

**EFFECT OF MARKING SCHEME ON THE CONSISTENCY OF
SCORING MATHEMATICS EXAMINATIONS**

PAUL MUNGAI NG'ANG'A

**UNIVERSITY OF NAIROBI
SCHOOL OF EDUCATION
DEPARTMENT OF PSYCHOLOGY**

2014

**A Research Project Submitted in Partial Fulfillment of the Requirements for the Award
of Degree of Master Education in Measurement and Evaluation to the Department of
Psychology, University of Nairobi**

Copyright

**All rights reserved. No part of this work may be reproduced, stored in retrieval or
transmitted in any form or means whether electronic, photocopy, recording or
otherwise without prior permission of the author or university of Nairobi**

DECLARATION

This research report is my original work and has not been presented for any academic award at any other university.

Sign

Date

Paul Mungai Ng'ang'a

Reg. No. E58/71543/2011

This research report has been submitted for examination with my approval as University of Nairobi supervisor.

Sign

Date

Dr. Japheth Origa

Department of Educational Communication and Technology

University of Nairobi

DEDICATION

The research is dedicated to my dear wife Catherine, Our children Collins and Rachael, My parents Teresiah and James. Thanks for your patience, Support, advice and encouragement.

ACKNOWLEDGEMENT

I wish to express my sincere gratitude and appreciation to:

My supervisor Dr. Japheth Origa for his patience, guidance and scholarly advice which led to strengthening of the arguments presented here.

Special thanks to the lecturers who taught me and ensured that I acquired the necessary skills to make this possible and most specifically the course coordinator Dr. Karen T. Odhiambo.

The study respondents, who gave their time to be interviewed, mark and fill the questionnaires. Without their cooperation, this work could not have been completed.

Lastly, My colleagues at the University for their encouragement and contributions during group discussions.

ABSTRACT

It is common practice for secondary schools in Kenya to use common examinations to assess students' achievement. Proper monitoring of how formative assessment tests are typed and administered is usually done by school administrators. The results of these examinations must be fair and reliable as it is the only feedback given to students and parents. Little attention however has been given to how teachers mark these school examinations and therefore the validity of the scores given is always questionable. The study sought to examine the effects of a marking scheme on the consistency of scoring mathematics examinations. The objectives of the study were; to investigate the effect of a marking scheme on consistency of scoring mathematics examinations in schools; to establish if moderation of a marking scheme improves the consistency of scoring and to determine students' factors that may have any effect on consistency of scoring mathematics examinations. The study was guided by classical test theory and positivist perspective of assessment. The research employed descriptive research design. The target population was 156 mathematics teachers, 23 deans of studies and deputy principals in case a school did not have a dean of studies. A sample of 57 teachers, 10 deans of studies and 2 deputy principals was taken for the study. Descriptive statistics such as frequencies, percentages, means and standard deviations were used to report the data. Cronbach alpha a coefficient for internal consistency was also computed. The analysed data was presented in the form of frequency tables, bar graphs and pie chart. The study revealed that when a common moderated marking scheme was used, consistency of scoring mathematics improved. Organization of answers on the script, Handwriting of the student, knowledge of student and previous performance affected how teachers marked student work. However, personal feelings, gender, race, social class and physical attractiveness of the students did not affect the way teachers in Makadara Sub-county marked mathematics examinations. The study concluded that a well prepared marking scheme enhances consistency of scoring especially if it is prepared and moderated by all teachers marking the examination. The consistency could further be enhanced if a marking scheme is piloted and any difference in interpretation is discussed and adjusted accordingly. The study recommends proper training of teachers on essential assessment concepts, principles, techniques and procedures by teachers training institutions. In addition, Ministry of Education (MoE) in collaboration with Kenya National Examination Council (KNEC) should organize workshop on setting and marking of examinations. Finally, research should be conducted on the effects of feedback to examiners on consistency of scoring examinations. This study should also be replicated in other sub-counties for the purpose of comparing the study results.

TABLE OF CONTENTS

Declaration	iii
Dedication	iv
Acknowledgement	v
Abstract	vi
List of Tables	xi
List of Figures	xii
List of Abbreviations and Acronymns	xiii

CHAPTER ONE

INTRODUCTION

1.1 Background to the Study.....	1
1.2 Statement of the Problem.....	3
1.3 Purpose of the Study	4
1.4 Objectives of the Study.....	4
1.5 Research Questions	5
1.6 Significance of the Study	5
1.7 Limitation of the Study	5
1.8 Assumptions of the Study	6
1.9 Delimitation of the Study.....	6
1.10 Definitions of Key Terms	7

CHAPTER TWO

REVIEW OF RELATED LITERATURE

2.1 Introduction.....	8
2.2 The Effects of Marking Scheme on consistency of scoring.	8
2.3 Consistency of Marking Subjective Tests.....	9
2.3.1 Consistency of scoring and its impact to learners.....	11
2.4 Moderation of marking scheme and its effects on scoring	12
2.4.1 Moderation of student assessment	13
2.4.2 Consensus moderation	14
2.4.3 Multiple marking.....	14
2.5 Students’ factors and consistency of marking	15
2.5.1 Handwriting style and organization of students’ work.	15
2.5.2 Contrast Effects and consistency of scoring	17
2.5.3 Examiners’ factors and consistency of scoring.....	18
2.6 Improving Marking Consistency	22
2.7 Theoretical Framework.....	25
2.8 Conceptual Framework.....	27
2.9 Summary of the Reviewed Literature	29

CHAPTER THREE

RESEARCH METHODOLOGY

3.1 Introduction.....	30
3.2 Research Design.....	30
3.3 Target Population.....	30

3.4 Sample Size and Sampling Procedures.....	31
3.5 Data Collection Instruments	31
3.6 Validity of the instruments.....	32
3.7 Data Collection Procedure	32
3.8 Data Analysis Procedure.....	32
3.9 Ethical considerations	33

CHAPTER FOUR

DATA ANALYSIS, PRESENTATION AND INTERPRETATION

4.1 Introduction.....	34
4.2 Demographic Information of the Respondent.....	34
4.3 Marking Scheme and Consistency of Marking.....	36
4.3.1 Marking Scheme properties.	42
4.3.2 Marking scheme preparation and usage.....	43
4.4 Moderation and consistency of marking.....	44
4.4.1 Reducing scores variation among teacher marking same scripts.....	46
4.5 Students’ Factors affecting consistency of scoring.....	48

CHAPTER FIVE

SUMMARY, CONCLUSIONS AND RECOMMEDATIONS

5.1 Introduction.....	50
5.2 Summary of the study findings.	50
5.3 Conclusion	52
5.4 Recommendations for policy.....	53

5.5 Recommendations for further research.....	53
REFERENCES	54
APPENDICES	60
Appendix A; Questionnaire for Teachers.....	60
Appendix B: Interview Schedule for deans of studies/deputy principals.	65
Appendix C: Mathematic Examination Paper.....	67
Appendix D: Mathematic Examination Paper Marking Scheme.	72

LIST OF TABLES

Table 3-1 Categories of the school sampled.....	31
Table 4-1 Gender of the teachers' respondents.....	34
Table 4-2 Academic qualifications of the respondents.....	35
Table 4-3 Mathematics Scores marked out of 100	37
Table 4-4 Group Statistics for the scores awarded	38
Table 4-5 Internal consistency reliability coefficients.....	39
Table 4-6 Time of preparation of marking scheme	41
Table 4-7 Evaluation of marking schemes properties.....	42
Table 4-8 Marking scheme moderation	44
Table 4-9 Ways of improving consistency of marking.....	45
Table 4-10 Factors affecting consistency of scoring	48
Table 4-11 Student's characteristics that can influence scoring	49

LIST OF FIGURES

Figure 2-1 Conceptual framework	27
Figure 4-1 Teaching experience bar chart	36
Figure 4-2 Marking schemes preparation	40
Figure 4-3 Reasons why teachers do not prepare marking schemes.	43
Figure 4-4 Ways of reducing scores variation	47

LIST OF ABBREVIATIONS AND ACRONYMS

CATs	Continuous Assessment Tests.
CBS	Conveyer belt marking system.
CE	Chief Examiner.
GCSE	General Certificate of Secondary Education.
HODs	Head of Departments.
IMAP	Integrated Moderation of Assessment.
KCPE	Kenya Certificate of Primary Education.
KCSE	Kenya Certificate of Secondary Education.
KNEC	Kenya National Examinations Council
MC	Multiple Choice.
MoE	Ministry of Education
OP	Open Ended.
PGCE	Post Graduate Certificate in Education.
QCA	Qualification and Curriculum Authority.
WWW	World Wide Web.

CHAPTER ONE INTRODUCTION

1.1 Background to the Study

The Kenya education system provides 12 years of general education – 8 years in the primary school and 4 years in secondary school. The delivery of the curriculum is assessed through school based and end of the school cycles examinations (Sessional paper No. 14, 2012). The school based examinations are both diagnostic and for placement to determine progression from one level to another. Undoubtedly, assessment is an integral part of teaching and learning process as it provides teachers with vital information about students learning progress. McCormack and Jones (1998) views assessment as any mechanism involving information gathering that can be used to improve instruction and learning. At school level, two categories of testing-continuous assessment tests (CATs) and End of term examinations form the basis of ascertaining students learning progress through their school years. The CATS and end of term examination are set, administered and scored by teachers. The continuous assessments are meant to ascertain how well or otherwise a unit of learning has taken place, and to assist teachers in the pacing of the contents of their lessons and to design and implement remedial strategies when needed.

The suite of public examinations in Kenya includes Kenya Certificate of Primary Education (K.C.P.E), Kenya Certificate of Secondary Education (K.C.S.E) and a range of applied and vocational qualifications. The results of these exams are used as entry “tickets” for higher education training or employment opportunities. It is therefore essential that these public examinations and all examinations be marked as accurately as possible, ensuring fair result for all. The Kenya national examination council (KNEC) administers these national examinations at the end of a cycle which are uniform throughout Kenya irrespective of region or the cultural / urban divide (Sessional paper No. 14, 2012). KNEC secondary school examinations are marked by teachers who usually have undergone training and issued with certificate. KNEC usually ensures that marking is standardized. The purpose of the standardization along other control measures is to enable valid and reliable marking. Moderation takes place during standardization meetings. When marking start, chief examiners monitor marking to ensure consistency in marking. In most cases, the Chief

examiner's mark and teachers mark vary with an error of ± 3 but anything above that will trigger a remark of the entire script. If school based tests score are to be used to improve teaching and learning, test scores must be reliable. One way to improving the reliability of this test score is to use a common marking scheme

Teachers and administrators need to not only know that there is error in all classroom and standardized assessments, but also more specifically how reliability is determined and how much error is likely to occur (McMillan, 2000). With so much emphasis today on high – stakes testing for promotion, graduation, teachers and administrators accountability, and school accreditation, it is critical that all educators ensure that test score are reliable. According to Brimi (2011), student grades present quantifiable evidence of student achievement, open the doors to higher education and still frequently determine how students view themselves. Standardized achievement tests are administered and scored under uniform conditions making it possible to compare the achievement level of one group against that of a norm or reference group or against some standard of mastery (Sax, 1997). In a study by Brown and Hudson (1998), it was found that the problem with formative assessments in schools lies with their validity and reliability. The tasks are often not tried out to see whether they produce the desired information. Marking criteria are often not investigated to see whether they work and raters are not often trained to give consistent marks. Some of the assessment used in schools may have high face validity – they may look excellent to the unformed but they may be marred by inappropriate marking schemes and rating inconsistencies. This may produce invalid results which are unfair to student and teachers alike (Clapham, 2000)

The validity and reliability of the school based assessment has been questioned as most of these tests are not standardized. According to Meadows and Billington (2005) there has been an assumption on the part of the public that the marks awarded to candidates in high stake examination such as K.C.S.E and G.C.S.E are highly reliable and a true reflection of the candidates' abilities. There is however a long history of research findings to the contrary. As early as 1912 Starch and Elliot reported a study of the reliability of the marks assigned by teachers to English examination scripts. There was wide variation in the marks given to the

same script. Starch and Elliot replicated their study in the marking of mathematics (1913 a) and History (1913 b).The researchers expected to find more consistency in marking mathematics than English, but found that the marks varied even more widely. The researchers concluded that this was due to some teachers taking into account poor appearance of the scripts, which others ignored. Teachers also came from schools with varying levels of achievement, which they believed could have affected their grading (Brimi, 2011).

According to Harvey (2012), teachers should be observed, but watching teachers teach is not enough. School principals should know how teachers are assessing and grading students. Continuous assessment must inform teaching and learning processes that are in tandem with Kenya's vision 2030 and also enhance learner achievement skills and competences (Sessional paper no.14, 2012).Harvey (2012) point out that of the many tasks that teacher performs, assessment is one of the most essential. Principals should therefore ensure that teachers offer valid tests and quizzes as well as formative assessments. They must be able to offer teachers suggestions for improving assessments as needed. Several authors have argued that there are a number of essential assessment concepts, principles, techniques and procedure that teachers and administrators need to know about. There is however little emphasis on assessment in the preparation of or professional development of teachers and administrators (Stiggins 2000).

