alternatives first before

choosing an answer.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

33) Most of the students in the school come from single parent families.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

34) Most of the students in the school come from two parent families

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

35) My students' parents have a low level of education.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

36) My students' parents have a moderate level of education.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

37). My students' parents have a very high level of education.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

38). Parents are involved in their children's learning in school.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

39) .The students in my school are often sent home due to nonpayment of school levies.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

40).There are students' dropout cases in my school.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

Appendix 111

STUDENTS QUESTION PAPER

This question paper is designed to gather information on Reliability, Item difficulty and Discrimination indices of science test items in commercial test papers and their correlation to student KCPE performance in science in Limuru Sub-County. You are kindly requested to tick (√) the appropriate response or respond as indicated. **Do not put your name or any other form of identification. The information obtained is strictly for research purposes and will be treated with confidentiality.**

SECTION 1

1) What is your gender?

   i)   Male   (   )

   ii)   Female   (   )

2) How old are you?

   i)   Below 14 years   (   )

   ii)   14 years   (   )

   iii)   Above 14 years   (   )

3) Do you live with both your parents?

   i)   Yes   (   )

   ii)   No   (   )

           If no specify _____

4) What level of education did your parents complete?

   i)   No education   (   )

   ii)   Primary   (   )

iii)   Secondary   (   )

iv)   Higher than secondary   (   )

v)   Don't know   (   )

5) How many siblings do you have?

i)   None   (   )

ii)   1-2   (   )

iii)   3-4   (   )

iv)   Above 5   (   )

6) Do your parents help in your school work?

i)   Yes   (   )

ii)   No   (   )

       If yes explain_____

7) Have you ever been sent home from school due to non-payment of school levies?

i)   Yes   (   )

ii)   No   (   )

SECTION II

Please answer all the questions by drawing a dark line across the correct response.

1. Which one of the following **CORRECTLY** shows the blood flow from the lungs to the body?

    A. pulmonary artery→ left auricle   right ventricle   aorta

    B. pulmonary artery   left auricle   left ventricle   aorta

    C. pulmonary vein   right auricle   right ventricle   aorta

    D. pulmonary vein   left auricle   left ventricle   aorta

92

2. Which one of the following is **NOT** a proper use of medicine?

    A. Always complete the dosage

    B. Do not expect to be given medicine every time you visit the hospital

    C.Instructions and directions for use should be kept with each type of medicine

    D.Do not buy medicine from a shop of your own.

3. Which one of the following features does not help to minimize resistance in water as fish

moves in water?

    A. Scales pointing backwards

    B. Swim bladder

    C. Mucus layer

    D. Fins facing behind.

4. The reason why some plants have silvery hair on their leaves is to_____

    A. Reflect light and heat

    B.Trap more sunlight

    C.Reduce water loss.

    D. Allow stomata to open

 5. Which one of the following consists of groups that contain animal fats only?

    A.Butter, Bacon, Cheese

    B.Cheese, Salads, Cookingfat

    C. Bacon, Salads, Cooking fat and Salads.

    D. Margarine, Cooking fat and Salads.

6. Which one of the following is the first step in obtaining salt from a mixture of salt solution,

aluminium wire and copper wire?

A.Using a magnet

B.Evaporation

C.Filtering

D.Sieving.

7. Std 5 pupils were asked by their science teacher to bring the following materials.

i) bottle tops   ii) water   iii) kerosene iv) methylated spirit v) petrol

Which one of the following were the pupils likely to investigate?

A) Characteristics of liquids

B) Rate of evaporation in different liquids

C) Convection in liquids

D) Expansion and contraction of liquids.

8)The main reason why pregnant mother should feed on food rich in calcium is in order to _____

A) Improve the immunity of the baby

B) Enable the mother to have strong bones and teeth

C) Enable the mother to have strength during birth.

D) Strengthen the formation of bones and teeth in unborn babies.

9. Which one of the following is the most effective way of controlling the spread of HIV/AIDS among the infected married couples?