1.2 Statement of the Problem

School examinations are routinely designed and administered by teachers to assess students learning capabilities .Output from these examination are often used to support decision making such as giving grades to students or assigning students to remedial classes (Alias, 2005).In order to be of real value in decision making, these school based examinations must possess two important characteristics namely: validity and reliability. The researcher had noted that when two or more teachers are teaching the same form, the school administration usually require them to use common examination to assess student's achievement .There was however no rule that required teachers to use common marking schemes or even moderate the marking schemes. This resulted to score variation among secondary school teachers marking same script.

The researcher had also noted the tendency of some teachers setting items that were not workable or items they do not understand especially in mathematics .As a result, the marking criteria for such items were many depending on the number of teachers marking the items. Brimi (2011) pointed that a more difficult obstacle to improving assessment lies in the teachers own content knowledge or lack thereof. Stiggins (2000) actually questioned the ability of teachers to assess content they had not themselves ‘mastered’. This study sought to examine the effect of a marking scheme on the consistency of scoring mathematics examinations.

1.3 Purpose of the Study

The purpose of the study was to investigate the effects of a marking scheme on consistency of marking mathematics examinations in secondary school in Kenya. The study explored the consistency with which teachers applied the marking schemes. This was done by measuring teachers’ divergence from the ‘true’ script score. The true score in this study was taken as the score given by the chief examiner who was the setter of the examination paper.

1.4 Objectives of the Study

The objectives of the study were:

- i. To investigate the effects of a marking scheme on the consistency of scoring mathematics examination.
- ii. To establish if moderation of a marking scheme improves the consistency of scoring mathematics examinations.
- iii. To determine students’ factors that may have any effects on consistency of scoring mathematics examination in schools.

1.5 Research Questions

The study sought to answer the following research questions.

- i. What effects does a marking scheme have on consistency of scoring mathematics examinations?
- ii. Does moderation of a marking scheme improve the consistency of scoring mathematics examination?
- iii. Which students' factors affect consistency of scoring mathematics examination in schools?

1.6 Significance of the Study

This study exposed essential assessment concepts, principles, techniques and procedures that teachers and administrators should know about school assessment. The policy makers could draw from the findings of this research report to reflect on issues relating to school based examinations and make appropriate policy, such as training and in – servicing teachers on assessment techniques. This could ensure that both teachers and administrators have necessary knowledge, skills and attitudes on marking examinations which is important in improving teaching and learning. The study also add to the body of knowledge on assessment at school level especially how teachers should mark mathematics examinations and role of school principal when it comes to setting and marking exams

1.7 Limitation of the Study

The study was limited to respondents' bias and truthfulness. The researcher depended on the cooperation of the respondents to mark examination script provided and later fill the questionnaires. Some of the teachers felt that the task of marking was too demanding for them. Three teachers did not return the mathematics script and the questionnaires . Time was also a limiting factor considering that the study needed to be concluded within short time. The researcher was also financially constrained as no sponsor funded the study.

1.8 Assumptions of the Study

The study was based on the following assumptions; the respondents cooperated and gave accurate information; sometimes teachers use different marking schemes to mark mathematics examinations. If common marking schemes are used, they are sometimes interpreted differently and this affects the consistency of scoring among teachers.

1.9 Delimitation of the Study

The study was conducted in Makadara Sub-county in Nairobi, Kenya. The researcher restricted the study to finding the extent to which examination score are affected when secondary school teachers used different marking schemes to mark same mathematics examination script. Effects of moderating marking scheme on consistency of marking were also explored. Finally the study sought to investigate possible causes of scoring inconsistencies among teachers like training, student handwriting and organization of the answers on the scripts.

1.10 Definitions of Key Terms

Mathematics examinations: - These are test used by teachers to generate information on the level of acquisition of the subject matter by the learner in schools.

Marking scheme:-marking scheme is a set of criteria used in assessing student learning.

Scoring:-A process in which a teacher or a marker gives a letter or number to show how good a student work is.

Grading: - A process in which a teacher or a marker gives a letter or number to show how good a student work is.

Consistency; - refers to the degree of similarity between different examiners: can two or more markers, without influencing one another, give the same marks to the same set of scripts.

Double marking:-Is defined as a method of marking assignments where scripts are marked independently by two markers who then meet and arrive at agreed marks.

Moderation:-A process aimed at ensuring that marks and grades are as valid, reliable, and fair as possible for all students and markers.

Students' factors:-These are Learners' characteristic that may affects how student work is scored.

CHAPTER TWO REVIEW OF RELATED LITERATURE

2.1 Introduction

This chapter reviews literature related to study objectives. It includes literature on how a marking scheme affects consistency of scoring, effects of moderation on consistency of marking and factors that affects consistency of scoring. Theoretical and conceptual framework for this study is also discussed in this chapter.

2.2 The Effects of Marking Scheme on consistency of scoring.

Examiners' marking to common standard and common interpretation of marking scheme is important to ensure no student is disadvantaged or favoured. In addition to disadvantaging or benefiting those sitting an examination, aberrant marking can also affect the integrity of an award and / or qualification by inflating or deflating pass rates (Powell – Howard, 2009). The initial development of marking guidelines with input from the marking team may enhance the breadth and depth of the guidelines and increase the commitment and understanding of the marking guidelines (Rone-Adams and Naylor, 2009). However, O'Donovan (2005) noted that making schemes in the Humanities are more complex to apply as they are 'content-advisory' rather than 'content specific'. This is in contrast to examination papers in sciences that tend to use questions that are more clearly right or wrong.

Research has revealed that an unsatisfactory marking scheme can be the principal source of unreliable marking. Delap (1993) conducted marking reliability studies to determine the extent of any unreliability in marking and to provide diagnostic information useful for examiners to minimize the source of variation in the marking between examiners. Following the re-marking of Business studies and Geography scripts, meeting were held with examiners to discuss the results and any difficulties they experienced during marking. In both subjects, the source of most difficulties was traced back to the marking scheme. In particular, there was widespread confusion amongst examiners over the use of a 'level of response' marking scheme in which examiners were required to place a candidate within a specific level based

on level descriptors. Price and Rust (1999), argued that with some exceptions, the introduction of detailed assessment criteria leads to improvement in marking consistency.

Moskal and Leydens (2000), in their work on how teachers can improve reliability of their assessment of students work, they argued that improving the scoring rubric is likely to improve both inter rater and intra rater reliability. They postulate several questions that may be useful in evaluating the clarity of a given rubric: are the scoring categories well defined? Are the differences between the score categories clear? And would two independent markers arrive at the same score for a given response given the scoring rubric? Moskal and Leydens suggested that if the answer to any of these questions is no, then the unclear score categories should be revised. They recommended the use of exemplar. These are a set of score response illustrating the nuances of scoring rubric. The marker may refer to the exemplar throughout the scoring process to illuminate the difference between the score levels. They also suggested that the rubric be piloted. Any difference in interpretation should be discussed and adjusted to the rubric negotiated. This can take time but greatly enhance reliability.

2.3 Consistency of Marking Subjective Tests.

Marking subjective tests can be extremely challenging. Numerous factors influence the reliability of scoring (or marking) subjective tests. Meadows and Billington (2005) warns against confusing examiners reliability and examination reliability. According to Aslett (2006) reliability relates to the fairness and consistency of assessment. Unreliability in marking is only one factor influencing the overall reliability of the assessment.

Whereas the scoring of multiple- choice (MC) item is considered objective and highly reliable, the scoring of open –ended (OE) item has a subjective component; this kind of scoring is less reliable than MC because OE involves human raters and is affected by their input (Allalouf, Klapter & Fronton, 2008).As Sandler (2009) noted, subjective judgments’ made by different markers about the same pieces of student work differ from marker to marker, sometimes wildly. According to Sandler, some markers are characteristically generous, some are strict and others may be inconsistent. Some markers are influenced by aspects other than the actual quality of the student work. In order to reduce subjectivity inherent in the scoring of OE items and improve its reliability, Allalouf, Klapter & Fronton

(2008) suggests the engagement of professional raters, using comprehensive rating instructions, training the raters, monitoring the rating process, using retraining when drift is detected, having multiple raters and engaging the services of an additional raters incase of discrepancy between the raters. Bond (2009) suggests the use of common examinations and formal guidelines for the distribution of grades to lower grading variability across scorers. Sandler (2009) has argued that the influence of assessors' personal standards, tests and preferences p. (809) should be barred from the marking process in order to avoid a wide variety of sub – optimal practices.

William (2000) argues that, although test can be made reliable by improving the items included, and by marking more consistently, the effect of such changes is small. The most effective way of increasing the reliability of a test is to make the scope of the test narrower, or make the test longer. It is however important to enquire into the basic underlying problems in terms of the causes and reasons why different assessors give different marks.

Marking reliability studies are an important aspect of quality control of an assessment process that affects candidates' life chances and has implications for teachers and school. According to Meadows and Billington (2005) by 1970s it was clear that marking reliability was dependant on the subject area being assessed. The least reliably marked examinations tended to be those that placed the most dependence on essay-type questions and the most reliably marked tended to be those made up of highly structured questions. Taylor (1992) considered the reliability of marking GCSE English, History, Mathematics and GCE Psychology coursework. In each subject, previously moderated work was re-marked by two further moderators (thus four marks were available for each candidate: the centre mark, the original moderator's mark and the marks awarded by the two 'project' moderators).In mathematics, despite the fact that coursework was not as highly structured as the traditional written papers, the correlation coefficient between two moderators re-marking coursework folders ranged between 0.91 and 0.97 for different pairs of moderators. The coefficients were similarly high for English, ranging between 0.87 and 0.97.Despite these high coefficients, it was found that, if candidates involved in the study were re-graded on the basis of their re-mark scores, approximately 20% would have received different grades.

2.3.1 Consistency of scoring and its impact to learners

Several limitations exist among all forms of assessment. Guskeys (1996) noted that no one method of grading and reporting serves all purposes well. Grading and reporting will always involve some degree of subjectivity. Teachers and administrators need to not only know there is error in all classroom and standardized assessment, but also more specifically how reliability is determined and how much error is likely (McMillan, 2000). According to Stiggins (1999) there have been several impediments to progress in overcoming the assessment illiteracy that has hindered educators, while teachers have at times realized their own limitations in the field of assessment; they have been largely incapable or simply unwilling to make changes to their current practices. Stiggins (1986) commented on the stark discrepancy between recommended practices and what actually occurs in classrooms. In terms of writing assessments, Hillocks (2006) attributed this reality to a combination of teacher's ignorance of research and or their indifference towards research. Brimi (2011) noted that teachers often lack the time or administrative support essential for professional development. Teachers have demonstrated that they use assessment as a motivator, not as an instrument for student learning i.e. they find that students are more willing to pay attention, to "learn" if they know that class material will be tested (Kahn, 2000; Stiggins, 1999). Brimi argues that a more difficult obstacle to improving assessment, lies in the teachers own content knowledge, or lack of thereof. Stiggins (1999) questioned the ability of teachers to assess content that they have not themselves "mastered,"

In a study by Brimi (2011) to investigate the reliability of grading high school work in English, 90 teachers were trained to score single essay paper, despite several sessions of training in using the same grading methods, the teachers awarded final scores that were discrepant. In that study, the range of score for the single paper within the school district was 46. Brimi concludes that English teachers within this district evaluated writing differently and as a result, a wide range of scores exist for the same quality of work. Wang (2009) chose one composition from examinees' and eight examiners to mark the composition individually. The raters who marked the examinee's writing were all experienced teachers and specialists in the field of English. Ratings of examinee's essays were carried out using holistic and the analytic methods. The results showed that even experienced raters give different scores

although the range of score was small. Wang also noted that analytic methods scores were usually higher than that of holistic scoring. Rone-Adams and Naylor (2009) investigated if there was significant difference between marks as well as the comments given on research proposal that was double marked by two faculty members, where faculty were blind to the other marker's mark and comments. The study indicated that there was a wide discrepancy between the classification of the research proposals, the marks given and the comments given by two independent markers.

Brimi (2011) concluded that there are several large-scale implications of subjectivity in grading. Grades help to determine which student's colleges admit and which student receive scholarships. If students qualify for scholarships or admission based on inflated grades, then their college experience may be marked by futility and the funding effectively rendered a lost investment when such students fail to earn a degree. Also, as teachers garner reputations as easy or hard graders, students may increasingly pressure administrators and guidance counselors for preferable placement. According to Rone-Adams and Naylor (2009) most research in these areas has concluded that there are ways to obtain better argument between makers.

2.4 Moderation of marking scheme and its effects on scoring

Meadows and Billington (2005) observed that despite the pervasive view that a clear and detailed marking scheme result in higher marking reliability, intended improvement to marking scheme do not always bring about expected improvement in reliability. In the Braid, Greator, and Bell (2004) study, the impact of standardization meeting were investigated. In the study, examiners were provided with marking scheme and some examiner were provided with exemplar script and given the feedback about the marking of those scripts. In the second study, the effects of the discussion of the marking scheme were explored. All examiners received marking scheme and exemplar script, but some examiners did not attend a standardization meeting. The study found that neither the process (use of exemplar script or discussion between examiners) demonstrated an improvement in marking reliability. However, these findings contradict the research undertaken by the same researcher Greator, Baird, and Bell, in 2002. They noted that examiners think that the standardization

meetings are reliable because it help them understand the marking scheme and makes the principal examiner's interpretation of the marking scheme clear. Some examiners also noted that attending standardization meeting gave them confidence to know that they were marking appropriately and had the same understanding of the marking scheme as that of their fellow examiners.

2.4.1 Moderation of student assessment

Moderation of student assessment is a process aimed at ensuring that marks and grades are as valid, reliable, and fair as possible for *all students* and *all markers* (ALTC, 2012c). Moderation of assessment checks that marking is consistent such that an assessment item would be awarded the same mark by any marker. Moderation is necessary whenever more than one person marks assessment items in a unit and when a unit is taught in more than one school. According to (Sadler, 2009) the task of moderation is to minimize discrepancies among assessors *before* students receive their marks. Moderation is more than checking the marks; it is the checking of assessments from the development of each item to ensure that the whole assessment process is fair, valid and reliable enabling equivalence and comparability (ALTC, 2012a).

The relationship between student assessment and grading, quality assurance and academic standards has been a major issue (James, 2003). Ensuring consistency of assessments in a unit, and even moderation of these assessments, is a challenge when a unit is offered on more than one campus and also on-line (Kuzich, Groves, O'Hare, & Pelliccione, 2010). Marking and grading in most disciplines is inevitably subjective (Hughes, 2011) but a systematic approach to identifying significant tacit beliefs may assist in reducing the effect on grader variation (Hunter & Docherty, 2011). Conversations amongst markers assessing student performances influence how the group of markers reach to an agreement (Orr, 2007)

If broad categories are used as the basis for grading students' work or attainment of skills and knowledge, then grading becomes overtly judgmental and subject to many psychosocial pressures (Yorke, 2010). After markers had participated in a professional development using an integrated moderation of assessment program (IMAP), variation between markers tended to decrease (reliability increased) particularly when they were divided into novice and

experienced groups (Bird & Yucel, 2010). Also time taken to mark tended to decrease so efficiency of marking increased after participation in the professional development (Bird & Yucel, 2010).

2.4.2 Consensus moderation

Academics who share the marking of large batches of student works can collaborate on how marks are allocated. This is the principle behind the approach known as *consensus moderation* (Sadler, 2009). In its most common form, consensus moderation requires that all assessors mark the same sample of student responses with, or without, prior consultation among themselves. They then discuss the results of their marking in order to arrive at a common view about the grading 'standards' to be used for the whole student group. After the trial marking and conferring, the bulk of the marking may be carried out more or less independently, with only occasional cross-checks.

2.4.3 Multiple marking

This approach also applies to student responses to a single assessment task within a course, but it does not depend on professional consensus. Two or more markers score all students' responses. The separate scores are then subjected to statistical or mechanical 'moderation', which is simply a method of combining them. The simplest method is to average the scores from different markers, with no attempt made to arrive at inter-markers consensus. With three or more markers, a variation on this rule can be to first eliminate the most extreme score (if any) and then average the remainder. (This process is similar to that used in judging and scoring individual performances in certain competitive sports.) Statistical moderation can be – and usually is – carried out without scrutinizing and discussing actual students' responses. In some implementations, the specified calculations are implemented automatically on a mark spreadsheet as soon as the numbers from different markers are entered. In some UK institutions, double marking followed by averaging is standard practice for all extended student responses. Multiple marking is labour intensive (and therefore relatively expensive) for large course enrolments (Sadler, 2009).

2.5 Students' factors and consistency of marking

Research has shown that examiners can be influenced in their judgment by characteristics of the candidate, as well as order of marking and script presentation (Wade, 1978). Candidate's characteristics include gender, race, social class and physical attractiveness. Studies by (Aslett, 2006, Owen et al 2010) have shown that we tend to attribute more favourable character traits to people who are physically attractive e.g. more intelligent, friendly, sensitive etc. When David Laudy and Harold Signall (1974) attached a photograph of an attractive student to an essay, they found it received a higher grade than the same essay with a photograph of an unattractive student or "no photograph" attached. Knowledge of the student and previous performance can also affect a student's score. Aslett (2006) noted that if a student who normally gets high marks submits a poor assignment, allowances may be made due to halo effect- which is why many courses use 'anonymous marking' where the identity of the student is not included with the assignment. In contrast, Baird (1999) found examiners could identify the gender of a candidate from their handwriting style with an accuracy rate of 75 percent. This has led to reservation about the effectiveness of blind marking for completely eliminating gender bias. A feasibility study of anonymous marking in GCSE English, conducted by Baird and Bridle (2000) concluded that concealing a candidates' name from examiners is far from a panacea for marking bias, as handwriting style, the content and the style of the language used reveal personal characteristics of the candidate. Perhaps, a more effective solution for a gender bias in marking would be to provide an examiner with detailed evaluating criteria (Bridle, 2000).

2.5.1 Handwriting style and organization of students' work.

Evidence suggests that the marks teachers' award to pupils' work is at times influenced by neatness of the handwriting. Whereas good handwriting enables the teacher to discern easily what the pupil is trying to communicate, poor handwriting makes the task of reading more difficult. Henry (2012) noted that a neat handwriting bring students extra marks. Henry also noted that students who provide longer stories or had very neat writing were also more likely to receive better marks, regardless of the quality of their writing. According (Hart, 2010) poor handwriting sometimes includes sentences that are poorly punctuated, misuse of upper and lower case letters, words with incorrect spacing and sentences that do not make sense.

Essays that are harder to read require more time and effort. Hart noted that an essay may be misinterpreted entirely simply because the teacher misunderstood a word or a sentence. Sometimes the teacher might even give up on trying understanding what the student is saying. These are the sorts of paper that received a bad grade.

Chase (1983) compared scores on two essays, each correct in spelling and grammar, but one constructed to be at a difficult reading level, the other at a less difficult level, but with a common text base, to see how different levels of readability influence essay test scores. Although the readers were all graduate students who had experience with reading material that ranged in difficulty, the essay written at a difficult reading level was scored lower than the essay written at an easier reading level. Chase concluded that variables that complicate the reading of an essay, spelling errors, grammar errors, and poor handwriting and so on reduce the marks assigned to the work. Massey (1983) explored whether text effects are confined to teachers marking or whether they also affect the marking of experienced examiners. The findings suggested that experienced examiners are not susceptible to the biasing effects of handwriting style and presentation. According to Massey a well defined marking scheme and good community of practice brought about by well-managed standardization meetings, found in today's public examination might reduce the effect of presentational style. Nonetheless one obvious counter measure to allay concerns over the effects of handwriting style and presentation on the marks awarded is to have candidate type their work where possible. One of the benefits of teachers requiring a typed assignment is that everyone can profit from aesthetically similar papers. With hand written paper, this is not the case (Hart, 2010). There is evidence however that assessors judge typed scripts more harshly than handwritten scripts (Graig, 2001; Russell, 2002).

Graig (2001) investigated the issue of handwriting quality and word-processing as biasing factors in English as a second language testing. Four experts rated 40 essays, 20 original and 20 transcribed in either messy or neat handwriting or on a word processor. Word processed essays were scored lower than their handwritten counterparts. There was no effect of handwriting legibility. Other studies have shown that a small but consistent effect when marking handwritten originals and their typed transcripts (Powers et al 1994, Russell & Tao,

2004). Russell and Tao (2004) concluded that computer printed script would score on average 1.3 point less than the same words in a handwritten script. This study agrees that markers may indeed be influenced by format and that difference might be worth almost 2 marks to the average students. Such variability could of course be controlled by ensuring all markers are only given script in one format but the cost of transcribing large numbers of scripts almost certainly render this impractical. Russel and Tao however demonstrated that marking essays in cursive font, and alerting the markers to the format effect, had the effect of reducing difference in the score, and both approaches may be practical to implement.

2.5.2 Contrast Effects and consistency of scoring

Several studies have found that marks awarded to an essay may depend on the quality of those marked before it. If it follows a poor assignment, there is a tendency to award a higher mark, but if it follows a really good assignment, there is tendency toward a lower mark (Aslett, 2006; Owen et al, 2010). Spear (1996, 1997) found that good work tend to be assessed more favourably when it followed work of a lower standard than when it preceded such work. Poor quality work was assessed more severely when it followed work of higher quality. Hughes, Keeling and Tuck (1980a and b) found that good and poor essays were susceptible to contrast effects than were average quality essays. They also found that contrast effects tend to disappear after a number of essays have been marked. Hughes et al believed that by this time marking standards had become established and consequently markers were less susceptible to contrast effects. Vaughan (1991) provided qualitative evidence of contrast effects. Vaughan made raters read through and holistically grade essays whilst verbally commenting into a tape recorder. Analysis of the transcribed tape revealed a tendency for the essay to become one long discourse in the rater's mind. Rater made comparative statements such as "this essay is better/worse than the previous one or than others" as they led.

A study by Hughes et al (1983) sought to eliminate context effects by explicitly warning markers about their influence and also requesting that markers to categorise essays qualitatively before re-reading them and awarding final grade. The results of these procedures were compared with those obtained by markers who were merely warned of the existence of context effects and with those obtained by makers who were given no

information about the influence of the context. The result showed that all the three groups were influenced by context effect and to about the same extent.

In a final attempt to control context affects, Hughes and Keeling (1984) provided markers with model essays. Context effects persisted despite the use of a model essay during marking. Although the possibility remains that the provisions of models may lessen the influence of context on the marking of essays in subject areas where factual accuracy rather than written communication is being assessed. Hughes and Keeling concluded that, when written expressions are the primary focus of assessment, “We may be forced to accept context effects as an unavoidable concomitant of essay scoring” (p. 238).

2.5.3 Examiners’ factors and consistency of scoring

A number of studies have attempted to identify factors which allow awarding bodies to predict those examiners who are likely to mark most reliably and those who are likely to require additional training or monitoring. Meadow and Billington (2013) have noted that compared to experienced markers, inexperienced markers tend to mark more severely and employ different rating strategies. Training may however remove these differences. Not all studies have replicated the relationship between inexperienced and marking severity, for example. Meyer (2000a, 2000b) investigated marking in GCSE English Literature and Geography and found that the length of experience and senior examiner rating of the markers’ performance rarely proved useful as predictors of whether an examiner’s marks would require adjustment to correct for severity or generosity.

Meadow and Billington (2013) investigated the effect of markers background and training on quality of marking in GCSE English. The study was conducted in a marking center. Participants marked 100 part-scripts using the marking scheme. They then received the standardization training which replicated as closely as possible the training used in the live examination. Participants then marked another 99 part-scripts after training. The group of markers that participated in the study included; experienced GCSE English markers, postgraduate certificate in education (PGCE) English teachers, English/linguistic undergraduates and undergraduates of other disciplines. The result of the study showed that

before training, the marking of examiners and PGCE student was on average half a mark more generous than that of undergraduate group. After training, all groups become more generous to an equal extent. The range of marks awarded to part-scripts marked following training was slightly less than those marked before training. The researchers concluded that background had no effect on the marking accuracy of part-scripts. However, background was noted to affect marking consistency. Training improved accuracy to the same extent whatever the markers' background was but the effect was small.

In a study commissioned by the department for education and conducted by the national centre for social research in United Kingdom, it was found that staff allows "bias" and "personal feelings" to influence their marking (Henry, 2012). The research involved more than 2000 teachers judging essays written by 11 years old pupil over the course of a year. The overall marks awarded to pupil were then double checked by specially trained external "moderators". They discovered that one in ten cases; teachers had marked the work too favorably. In a 5 percent of cases, the work was marked too harshly. Nearly two thirds of the moderators said they thought that teachers' personal feelings about particular pupils influence their assessments on some occasions or on a regular basis. According to Henry, the findings cast doubt on teacher's objectivity and undermine calls from teachers unions and some academics for internal assessment to replace external test at primary school level. Henry argues that even if teachers were aware of their prejudices, trying to compensate for them would not make their assessments reliable. The only fair way to test children is through externally set and externally marked exams.

Aslett (2006) found that there are various psychological and physiological variables that affect examiners reliability. These include; mental fatigue due to monotony and lack of interest in a task which had severe implications with regards to task performance and accuracy. Aslett also found that lack of sleep can affect vigilance, attention, logical reasoning and rational thinking, Wolfe, Moulder and Myford (2001) developed the term differential rater functioning over time (DRIFT) which was used to describe how the accuracy of a single examiner decreased over time due to fatigue and lack of attention control as a result of drift condition. Equivalent answers marked earlier by an examiner were found to receive

significantly different marks to answer marked later. In a study by Klein and El (2003) they also noted that papers marked earlier in a marking session were awarded significantly lower marks than later marked papers. Emotional factor can play a part in the marks that examiners award (Aslett, 2006). This was demonstrated to be most obvious when examiner were aware of the identity of the student whose work they are marking. As Aslett noted, “whilst an assessor would hope to remain as objective as possible throughout the assessment process, where a marker is aware of a student’s identity their marking can potentially be profoundly affected” (p. 87)

Suto and Nadas (2008) found that the level of a marker’s highest education achievement (either in general or in a relevant subject) is essentially a better predictor of accuracy than either teaching or marking experience. Graduate in relevant subjects but with neither teaching nor marking experience were able to mark as accurately as individuals with both teaching and marking experience. Suto and Nadas concluded that education of an examiner is more important than experience (p. 10). They however noted that the key to successful marking is being able to follow marking instructions and interpret the marking scheme in the way its author intended” (p.10). According to (Powell-Howard, 2009) somebody may have a high level of qualification but still need some form of instruction and training on how to apply a marking scheme .Some degree of aptitude for the role of being an examiner was also necessary.