A) Use of condoms

B) Abstinence

C) Medical attention

D) Being faithful

10. Which one of the following water will be **BEST** for drinking?

A) Soft clean water

B) Hard clean water

C) Soft safe water

D) Hard safe water

12) The following are signs and symptoms of a nutritional deficiency.

i) muscles disappear   ii) visible bones   iii) the child cries often

The best way to prevent the diseases is by _____

A) Eating enough balanced diet

B) Eating enough protein

C) Eating enough food rich in calcium

D) Eating enough food rich in iron

13. Which one of the following consists only of animals that have varying body temperature?

A) Shark, Tortoise, Whale

B) Ostrich, Shark, Chameleon

C) Salamanders, Trout, Turtle

D) Bat, Ostrich, whale

14. The following experiment was done by STD 5 Pupils in the comparison of capillarity in different soils.

In carrying out the experiment above which one of the following need not to be the same?

A) Amount of each type of soil

B) Amount of cotton wool.

C) Size of tubes

D) Amount of water

15. Which one of the following is true about the experiments above?

    A).Both candles went off at the same time

    B).Water rose by a fifth in both basins

    C). Both the candles went off immediately

    D). The candles continued burning until they got finished

16. In which one of the following positions from the sun is planet earth?

    A). Second

    B). Third

    C). Fourth

    D). Fifth

17. Which one of the following crops are **CORRECTLY** represented by W ,X , Y, Z.

W	X	Y	Z.
A). Spinach	groundnuts	sunflower	sisal
B). Cauliflower	arrowroots	sunflower	palm tree
C). Peas,	sweet potatoes,	coconut	sisal
D). kales,	sunflower,	greengrams,	palmtree

18. The following is a feeding relationship that was observed in a certain habitat.

Algae   Small fish   Nile perch   K

Which one of the following is **CORRECTLY** represented by K?

    A) Snakes

    B) Humming bird

    C) King fish

D) Salamander

19. Std 5 pupils went for a nature walk with their science teacher and observed an animal with the following characteristics.

i) Independent body temperature

ii) Lay eggs

iii) Body covered with fur

Which one of the following animals could be placed in the same group as the animal above?

    A) Salamander

    B) Weaver bird

    C) Chameleon

    D) Whale

20. In finding out the relationship between pressure and depth in liquids, Std4 Pupils carried out the experiment as shown below.

Which one of the following holes below poured water least?

21. Std 5 Pupils were investigating certain aspect in science as shown belo

Which one of the following aspects were they likely to be investigating?

    A) Effect of heat on temperature

    B) Expansion and contraction in liquids

    C) Expansion and evaporation of gases

    D) Expansion and contraction of gases.

22. The diagram below represents a set-up that was used to demonstrate a certain aspect of heat.

The aspect demonstrated was_____

    A) Expansion in different solids

B) Rate of melting in different solids

C) Rate of conduction in different solids

D) Effect of heat on matter.

23. Which one of the following does **NOT** work the same way as a bicycle dynamo?

    A) Wind driven turbines

    B) Car battery

    C) Geo thermal generators

    D) Petrol and Diesel generators

24. Which one of the following is **NOT** a safety measure against electricity?

    A) Avoid carrying metallic and sharp objects when it is raining.

    B) Avoid using appliances with damaged wires

    C) Avoid operating electrical equipment with wet hands

    D) Avoid playing near equipment with the sign Danger or HATARI.

25. Which one of the following will be the best way of reducing friction in liquid and gases?

    A) Using Rollers

    B) Streamlining

    C) Smoothening

    D) Using lubricants.

26. Which one of the following is **NOT** an importance of cleaning the teeth?

    A) Remove food particles

    B) Prevent tooth decay

    C) Make the teeth strong

    D) Give fresh breath

27. Which one of the following is **NOT** a function of stomach as part of the alimentary canal?

A) Storage of food

B) Production of digestive juice

C) Breakdown of food

D) Digestion of glucose

28. Which one of the following statements about reproduction in human beings is **CORRECT**?

A) Conception means becoming pregnant

B) Fusion takes place in the vagina

C) Implantation takes place before fertilization

D) The embryo develops into a zygote

29. In which of the following activities is water not used for recreation?

A) Surfing

B) Water skiing

C) Boat racing

D) Fountain

30. Which one of the following is a minor nutrient?

A) Zinc

B) Manganese

C) Magnesium

D) Molybdenum

31. Std 5 Pupils constructed the weather instrument as shown below.In which direction will the

wind vane face?