In England, Royal–Dawson and Baird (2009) explored whether teaching experience was a necessary recruitment criteria for marking national curriculum test in English taken at age 14. They compare the marking accuracy of four types of markers with an academic background in English but different years of teaching experience. English graduate, trainee teachers, teachers with three or more years’ teaching experience, and experienced markers. Accuracy was defined in two ways: the absolute difference between the participants’ marks and those of the most senior examiner, and the absolute difference between the participants’ marks and the mean mark awarded by all participants. This Dual approach had an advantage of not assuming perfect marking by a senior examiner and that there is only one valid judgment about the mark that any response is worth. Overall, whichever definition of accuracy was

used, there was little difference in the accuracy of the different types of markers. There was however evidence to suggest that classroom experience was needed to accurately mark curriculum specific item relating to Shakespeare. Interestingly, experience of marking seemed to reduce the accuracy of marking of a reading task. Royal-Dawson and Baird suggested that the task might have varied from that of previous years in which the markers had experience and that there had been some negative transfer of early training. They went on to propose a rudimentary model for the allocation of markers with varying levels of expertise to different item type. The model was based upon the level of detail encompassed by the scoring criteria and the level of curriculum specificity of the items.

Similarly, Suto and Nadas (2008) compared the marking accuracy of experts and graduate markers in GCSE physics and mathematics. Experts had experience of both teaching and marking. Graduates had no teaching or marking experience but both groups had a relevant degree. Accuracy was defined as the proportion of raw agreement between the participant marks and those of the most senior examiner, although the study also reported relative and absolute difference between the participants' marks and those of the most senior examiner. There were very few differences in the accuracy of experts and graduates for either subject. The groups significantly differed on just one question (out of twenty) for mathematics and two questions (out of thirteen) for physics. In any case, the difference in accuracy was small. They came to a similar conclusion to that of Baird and Royal-Dawson (2009), that the selection criteria for GSCE mathematics and physics examiners could be relaxed. Questions requiring markers to use more complex reflective thoughts processes were marked less accurately than those entailing only simple "intuitive" judgments. Such differences could form the basis of a rationale for assigning particular questions to different marker groups with different levels of expertise.

Suto, Nadas and Bell (2009) conducted one of the most comprehensive investigations into the factor affecting marking accuracy. Their study focused upon an international Biology examination designed for 16 year-old. Forty-two markers participated, comprising five groups: experienced examiner, Biology teachers, graduates in Biology, graduate in other subjects and non-graduates. The design of their study enabled the investigation of the relative

effects on marking accuracy of marking experience, teaching experience, highest education in a relevant subject, and highest education in any subject. In addition, they explored three aspects believed to impact upon the demands of the marking task: cognitive marking strategy, complexity of the maximum mark the question is worth and difficulty of the question for the examinee. In general, marking accuracy was high and all the five groups marked questions requiring simple marking strategy extremely accurately. For those questions requiring more complex marking strategies, highest general education, highest relevant education, target grade and total mark were found to be the most important predictors of accuracy. However, with sufficient training, the accuracy of some markers with only GCSE or A level education was found to be comparable to that of many markers with highest qualifications. Having teaching experience and marking experience were significant but less important predictor of accuracy although highest general education and marking experience were closely associated making it difficult to partial out their effects. Suto et al recognized that a key limitation of the study was its reliance on a single definition of marking reliability, the proportion of exact agreement with the marks awarded by most senior examiners of the assessment. Thus the senior examiner was considered infallible and there was only one valid judgment about the marks that a response is worth. These assumptions are more likely to hold for the point based marking of relatively short answer in say mathematics than the level of response based marking of longer answer in English for example.

2.6 Improving Marking Consistency

A number of studies have attempted to identify factors that might and the identification of those individuals likely to mark most reliably and those who are likely to require additional training and monitoring (Meadow & Billington, 2013). While grade inflation has been a topic of much discussion (Johnson, 2003; Millet, 2010), one reason that not much has been done about grade inflation is that a unilateral lowering of grades might hurt the prospects of students (Rojstaczer, 2009). In contrast, lowering grading variability across faculty members may actually benefit students by facilitating unbiased choices of elective areas of study (Felton & Koper, 2005). One approach for lowering grading variations across faculty members is to use common examinations (Bond, 2009). However, as acknowledged by Bond “programs that regularly employ common examinations are still rare, primarily because they

require a significant investment of faculty time and effort.” Another approach for increasing grading consistency is to create formal guidelines for the distribution of grades. Bukenya (2006) suggested use of conveyer belt marking system (CBS) where each marker marks only a set of questions and then passes the candidate’s answer script to the next marker who will also just mark the set of questions allocated.

Wang (2009) cited seven ways in which high inter-rater reliability can be established; first is setting the standards. In a test with a large number of examinees, it is impossible for all examiners to have an equal say in determining scoring policy. This description assumes that there is a “chief examiner (CE)”, either alone or with a small group of colleagues set standards for marking and passes these onto the examiners who may mark centrally or individually in their homes. The second way is Training the scorers. The scoring of subjective tests should not be assigned to anyone who has not learned to score accurately from past administrations. After each administration, patterns of scoring should be analyzed. The individuals whose scorings deviate markedly and inconsistently from the norm should not be used again. The Third way is identifying candidates by number, not name. Scorers inevitably have expectations of candidates that they know. This will affect the way they score especially in subjective marking. Studies have shown that even where the candidates are unknown to the scorer, the name on scripts will make a significant difference to the score given. For example, a scorer may be influenced by the gender or nationality of the name into making predictions which can affect the score given. The identification of the candidate only by number will reduce such effects.

The fourth way of achieving high inter-rater reliability is setting the specific standards before the real scoring. After the test has been administered, the CE should read quickly through as many scripts as possible to extract scripts which represent “adequate” and “inadequate” performances as well as scripts which presents problems which examiners are often faced with but which are rarely described in rating scales; bad handwriting, excessively short or long responses which indicate that the candidate misunderstood the task etc. The next step is for CE to form a standardizing committee to try out the rating scale on these scripts and to set and record the standard. All of the marking members should be given copies of the scripts

selected by the CE, in random order, and each member should mark all of these scripts before the committee meet to set standards.

Wang (2009) also cited Sampling by the chief examiner or team leader as another way of improving consistency of scoring. Each examiner is expected to mark a certain number of scripts on the first day of marking. The team leader collects a small percentage of marked scripts from the examiners (often 10-20%) and reads through them again in order to be given an independent mark (that is called blind marking) to find whether the examiners are marking properly. The process of sampling should be continued throughout the marking period in order to narrow the differences in examinees.

The sixth way was the Use of “reliability scripts”. Examiners are asked to independently mark the same packet of “the reliability scripts” which have been marked by the standards committee earlier. The reliability exercise should take place after the examiners have begun marking “for real”, but early enough in the marking period for changes to be made to scripts which may already have been marked incorrectly by unreliable examiners. The afternoon of the first day of marking or the second morning would be suitable time. Last but not least, Wang (2009) cite Routine double marking for every part of the exam that requires a subjective judgment. This means that every composition should be marked by two different examiners, each working independently. The mark that the candidate receives for a piece of writing is the mean of marks given by the two examiners.

In order to aid reliability of the marking process, the Qualification and Curriculum Authority (QCA) (2009) GCSE and GCE code of practice suggests that marking schemes should include general instructions on marking, be clear and designed so that they can be easily and consistently applied. It should also allocate marks commensurate with the demand of question/tasks and include the marks allocated for each question/task and part of a question/sub-task, with a more detailed breakdown where necessary. QCA also suggests that a marking scheme should State the acceptable responses to each question/tasks or part thereof, with detail that allows marking in a standardized manner.

2.7 Theoretical Framework

Psychometric is a field of psychology that deals with a measurement of individual differences in terms of traits, skills and other character. Classical test theory is the most common measurement theory used. It is usually represented by the formula $X = T + E$ where X is the observed score, T is the true score and E is the error score (the influence of error on the measurement, also known as 'measurement error'). Wiliam (1993) see classical test theory as an attempt to capture the idea of a 'signal-to-noise-ratio' for assessment. It is based on the assumption that an individual's score contain error (noise) which can be decreased but never totally eliminated. The theory assumes that the error is random and normally distributed (Meadows and Billington, 2005). In an article focusing on the theoretical and philosophical context of assessment, Orr (2007) considers assessment research from both the traditional positivist perspective and the emerging poststructuralist perspective. Researchers with a positivist perspective believe that assessment can be objective, transparent and reliable, often considering mismatches between markers to result from an 'error of measurement'. Researchers with a poststructuralist perspective, on the other hand, believe that assessment is '...co-constructed in communities of practice and standards are socially constructed, relative, provisional and contested'. As a result, Sadler (2009) is perhaps representative of a researcher with a positivist view of assessment. For instance, in his article on the importance of the integrity of grades as representations of academic achievement, he argues that the influence of assessors' 'personal standards tastes and preferences' (p.809) should be barred from the marking process. He instead calls for the employment of 'standards referencing as the grading principle' and strategic policy decisions at an institutional level in order to avoid 'a wide variety of sub-optimal practices' that could emerge if policy decisions were left to individual academics or departments.

In contrast to Sadler, Medland (2010) seem to adopt a poststructuralist perspective of assessment. Medland research aimed at illuminating some of the values and beliefs that shape the professional judgement informing the way marking is undertaken by academic developers as a means of exploring why mismatches between markers in the same team occur. Participants' comments indicated the nebulous nature of academic development as a 'discipline', if indeed it can be described as such. This may be attributable, in part, to the

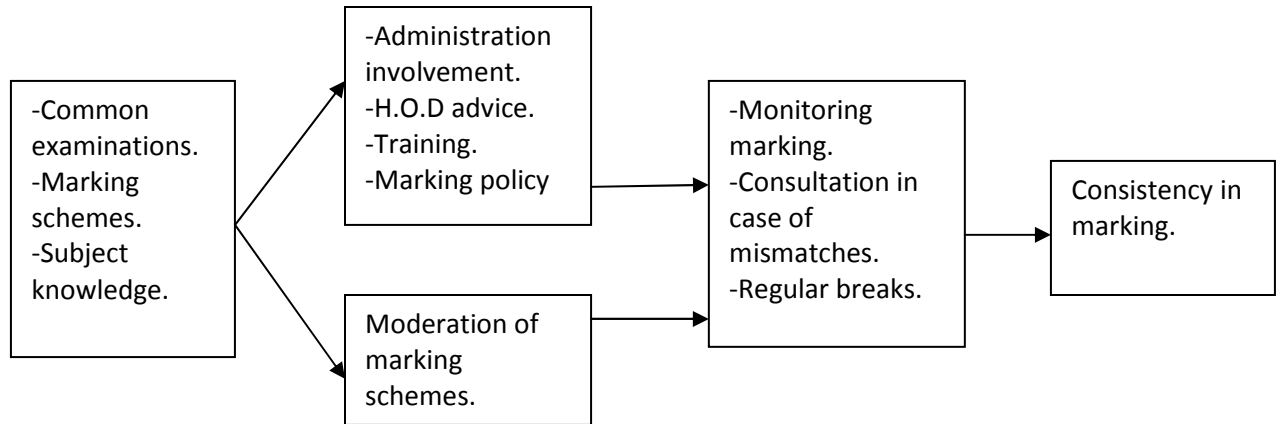
diversity of views concerning the underlying values and traditions of academic developers, who often hail from a variety of disciplinary backgrounds. However, differences in perspective were also believed to have a positive impact, often encouraging the articulation of personal values and beliefs and leading to the development of increasing self-awareness and a greater level of shared understanding amongst a team.

It is acknowledged that a large proportion of disagreement between markers can be problematic and result, as the external examiners commented, in increasing time being invested in marking. However, this research also indicated that the outcome of this disagreement can be highly useful in initiating discussions surrounding the personal values and beliefs that shape the manner in which one marks. Mismatches between markers can, therefore, be rewarded with an insight into the subjectivity that implicitly pervades the discourse of assessment. In other words, rather than the traditionally positivist perspective of the role of subjectivity in the assessment system as somehow compromising the integrity of the mark, it could instead be viewed as a tool for clarifying why mismatches between markers occur. When marking is viewed through this lens, subjectivity may then be reconceptualised as a potentially useful tool in developing greater levels of coherence between teams of markers Medland (2010). This research investigated the effect of a marking scheme on the consistency of scoring mathematics examination. The aim of the study was to find ways of reducing measurement errors in mathematics examination by using moderated common marking scheme in schools. Classical test theory and Positivist perspective of assessment therefore frames this research.

2.8 Conceptual Framework

Many secondary school in Kenya use common examination to assess student's achievement. There is also an attempt by most schools to standardize the examination using the subject teachers. All this is aimed at increasing the validity and reliability of school based examinations. However, if the marking schemes are not prepared and moderated then, examinations will not be marked accurately and the error in the test score will be high. This will result to wrong inferences and making a test that was initially valid to be interpreted as unreliable. **Figure 2-1** conceptualizes some key factors identified to contribute to marking accuracy of school based examinations.

Figure 2-1 Conceptual framework



In order to reduce the error in test score, school administrators must be involved to ensure that all the steps in examinations cycle are adhered to. One way of ensuring this is to have school based examinations policy to guide the whole process from setting, administration, marking and processing of the examination results. They should also ensure that common marking schemes are used and moderated to enhance consistency between teachers marking the same exam scripts. The head of department (HODs) should also advise the administration on ways of improving examinations marking or outsource human resource to train the department on best assessment practices .When marking, teachers should be monitored and should consult incase of mismatch.

As Suto and Nadas (2008) found out, graduates in relevant subjects but with neither teaching nor marking experience were able to mark as accurately as individuals with both teaching and marking experience. They therefore broadly suggested that, when it comes to marking; "... education of an examiner is more important than the experience" (p.10). They however asserted that "...suggesting that a marker's highest level of education in any subject is a better predictor of accuracy than his or her highest level of education in a relevant subject is open to a number of interpretations". To them, the key to successful marking was being able to follow the marking instructions and interpret the instructions in the way its author intended (p.10).

2.9 Summary of the Reviewed Literature

After the review, it was clear that when considering examinations and marking, teachers and administrators need to know that there is always an error in all classroom and standardized assessments. As Brown and Hudson (1998) found out:

“The problem with formative assessments in schools lies with their validity and reliability. The tasks are often not tried out to see whether they produce the desired information. Marking criteria are often not investigated to see whether they work and raters are not often trained to give consistent marks”

Researchers were in agreement that marking scheme improves marking reliability if it is clear and detailed. Mathematics scores awarded by teachers in schools will therefore be unreliable if marking schemes are not prepared and moderated. The researchers however held different views on whether moderation of marking scheme or any intended improvement enhances consistency of scoring.

Several factors were identified that affects consistency of scoring student work. Student factors that affected consistency of scoring included: student handwriting, physical attractiveness, race, and knowledge of student previous performance. It was also clear that when examiners are fatigued the consistency of marking declined. Personal feeling about a particular student and quality of scripts marked before it (contrast effects) also affected the way teachers marked.

In order to reduce subjectivity inherent in the scoring of OE items and improve its reliability, Allalouf, Klapter & Fronton (2008) suggest the engagement of professional raters, using comprehensive rating instructions, training the raters, monitoring the rating process, using retraining when drift is detected, having multiple raters and engaging the services of additional raters incase of discrepancy between the raters

CHAPTER THREE RESEARCH METHODOLOGY

3.1 Introduction

This section explains the methodology used in carrying out the research. It encompasses the research design, target population, sample and sampling procedures, research instruments, data collection and data analysis procedures. Ethical considerations of the study were also discussed.

3.2 Research Design

The research design used in this study was a descriptive survey. Kombo & Tromp (2006) points out that descriptive studies are not only restricted to fact findings, but may often result in the formulation of principal of knowledge and solution to significant problems. They are more than just a collection of data. They involve measurement, classification, analysis, comparison and interpretation of data. The study's aim was to collect information from respondents on consistency of scoring school based examinations when common and different marking schemes are used. The respondents were given copies of the same student mathematics script to score. Markers were divided into two groups one group was provided with marking schemes while the other group was not. They were given enough time to complete the task. The respondents had an opportunity to give some of the reasons why teachers do not use common marking schemes, and suggested ways in which variation of examinations scores could reduce among markers.

3.3 Target Population

There are 6,448 Kenya Certificate of Secondary Education (KSCE) examination centres in Kenya. Out of this 4769 are public secondary schools and 1679 are private secondary schools (KNEC report, 2011). This study was conducted in Makadara Sub-county in Nairobi County, Kenya. The target population for this study was 156 mathematics teachers, 23 deans of studies and deputy principals in case a school did not have a dean of studies

3.4 Sample Size and Sampling Procedures

According to Best and Kahn (2011) “there is no fixed number or percentage of subjects that determines the size of an adequate sample” (p. 21). Rather, an ideal sample is large enough to serve as a representation of the target population and small enough to be selected economically in terms of subject availability and expense in both time and money. The study sample was selected from Secondary schools in Makadara Sub-county. In Kenyan secondary schools, teachers are required to set, mark, grade and prepare marking schemes for school based examinations. In primary schools however, most schools buy exams from test publishers.

Table 3-1 shows categories of the school sample.

Table 3-1 Categories of the school sampled

Category of school	Number of schools	Sample size
Public	11	6
Private	12	6
Total	23	12

Source: Makadara Sub-county education office (2013)

Stratified random sampling technique was used to get the 12 school (six private and six public) out of the 23 schools provided by the Makadara Sub-county education office. From the twelve schools, a sample of 57 teachers, 10 deans of studies and 2 deputy principals was selected to take part in this study. The mathematics script that teachers marked was selected randomly from 37 form four student who had sat for the examination. Ten deans of studies and two deputy Principals also constituted the subjects for this study.

3.5 Data Collection Instruments

The data collection instruments included the mathematics script that respondents marked, semi-structured interviews and survey questionnaires. Interview schedule was used to gather information on school assessment policy, setting and marking examinations. The teachers marked one mathematics paper then filled questionnaires. This helped the researcher to get in-depth information on how teachers mark Mathematics examinations in the sample schools.

As a result, the researcher was able to get complete and detailed understanding of the way the schools marked formative assessment examinations.

3.6 Validity of the instruments

Validity in this context is concerned with the ability of an instrument to test or measure what is intended to measure (Kombo and Tromp). In the process of developing the instruments, the researcher consulted the supervisor who guided and verified that the instruments were appropriate for obtaining the needed information. The questions for the Mathematics script that teachers marked were set by the researcher as a mathematics teacher and were moderated by other mathematics department members. The content of the test was from secondary school mathematics syllabus. The students who did the paper had completed covering the syllabus.

3.7 Data Collection Procedure

The researcher randomly identified the schools where teachers were to be provided with a marking scheme or not. The selected schools were visited and the researcher gave teachers an examination script to mark. After marking, teachers were required to fill the questionnaires and put both the paper and the questionnaires in an envelope. Deputy Principals and deans of studies were interviewed. The respondents were assured of strict confidentiality in dealing with the responses. The questionnaires were collected after one week.

3.8 Data Analysis Procedure

The researcher analysed the data qualitatively and quantitatively. According to Walliman (2009), descriptive statistics provide a method of quantifying the characteristics of the data, where their centre is, how broadly they spread and how one aspect of the data relates to another aspect of the same data. The main purpose of statistical analysis is to examine qualities of a single set of data and also identify and quantify relationships between variables. By searching out patterns, surprising phenomena and inconsistencies, “the researcher can generate new concepts and theory or uncover further instances of those already in existence” p. 308. After all the data was collected, it was coded and entered in the computer for analysis using the statistical package for social sciences (SPSS 20) and Microsoft office excel 2010.

Descriptive statistics such as percentages, frequencies, mean and standard deviation was used to report the data. SPSS was also used to compute internal consistency reliability coefficient (Cronbach alpha).The findings of the study were represented in summary form using frequency distribution tables, pie charts and bar charts.

3.9 Ethical considerations

The respondents were assured that strict confidentiality was maintained in dealing with the responses. They were not required to write their names on the script that they marked and the questionnaires. They were also required to seal both the script and the questionnaire in an envelope that was provided. The researcher assured them that the information given was to be used for the purpose of this research project only.

CHAPTER FOUR
DATA ANALYSIS, PRESENTATION AND INTERPRETATION

4.1 Introduction.

This chapter presents data analysis of the study findings. The purpose of the study was to find out the effects of a marking scheme on consistency of scoring mathematics examination in schools. The findings of the research are presented based on the three research questions. Results and discussions derived from analysis of variables are presented in this chapter.

4.2 Demographic Information of the Respondents.

The study targeted 156 mathematics teachers in Makadara Sub-county and 23 deans of studies. The table below shows the summary of gender of the teachers' respondents.

Table 4-1 Gender of the teachers' respondents

Gender	Frequency	Percent	Valid percent
Male	42	73.7	73.7
Female	15	26.3	26.3
Total	57	100	100

Fifty seven respondents marked and filled questionnaires, three did not return the questionnaires; two males and one female. Ten Deans of studies; four males and six females were also interviewed. Two deputy principals were interviewed in the schools which did not have deans of studies. Majority of mathematics teachers in Makadara Sub-county were male at 73.7 percent compared to only 26.3 percent Females.

The qualifications of the respondents are shown in **table 4-2** below.

Table 4-2 Academic qualifications of teachers' respondents

	Frequency	Percent	Valid Percent
Diploma in Education	3	5.3	5.3
Bachelor of Education (science)	38	66.7	66.7
Bachelor of Education (Arts)	5	8.8	8.8
Master of Education	7	12.3	12.3
Other qualifications	4	7	7
Total	57	100	100

All the respondents had requisite knowledge to teach mathematics. The number of teachers who had bachelor of education (science) degree was 38 (66.7 percent) and 12.3 percent had a master of education degree. Those with other qualifications like Bachelor of Science degree had post graduate diploma in education. One teacher had higher national diploma and was teaching mathematics and physics. Three teachers (5.3 percent) had Bachelor of Arts and had majored in double mathematics. This was a clear indication that teachers in Makadara sub-county were well trained to teach mathematics.

Only 8 out of 57 respondents were trained examiners. However 66.7 percent of the entire respondents had attended a workshop/training in marking examinations. Most of the teachers who had some form of training had more than five year of teaching. Majority of teachers who had no training had served for less than five years. The data shows the need of increasing the number of trained teachers in Makadara Sub-County.

Figure 4-1 below shows the experience in years and frequency of teachers who marked the script.

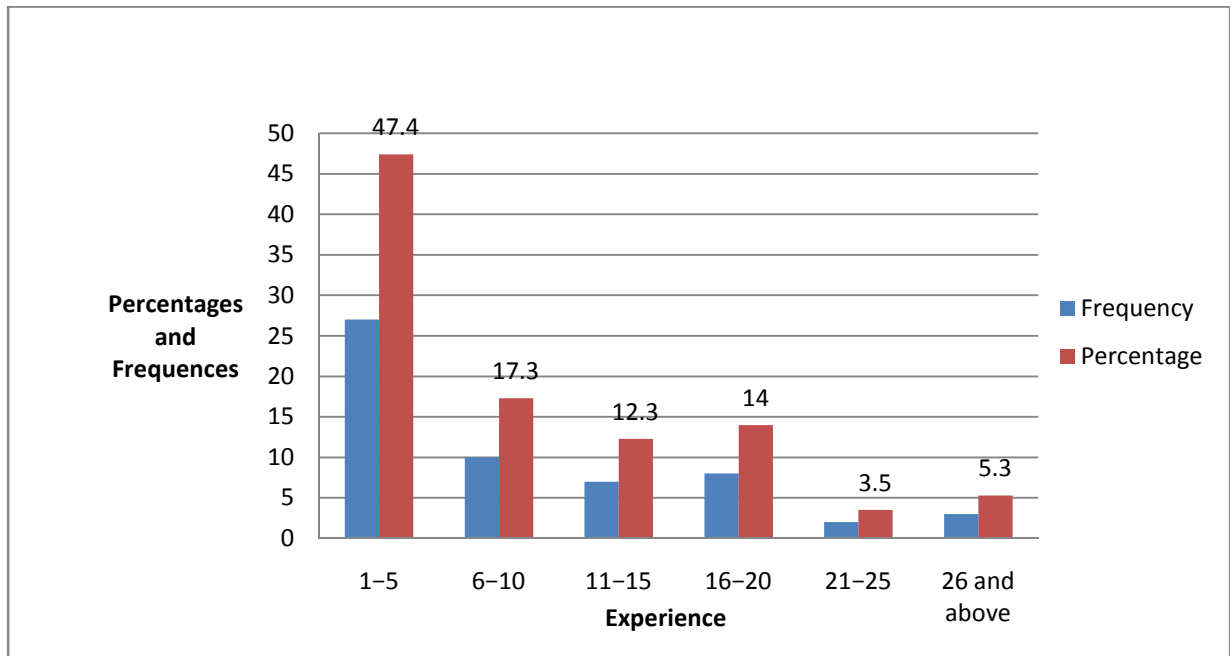


Figure 4-1 Teaching experience bar chart

The percentage of all the respondents who had served for less than five years was 47.4 while 17.5 percent had worked for less than 10 years but more than five years. This is an indication that most mathematics teachers in Makadara Sub-county are young professional mathematics teachers. Only 5.3 percent of teachers had served for more than 26 years.

4.3 Marking Scheme and Consistency of Marking.

The first objective of the study sought to investigate the effects of marking scheme on the consistency of marking school based mathematics examinations. To achieve this objective, teachers were given one mathematics examination paper to mark that was chosen from 37 learners who had sat for that examination. After marking, they were required to fill questionnaires. Twelve schools were sampled; six public and six private. 30 teachers were given marking scheme while 27 were not provided with the marking scheme. The score and the frequency of the score are shown in the table 4-3 below.