A) West

B) North

C) East

D) South

32. Which one of the following traffic lights will communicate to the pedestrians to cross the road?

A) Green

B) Red

C) Amber

D) Green and Amber

33. Which one of the following is an immediate sign of conception?

A) Breast become larger, tender and sensitive

B) Menstrual period stop

C) Morning sickness

D) Movement of the foetus

34. Which one of the following animals is correctly matched with the way it is adapted to feeding?

A) Humming bird-Long, thick slightly curved beak

B) Chicken-Short, straight strong and thick beaks

C) Duck-Broad, strong, slender and serrated

D) Hawk –Strong, curved broad and long beak.

35. Which one of the following does not describe the effect of soil pollution on plants?

A) Pollutants interfere with air circulation in the soil

B) Pollutants interfere with growth of roots of plants

C) Pollutants interfere with absorption of water and mineral salts.

D) Pollutants interfere with respiration of animals living in the soil.

36. Which one of the following pairs of objects will float on water?

    A) Candle wax and bottle opener

    B) Rubber band and ten shilling coin

    C) Spoon and biro pen casing

    D) Candle wax and rubber band

37. During a demonstration to investigate a certain property of matter, soil was put in a small container and heated. It was observed that the soil produced smoke. This was to show the presence of_____

    A) Water in soil

    B) Air in soil

    C) Manure in soil

    D) Soil in organic matter.

38. Which one of the following poses the greatest effect of soil erosion?

    A) Amount of rainfall

    B) Activities by man

    C) Types of soil

    D) Inadequate ground cover

39. Std 5 Pupils were investigating certain aspects on the properties of matter as shown below. If the correct procedure was followed which of the following was **UNLIKELY** to be observed?

    A) Water level in the straw dropped at initial stage

    B) Water level in the straw gradually dropped when the candle was withdrawn.

C) Coloured water improved visibility in the straw

D) Water level in the straw rose immediately heating began.

40. The set up below was used to investigate how electric current flows in different substances as shown below. If the switch is closed, which one of these substances will give the same results as silver spoon?

     A) Cigarette foil

     B) Plasticine

     C) Rusted nail

     D) A piece of glass.

41. Which one of the following is **CORRECT** about sound?

     A) Low height produces loud sound

     B) Sound is slowest in liquids

     C) Sound travels fast in air

     D) Minimal vibration causes soft sounds.

42. Which one of the following is **NOT** an artificial way of lighting the house?

     A) Using pressure lamp

     B) Using translucent roofs

     C) Using gas lamp

     D) Using electricity.

43. Which one of the following organs excrete toxic substances from the body?

     A) Skin

     B) Lungs

     C) Kidney

D) Anus

44. Which one of the following does **NOT** describe animal's reaction to changes in stimuli?

    A) A Snail glides to look for food

    B) A Dog pants on a hot day

    C) A Millipede recoils when it senses danger

    D) A Lizard basks on a cold day

45. Which one of the following is **NOT** a source of electricity?

    A) Hydro electric dam

    B) Dynamo

    C) Wind driven machine

    D) Bulbs

46. Pupils introduced a lighted Kerosene lamp in a dark room. The activity was done to demonstrate that light_____.

    A) Is dispersed

    B) Passed through transparent materials

    C) Travels in a straight line

    D) Travels in all directions from the source

47. The diagram below represents a simple balance made by a pupil.Which one of the following changes would cause an increase in the reading on the scale?

    A) Driving the screw nail away from the fulcrum

    B) Moving the fulcrum towards the screw nail

    C) Placing a small weight directly on top of the fulcrum

    D) Placing a small weight like the drinking straw near the scale

48. Standard 6 Pupils planted seeds in a container as shown below. After some days the pupils realized that only 3 of the 4 seeds germinated. Which one of the following is the reason why one of the seeds did not germinate?

    A) The seed did not get enough moisture

    B) The seed lacked warmth

    C) The seed had a problem with the embryo

    D) The seed lacked enough air to germinate

49. Which one of the following groups of animals have a certain similar characteristics?

    A) Mites, Salamander, Ostrich

    B) Mites, Ticks, Spiders

    C) Trout, Newt, Slugs

    D) Spider, Mouse, Octopus

50. Which one of the following pairs of animals have their teeth adapted to the same type of feeding?

    A) Chimpanzee and Fox

    B) Pig and Chimpanzee

    C) Mouse and Cat

    D) Cat and Antelope