Table 4-3 Mathematics Scores marked out of 100

Scores awarded	Frequency of teachers awarding the score and were provided with a Marking scheme	Frequency of teachers awarding the score and were not provided with a marking scheme
38	1	0
40	1	2
42	1	0
44	3	0
45	2	0
46	5	1
47	3	1
48	1	0
49	0	1
50	0	1
52	5	1
54	2	0
55	3	0
56	3	2
57	0	2
59	0	3
60	0	3
62	0	1
64	0	2
65	0	1
68	0	2
69	0	1
71	0	1
81	0	1
83	0	1
Total	30	27

From the table, it can be noted that teachers scored same mathematics paper differently even when a marking scheme was provided. Some questions were worked out using wrong methods and majority of teachers who marked these items right were teachers who did not have marking scheme. Some teachers even those with the marking scheme gave full score to item number six which had two values of x but the student had written only one value. Some also considered accuracy of the answer which others did not especially question number 18. Those who were provided with the marking scheme did not have an opportunity to discuss it with the examiner although the marking scheme had instructions on where to award marks or deny marks. The range of score for teachers who used different marking scheme was 43 while those who used same marking scheme was 18. The test developer who also scored the paper gave a score of 45. The mean of those who scored the paper using marking scheme was 48.88 which was almost within ± 3 range the score given by the test developer. This was an indication that teachers who used the marking scheme marked more consistently than those who did not use marking scheme. The mean, standard deviation and standard error mean are given by the group statistics table 4-4.

Table 4-4 Group Statistics for the scores awarded

Condition	N	Mean	Std. Deviation	Std. Error
				Mean
Without marking scheme	27	59.33	10.473	2.016
With marking scheme	30	48.73	5.132	0.937

The mean and standard deviation of the teachers who used a common marking scheme was lower than the mean and standard deviation of those who used a different marking scheme. This was a good indication that teachers who used a common marking scheme marked more consistently.

The table below presents internal consistency reliability coefficient (alpha).

Table 4-5 Internal consistency reliability coefficients

	Cronbach's Alpha	Cronbach's Alpha based on standardized items	No. of items
Without Marking Scheme	.609	.659	21
With Marking Scheme	.782	.781	21

***Note that a reliability coefficient of .70 or higher is considered “acceptable”.**

Cronbach alpha a coefficient for internal consistency was higher when a marking scheme was used than when marking scheme was not used. This implied that when common marking schemes were used, marking consistency was enhanced.

The research findings showed that in both cases teachers marked same scripts differently just like Sandler (2009) noted. Any discussion from the findings has to start with one asking why the respondents scored the script differently even those who were provided with the marking scheme. The respondents may be similar to those described by Sandler. According to Sandler, some markers are characteristically generous, some strict and others may be inconsistent. The range of scores and standard deviation was good evidence that some respondents marked the paper leniently while others were strict. The study findings concurs Wang (2009) findings who cited setting standards for marking and passing them to examiners who may mark centrally or individually in their homes as one way of improving consistency of marking. Respondents were also required to give their opinion on whether a well prepared marking scheme improves the consistency of scoring mathematics examination. Fifty five respondents (96.5 percent) agreed while only 2 respondents (3.5 percent) did not believe that a well prepared marking scheme can improve consistency of marking.

From the findings of this study it is evident that marking scheme is necessary if consistency of marking is to be enhanced. The researcher therefore sought to find whether teachers

prepared marking schemes and when they prepared them. After analysis of items in part 2 of teachers' questionnaires the results showed that majority of teachers who marked the paper taught two subjects and most schools where the study was conducted had more than two streams. It was also clear that most teachers shared classes where mathematics in one form was taught by more than one teacher. When teachers were asked whether they prepare marking scheme for all the mathematics examinations they set, the results are as shown in **figure 4-2**.

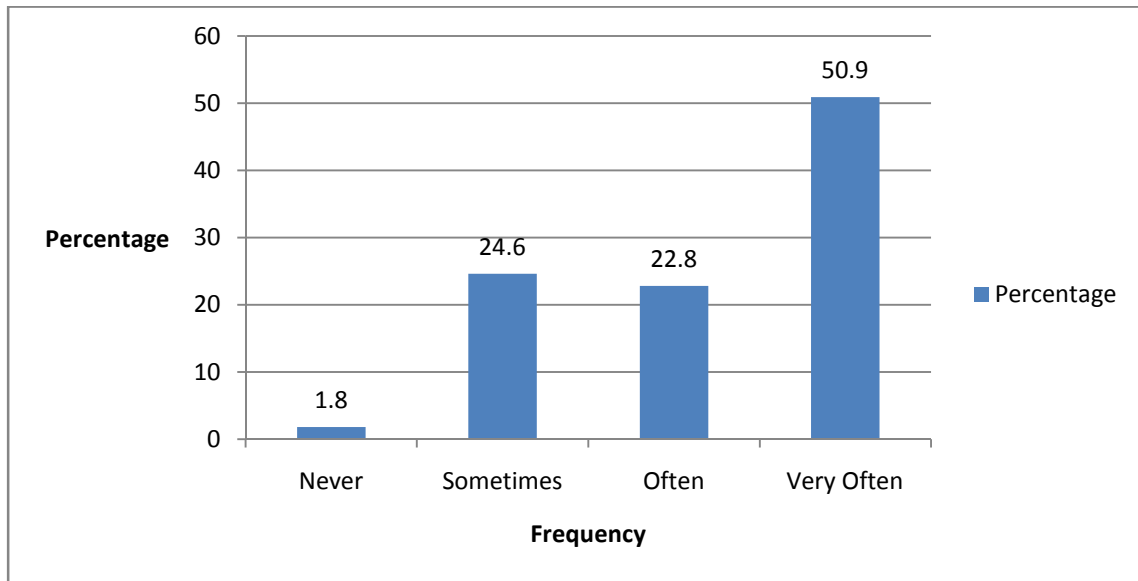


Figure 4-2 Marking schemes preparation

The figure above shows that slightly above 50 percent of teachers prepared marking scheme while 1.8 percent of the teachers did not prepare marking scheme at all. The rest prepared marking schemes but not all the time. When asked to explain why this was the case some cited high work load, lack of time while some said that there was no policy requiring them to prepare marking scheme. Those who prepare marking scheme were asked the time they prepare them. The results are shown in the **table 4-6** below.

Table 4-6 Time of preparation of marking scheme

Time	Frequency	Percent	Valid Percent
When setting exams.	28	49.1	49.1
After the exam is done before marking.	29	50.9	50.9
Total	57	100	100

The norm is that marking scheme should be prepared at the initial stage of test development. The data in table 4.6 indicates that more teachers prepared marking schemes after the exam was done but before marking. It was only 49.1 percent who prepared marking scheme when setting mathematics examinations. As seen earlier about 50 percent of the respondents did not prepare marking scheme all the time and when they prepared, it was after exams are done. This can explain why some questions are allocated more marks or fewer marks than expected when teachers set mathematics examinations. This however can be improved if marking schemes are prepared when examinations are set. No respondent in this study was preparing a marking scheme after marking some scripts. The findings reveal that most teachers in Makadara Sub-county lacked proper training on how to set and mark examinations. This was despite majority of them being professionally trained. The findings were in agreement with Stiggins (2000) who noted in his paper the little emphasis on assessment in the preparation of or professional development of teachers and administrators.

4.3.1 Marking Scheme properties.

Teachers were required to evaluate marking schemes prepared in their departments. The levels of respondents' agreement on the statements about the marking schemes prepared in their departments are captured in **table 4.7** below.

Table 4-7 Evaluation of marking schemes properties

Marking scheme properties	SA	A	U	D	SD	TOTAL %
Marking schemes prepared include general instructions on marking.	24.6	43.9	21.1	10.5	0	100
Are clear and designed so that they can be easily and consistently applied.	28.1	47.4	10.5	14	0	100
Allocate marks commensurate with the demand of question/tasks.	42.1	42.1	12.3	3.5	0	100
Include the marks allocated for each question with a more detailed breakdown where necessary.	26.3	40.4	17.5	14	1.8	100
State the acceptable responses to each question, with details that allow marking in a standardized manner.	24.6	35.1	22.8	17.5	0	100
Two independent markers will arrive at the same score for a given response given the marking scheme.	15.8	42.1	15.8	17.5	8.8	100
Teachers always discuss any difference in interpretation after moderation and make adjustments accordingly.	26.3	45.6	8.8	15.8	3.5	100
Teachers are unable to prepare marking criteria for some questions?	3.5	29.8	19.3	31.6	15.8	100

A bigger percentage of teachers agreed that marking schemes prepared in their departments included general instructions on marking however 10.5 percent disagreed. The marking schemes were also clear and allocated marks commensurate with the demand of question/tasks. Fifty nine point seven percent agreed that marking schemes prepared in their department allowed marking in a standardized manner, 22.8 percent were uncertain while 17.5 percent disagreed on the same. The percentage that agreed that teachers always meet to discuss any difference in interpretation after moderation was 71.9 while 19.3 disagreed. This was confirmed by deans of studies that teachers do not always meet to discuss marking schemes. Only 47.7 percent were certain that teachers are able to prepare marking criteria for all questions they set. The remaining percentage which was huge agreed that teachers

sometimes set questions they cannot prepare marking criteria. This however could have been minimized if marking schemes were prepared at the initial stage of test development. The findings concur with Stiggins (1999) who questioned the ability of teachers assessing content that they had not mastered.

4.3.2 Marking scheme preparation and usage

The respondents were also required to give reasons why teachers sometimes do not prepare marking scheme. The reasons they gave were varied and included lack of time, commitment, coordination by Head of departments (H.O.Ds) and team work. Some also mentioned heavy work load while others cited over confidence. Those who were over confident assumed they had mastered content being tested due to vast accumulated work experience. This made them believe that they can make marking scheme as they mark instead of wasting time making it. Due to different teaching styles, student worked different questions using various methods. Teachers therefore felt that there was no need of a common marking scheme. The chart below shows the percentages of various reasons given.

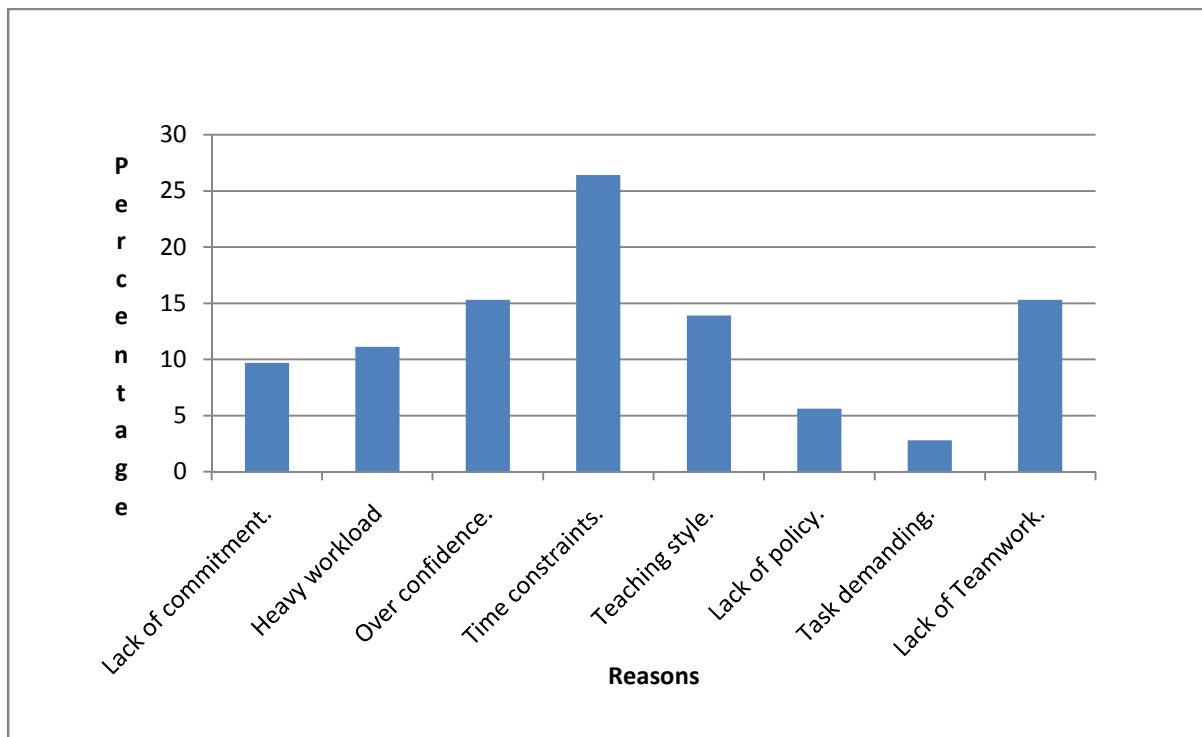


Figure 4-3 Reasons why teachers do not prepare marking schemes

From the bar chart, it can be noted that most teachers did not prepare marking scheme because they had limited time. This could have been as a result of heavy workload and teaching two subjects. Teachers also lacked common time to meet, set, prepare and mark examinations. They also mentioned that deadlines for setting exams were too short to wait for a common marking scheme to be prepared and moderated.

The reason why teachers do not use a common marking scheme was blamed first on a situation where a teacher set an examination and failed to prepare a marking scheme. Individual teachers sharing a class are therefore forced to prepare their own to meet set deadlines. The other possibility was where a teacher was provided with marking scheme and failed to use it because of teaching style. Some teachers also cited that their colleagues felt superior and therefore were not taking any advice from the perceived inferior. Final reason was because copies of the common marking scheme were few or were not made in time. In some schools, there was lack of coordination and a policy that teachers should use common marking schemes.

4.4 Moderation and consistency of marking.

The second research objective sought to establish if moderation of marking scheme improves the consistency of scoring mathematics examinations. To achieve this objective the respondents were required to answer three questions. The first item sought to know if teachers meet to discuss marking scheme prepared. The second item asked if moderation of marking scheme improves consistency of marking. Finally the respondents were required to give their opinion on ways in which consistency of scoring should be improved.

The data in the next **table 4-8** shows the frequency and percentage of those who meet or do not meet to discuss the marking scheme.

Table 4-8 Marking scheme moderation

	Frequency	Percent	Valid percent
Meet to discuss	39	68.4	68.4
Do not meet to discuss	18	31.6	31.6
Total	57	100	100

It was evident that most mathematics teachers in Makadara Sub-county meet to discuss the marking scheme. This was a good way of ensuring consistency of marking and improving the validity of scores awarded. However, the number of teachers sharing classes but did not meet to discuss (moderate) marking scheme was still high at 31.6 percent. It was also clear that teachers set questions that are not clear or sometimes not workable. This happened especially when teachers copied and pasted question from external examinations. The good thing however was that 93 percent of those sampled said they meet to discuss such questions. Only Seven percent did not meet to discuss such questions. This means that such questions were marked using different marking schemes equivalent to the number of scorers. The finding that not all teachers meet to discuss the marking schemes is in line with the earlier findings by Brown and Hudson (1998) that marking criteria in formative assessments are often not investigated to see whether they work and that raters are not often trained to give consistent marks. The marks awarded by mathematics teachers in makadara Sub-county were therefore questionable.

Table 4-9 shows respondents views on how consistency of marking among teachers marking same exams scripts can be improved.

Table 4-9 Ways of improving consistency of marking

Teachers opinion	SA	A	U	D	SD
Use of a common marking scheme.	77.2	17.5	5.3	0	0
Moderation of the marking scheme.	71.9	26.3	1.8	0	0
Discussing any difference in interpretation of a marking scheme.	61.4	36.8	1.8	0	0
Setting specific standards before real scoring.	28.1	45.6	14	8.8	3.5
Sampling by subject by subject heads.	21.1	36.8	21.1	12.3	8.8
Training of teachers as examiners.	52.6	43.9	0	1.8	1.8

From the table, it was clear that a large percentage of teachers believed that use of moderated common marking scheme improves consistency of marking. Only 1.8 percent of all the respondents were not sure if moderation actually improves the consistency of marking. Nearly all participants agreed that if teachers were to meet and any difference in

interpretation of the marking scheme discussed, consistency was to be greatly enhanced. These study findings agrees with the research by Greatorex, Baird, and Bell in 2002 that noted that examiners think that the standardization meetings are reliable because they help them understand the marking scheme and makes the principal examiner's interpretation of the marking scheme clear. The findings however contradict Meadows and Billington (2005) observation that the intended improvement to marking scheme do not always bring about expected improvement in reliability.

The table above also shows mixed views on whether sampling by subject heads improves the consistency of scoring. A huge percentage of 96.5 percent agreed that training teachers as examiners can improve consistency of marking mathematics exams in school. However as noted earlier only 14 percent of the respondents were trained KNEC examiners and 66.7 percent had attended a workshop on marking. Most of these trained examiners were from public schools.

4.4.1 Reducing scores variation among teacher marking same scripts.

Other than moderation and issues discussed above the respondents also gave their own views on how consistency of scoring mathematics in schools can be improved.

Figure 4-4 presents respondents' views on how score variation among teachers marking same scripts can be reduced.

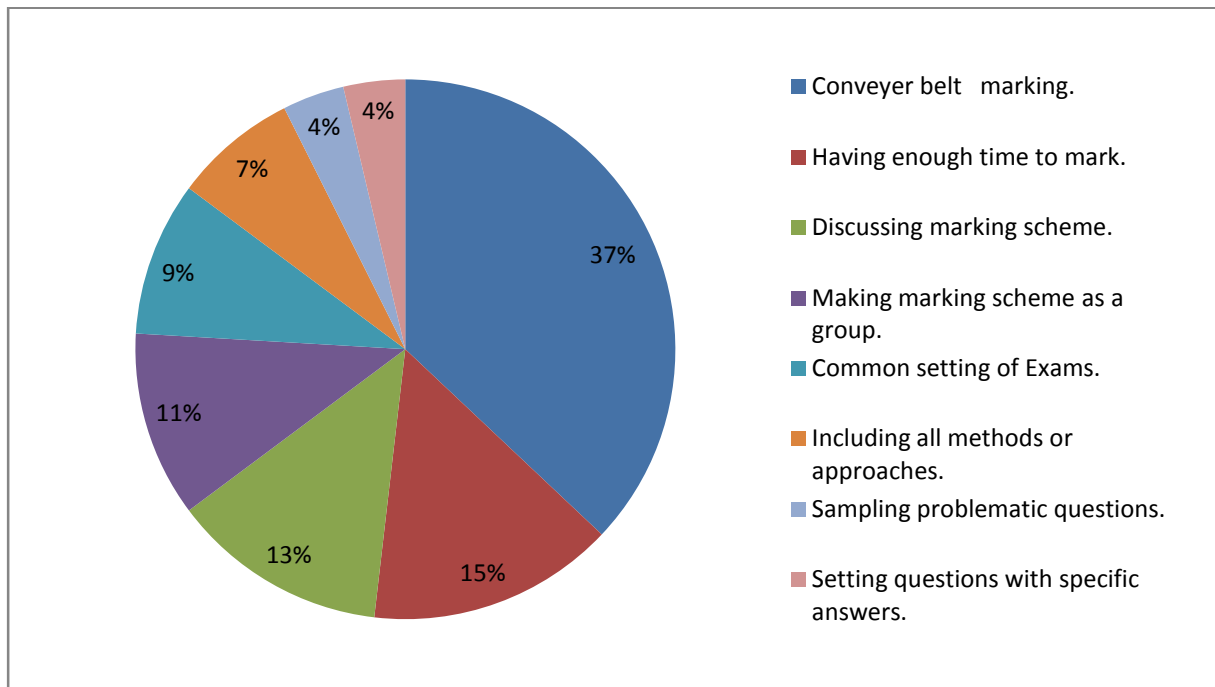


Figure 4-4 Ways of reducing scores variation

Conveyer belt marking system (CBS) was viewed by 37 percent of the respondents as one way in which variation can be reduced. This is in agreement with Bukenya (2006) suggestions that CBS can be used to reduce inconsistency of scoring. Fifteen percent however complained that there was limited time available for team marking. They felt that deadlines set were not realistic and therefore needed more time to mark consistently and as a team. About 33 percent of the respondents were of the view that setting examinations together, making marking schemes as a group during setting and discussing the marking scheme were other ways in which score variation can be reduced among teachers. Four percent suggested that sampling problematic questions and discussing how such questions should be marked can greatly enhance the consistency of marking. Some were also of the view that marking schemes should include all alternative methods/approaches of solving a certain question.

4.5 Students' Factors affecting consistency of scoring

The third and the final research objective sought to determine students' factors that may have any effects on consistency of mathematics examinations. The results of the findings are summarized in the **table 4-10** below.

Table 4-10 Factors affecting consistency of scoring

Teacher's opinion.	SA	A	U	D	SD	Total %
Organization of the answers on the scripts.	28.1	52.6	5.3	10.5	3.5	100
Handwriting of the students.	14	47.4	21.1	17.5	0	100
Personal feelings about particular student.	10.5	33.3	19.4	26.3	10.5	100
Decrease in markers accuracy over time due to fatigue and lack of attention.	14	50.9	17.5	15.8	1.8	100
Quality of scripts marked before it.	8.8	36.8	12.3	31.6	10.5	100

Over 70 percent of the respondents agreed that the way learners present their work and handwriting can cause score variation among teachers marking same script. Some respondents complained that poorly organized working plus limited time available may affect how one marks. This concurs with Henry (2012) observation that neatness of the handwriting and presentation bring students extra marks. The finding however contradicts Massey (1983) research findings that indicated that experienced examiners were not susceptible to biasing effects of handwriting style and presentation. There were mixed views on whether personal feelings about a particular learner and quality of scripts marked earlier can cause score variation. However, 64.9 percent agreed that lack of attention and marking accuracy decrease over time. Only a small percentage (1.8 percent) disagreed with this view.

The respondents were also asked to give their opinions on students' characteristics that can influence a markers judgment. The percentage results are shown in **table 4-11** below.

Table 4-11 Student’s characteristics that can influence scoring

Characteristics.	SA	A	U	D	SD	Total %
Gender.	3.5	8.8	10.5	36.8	40.4	100
Race.	1.8	5.3	8.6	40.4	43.9	100
Social class.	1.8	35	10.5	40.4	43.8	100
Physical attractiveness.	1.8	1.8	24.6	29.8	42	100
Knowledge of the student and previous performance.	26.3	45.6	14	10.5	3.6	100

Unlike in other countries, majority of the respondents in Makadara Sub-county, Kenya believed that gender, race and social class did not affect their judgments when marking. According to Wade (1978) candidate’s characteristics like gender, race, social class and physical attractiveness influence examiner’s judgments. In this study 24.6 percent were undecided whether physical attractiveness affects judgments but 71.8 percent disagreed. Majority (71.9 percent) agreed that knowledge of the student and previous performance affects how one marks. This supports Aslett (2006) finding that if a student who normally gets high marks submits poor assignments, allowances may be made to halo effect.

CHAPTER FIVE

SUMMARY, CONCLUSIONS AND RECOMMEDATIONS

5.1 Introduction.

The chapter presents the summary of the study findings, conclusion and recommendations arrived at. The recommendations are for policy and practice purpose as well as suggestions for related studies that could be carried out in future.

5.2 Summary of the study findings.

The purpose of the study was to investigate effects of marking scheme on the consistency of scoring mathematics examinations. Primary data for the research was collected from twelve schools where respondents were required to mark one examination script picked randomly from a mathematics examination administered to 37 students. Two deputy principals and 10 deans of studies were interviewed. Fifty seven mathematics teachers marked the script and then filled the questionnaires. These data was analyzed based on research questions using Microsoft office excel 2010 and SPSS 20.

The first research question sought to find what effects a marking scheme has on consistency of scoring mathematics examination. Would teachers using same marking scheme mark consistently the same way with teachers using different marking schemes? To answer the question teachers were divided into two groups. One group was provided with the marking scheme while the other was not. After analysis of the scores awarded by the two groups, it was noted that the standard deviation was lower at 5.132 for the group that was provided with marking schemes compared to 10.473 for the group that prepared their own marking schemes. Cronbach alpha a coefficient for internal consistency for the group using a common marking scheme was .782 while the other group that was not provided with marking schemes was .609. A reliability coefficient of .70 or higher is considered acceptable (Sax, 1997). This result was an indication that a marking scheme has effect on consistency of scoring mathematics examinations. If a common marking scheme is used, consistency of scoring mathematics examinations will be greatly enhanced.

With such findings the researcher was also interested at finding out whether mathematics teachers in Makadara Sub-county prepared marking schemes or not. The results showed that 72.7 percent prepared marking schemes while the rest prepared marking scheme occasionally. The study revealed that slightly above 50 of the respondents prepared marking schemes after examinations were done. This was despite majority of them being professionally trained. When asked why teachers sometimes do not prepare marking scheme, varied reasons were given like lack of time, Commitment, coordination by Head of departments (H.O.Ds) and team work. Some also mentioned heavy work load while other cited over confidence. Those who were over confident assumed they had mastered content being tested due to vast accumulated work experience

Teachers, deans of studies and deputy principals responses showed that majority of schools actually did not have policy on examination marking but expected teachers to do the right thing like preparing marking schemes when developing the test. The number of trained KNEC examiners in the Sub-county was relatively low at 14 percent. The other percentages of teachers had only attended workshops in marking examinations which were rarely organized.

The second research question sought to answer whether moderation of a marking scheme improves consistency of scoring mathematics examinations. When the respondents were asked if moderation of the marking scheme could improve the way they marked, 98.2 percent agreed. It was only 1.8 percent of all the respondents who were undecided. This study found that marking schemes prepared in Makadara Sub-county were rarely moderated and not all teachers managed to use the common marking scheme as few copies were made. Teachers were however meeting to discuss problematic questions. The respondents felt that discussing marking scheme, conveyer belt marking system, having enough time to mark, making marking scheme as a group and common setting of exams are some of the ways that can help in enhancing consistency of marking.

The third and the final research question sought to find students' factors that affect consistency of scoring mathematics examinations. Higher percentage of the respondents agreed that organization of the answers on the scripts, handwriting of the students, knowledge of student and previous performance affected the way teachers in Makadara Sub-county mark mathematics examinations in schools. There were mixed views on whether personal feelings about a particular learner and quality of scripts marked before could affect the way they marked. Gender, race, social class and physical attractiveness did not affect the way Makadara Sub-County teachers marked mathematics examinations. Other than the students factors that seemed to affect the way teachers mark, the respondents agreed that decrease in markers accuracy over time due to fatigue and lack of attention also affected the way teachers marked mathematics examinations.

5.3 Conclusion

The study concluded that marking scheme had effects on how teachers score mathematics examinations. A well prepared marking scheme enhances consistency of scoring especially if it is prepared and moderated by all teachers marking the examination. The consistency could be further enhanced if a marking scheme is piloted and any difference in interpretation is discussed and adjusted accordingly. However from the way the teachers marked the script the researcher concluded that score awarded by teacher 'A' may not be the same as the score awarded by teacher 'B' for the same script. The score variation among teacher can be reduced if common moderated marking schemes are used. The implications to such conclusion result to students classifying teachers based on how they mark. The students may also be affected on how they select elective subjects as the researcher had noted. A teacher who marks leniently may have the entire class selecting his/her subject .On the other hand, students may fail to select subjects for teachers considered as hard graders. Such a problem can be minimized if teachers sharing classes moderate marking schemes.

The study also concluded that teachers in Makadara Sub-County lacked proper training on how to set and mark examinations. Most schools did not have a policy on examination marking but expected teachers to do the right thing like preparing marking scheme when developing a test.

5.4 Recommendations for policy.

For teachers to improve how they mark mathematics examinations in schools the study recommends use of common marking schemes to mark same script. These marking schemes must be moderated. The preparation of marking schemes should be done when developing the test and it is the responsibility of the H.O.Ds and administration to ensure this is done. Enough copies of moderated marking schemes should be made available to all teachers marking common exams and on time. Although teachers complained of time constraints, the study recommends team marking for common exams especially the use of CBS.

Ministry of Education in collaboration with KNEC should ensure that the school administrators and teachers are trained on essential assessment concepts, principles, techniques and procedures. This can be done through in-service training and workshops. School administrators should support teachers to train as examiners as this could improve the way teachers mark. They should also encourage team work among teachers and should set reasonable deadlines that may help teachers mark consistency. In case of fatigue, teachers should have regular breaks. The administrators should also monitor the whole assessment process from setting, administration and marking examinations.

5.5 Recommendations for further research

Based on the findings of the study, the researcher makes the following suggestions for the research: The study should be replicated in other sub-counties in Kenya to compare the effects of marking schemes on the consistency of scoring mathematics examinations.

Similarly there is need to compare the effects of marking schemes on the consistency of scoring mathematics examinations and other examinable subjects examinations. A study on the effects of feedback to examiners on reliability of scoring examinations is also necessary. Finally, future researchers should consider conducting a study on reliability of formative assessment in schools.

REFERENCES

- Alias (2005) *Assessment of learning outcomes; validity and reliability of classroom tests*, world Transactions on Engineering and technology Education vol 4(2)2005. pp 235-238
- Allalouf, A. Klapper, G. & Fronton, M (2008). *Comparing vertical and horizontal scoring of open – ended Questionnaires*. Practical Assessment Research & Evaluation, 13(8) <http://pareonline.net/getvn.asp?v=13&n=8>.
- Aslett, J. (2006). *Reducing variability, increasing reliability: exploring the psychology of intra- and inter-rater reliability*. Investigations in university teaching and learning. Vol.4(1) autumn 2006.
- ALTC. (2012a). *Assessment moderation toolkit*, Retrieved from <http://resource.unisa.edu.au/course/view.php?id=285&topic=1>
- ALTC. (2012c). *Moderation checklist*, Retrieved from <http://resource.unisa.edu.au/course/view.php?id=285&topic=1>.
- Baird, J. & Meadows, M. (2009). *What is the right mark? Respecting marker's view in a community of practice*. Paper presented at the annual conference of the international association for educational assessment, Australia.
- Baird, J. (1999). *Blind marking in GCSE and A-level examinations*. AEB Research Report, RC/10.
- Baird, J., Greatorex, J. & Bell, J. F. (2003). *What makes marking reliable? Experiments with UK examinations*. AQA Research Report, RC217.
- Baird, J. & Bridle, N. (2000). *A feasibility study on anonymised marking in large-scale public examinations*. AQA Research Report, RC/91.
- Bird, F., & Yucel, R. (2010). *Building sustainable expertise in marking: integrating the moderation of first year assessment*. Paper presented at the ATN Assessment Conference University of Technology Sydney. <http://www.iml.uts.edu.au/atnassessment/poster.html>
- Bloxham, S., Boyd, P., & Orr, S. (2011). *Mark my words: the role of assessment criteria in UK higher education grading practices*. Studies in Higher Education, 36(6), 655-670. doi:10.1080/03075071003777716

- Bond, L. (2009). *The case for common Examination*. The carriage Foundation for the Advancement of teaching.
- Brimi, M. (2011). *Reliability of grading High school work in English*. Practical Assessment Research & Evaluation, 16(17)
- Brown, J.D. & Hudson. (1998). *The alternatives in language assessment*. TOESOL Quarterly. 32.653 – 675
- Bukenya, M. (2006) *Comparing the reliability of the conveyor belt marking system with the traditional marking system*. Uganda National Examinations Board.
- Chase, C. I. (1983) Essay test scores and reading difficulty. *Journal of Educational Measurement*, v20 n3 p293-297.
- Claupham, C. (2000). *Assessment and testing*. Annual review of applied linguistics, Cambridge university press 20, 147 – 161.
- Delap, M. R. (1993a) *Marking reliability study in Business Studies (665)* AEB Research Report RAC/609.
- Delap, M. R. (1993b) *Marking reliability study in GCSE Geography (1163)* AEB Research Report RAC/610.
- Felton J. & Peter T. Koper T. P. (2005). *Nominal GPA and Real GPA: A Simple Adjustment that Compensates for Grade Inflation*. Assessment & Evaluation in Higher Education, 30(6), 561-569
- Greatorex, J., Baird, J., & Bell, J.F. (2002) *'Tools for the trade': What makes GCSE marking reliable?* Paper presented at the EARLI Special Interest Group on Assessment & Evaluation, University of Northumbria, UK, August 2002.
- Hart, J. (2010). *The effect of handwriting on teachers grading of high school essays*.
- Harvey, C. (2012). *Principals should Evaluate Teachers' Assessment*.
- Henry, J. (2013). *Teacher 'bias' give better marks to favourite pupils*.

- Hillocks, G. (2006). Two Decades of Research on Teaching Writing in the Secondary Schools in the US. *LI-Educational Studies in Language and Literature*, 6 (2), 29-51.
- Hughes. (2011). Towards a personal best: A case for introducing ipsative assessment in higher education. *Studies in higher education*, 36(3), 353-367.
- Hughes, D. C., Keeling, B., & Tuck, B.F. (1980a) Essay marking and the context problem. *Educational Research*, v22 n2 p147-148.
- Hughes, D. C., Keeling, B., & Tuck, B.F. (1980b). The influence of context position and scoring method on essay scoring. *Journal of Educational Measurement*, v17 p131-135.
- Hughes, D. C., Keeling, B., & Tuck, B.F. (1983). *The effects of instructions to scorers intended to reduce context effects in essay scoring*. *Educational & Psychological Measurement*, v34.
- Hughes, D. C. & Keeling, B. (1984). The use of model essays to reduce context effects in essay scoring. *Journal of Educational Measurement*, v21 p277-281.
- Hunter, K. & Docherty, P. (2011). *Reducing Variation in the assessment of Student writing*. *Assessment & Evaluation in Higher education*, 26(10), 109-124. doi:10.1080/1026029309032158442
- James, R. (2003). *Academic standards and assessment of student learning: Some current issues in Australian higher education*. *Tertiary education and management*, 9(3), 187- 198. doi:10.1080/13583883.2003.9967103
- Johnson, V. (2003). *Grade inflation: A crisis in college education*. New York: Springer-Verlag.
- Kahn, E. A. (2000). A Case Study of Assessment in a Grade 10 English Course. *The Journal of Educational Research*, 93(5), 276.
- Kombo, D.K. & Tromp, D.A. (2006). *Proposal and Thesis Writing: An Introduction*. Paulines Publications' Africa, Nairobi.
- Kuzick, S, Groves, R., O'Hare, S., & Pelliccione, L. (2010). *Building team capacity: Sustaining quality in assessment and moderation practices in a fully online unit*. Paper presented at the ATN assessment conference, University of Technology Sydney

- Malkiel, N. (2009) *Grading Policy Letter*, The Daily Princetonian, Retrieved may 17, 2013 from: <http://www.dailyprincetonian.com/2009/04/23/23516/>
- Massey, A. (1983).The effects of handwriting and other incidental variables on GCE 'A' level marks in English Literature. *Educational Review*, v35 n1 p45-50.
- McCormack, C.,& Jones,D.(1998) *Building a web-based education system*. John Wiley & sons, Inc.
- McMillan,J.H.(2000).*Fundamental assessment principles for teachers and school administrators*.Practical Assessment Research & Evaluation.
- Meadows, M. & Billington, L.(2013).*The effect of markers background and training on the quality of marking in GCSE English*. Assessment and Qualifications Alliance, UK; University of Bristol, UK.
- Meadows, M. & Billington, L. (2005).*A review of the literature on marking reliability*. National Assessment Agency.4-5.
- Meyer, J.P. (2010).*Reliability*. Oxford University press Inc, 198 Madison Avenue,NewYork, New York 1006
- Meyer, L. (2000a).*The ones that got away - development of a safety net to catch lingering doubt examiners*. AQA Research Report, RC50.
- Meyer, L. (2000b) *Lingering doubt examiners: results of pilot modelling analyses, summer 2000*: AEB Research Report.
- Millet, I. (2010). *Improving Grading Consistency through Grade Lift Reporting*. Practical Assessment, Research & Evaluation, 15(4). Available online: <http://pareonline.net/getvn.asp?v=15&n=4>.
- Moskal, B. M. (2000) Scoring Rubrics: What, When and How? *Practical Assessment, Research & Evaluation*, v7 n3 p1-7.
- Powers, D., Fowles, M., Farnum, M. & Ramsey, P. (1994).*Will they think less of my handwritten essay if others word process theirs? Effects on essay scores of intermingling handwritten and word-processed essays*. Journal of educational measurement, 31(3), 220-233.

- Powell-Howard, J. (2009). What affects the reliability of vocational examiners' marking?
- Price, M. & Rust, C. (1999). The experience of introducing a common criteria assessment grid across an academic department. *Quality in Higher Education*, v5 p133-144.
- Rojstaczer, S. (2009), *Grade inflation gone wild*, Christian Science Monitor, March 24, 2009, Retrieved May, 2013 from: www.csmonitor.com/2009/0324/p09s02-coop.html
- Rone-Adams, S. & Naylor, S. (2009). *Examination of the inter-rater Agreement among faculty marking a Research proposal on an undergraduate health course*. The internet journal of allied Health course. The internet journal of allied health sciences and practice. vol.7 No.4.
- Royal-Dawson, L. & Baird, J. (2009). *Is teaching experience necessary for reliable scoring of extended English questions?* Educational Measurement: issues and practice 28:2-8
- Russell, M. & Tao, W. (2004). *The influence of Computer-print on rater scores*. Practical Assessment, Research & Evaluation, 9(10). Retrieved may 5, 2013 from <http://PAREonline.net/get/getvn.asp?v=9&n=10>
- Sandler, D.R. (2009). *Grade integrity and the representation of academic achievement*. Studies in higher education, 34,807-826.
- Sax, G. (1997). *Principles of educational and psychological measurement and evaluation*. Gilbert sax-4th edition.
- Sessional paper no.14. (2012). *A policy Framework for Education and training*. Reforming education and training in Kenya.
- Spear, M. (1996). The influence of halo effects upon teachers' assessments of written work. *Research in Education*, v56 p85-87
- Spear, M. (1997). The influence of contrast effects upon teachers' marks. *Educational Research*, v39 n2 p229-233.
- Starch, D. & Elliot, E. (1912). *Reliability in grading High school work in English*. School review. 20, 442-457.
- Stiggins, R. (1986). Inside high school grading practices. *The Northwest Regional Educational Laboratory Report*. Portland, OR: Northwest Regional Laboratory.

- Stiggins, R. (1999). Barriers to effective student assessments. *The Education Digest*, 64(6), 259.
- Stiggins, R. J. (2000). *Classroom assessment: A history of neglect, a future of immense potential*. Paper presented at annual meeting of the American Education research Association.
- Suto, I. & Nadas, R. (2008). *What determines GCSE accuracy? An exploration of expertise among Maths and Physics teachers*. *Research papers in Education* 23(4):477-497
- Suto, I., Nadas, R. & Bell, J. (2009). *Who should mark what? A study of factors affecting marking accuracy in biology examination*. *Research paper in education* 1-31. University of Tasmania, UTAS, (2011). *Guidelines for good assessment practice*
- Van der schAAF, M., Baartman, L. & Prins, F, (2011). *Exploring the role of assessment criteria during teachers' collaborative judgement process of students' portfolios*. *Assessment & Evaluation in higher education*, 1-14. doi:10.1080102602938.2011.576312
- Vaughan, C. (1991) *Holistic assessment: What goes on in the rater's mind?* In L. H.-. Lyons (Ed.), *Assessing Second Language Writing in Academic Contexts*. Norwood, N.J.: Ablex Publishing Corporation.
- Wade, B. (1978) Responses to written work. *Educational Review*, v30 p149-158
- Wang, P. (2009). The inter-rater reliability in scoring composition. *English language teaching*, vol.2, No.3. <http://www.ccsenet.org/journal.html>.
- William, D. (2000) Reliability, validity, and all that jazz. *Education review*, v29 n3 p9-13
- Walliman, N. (2009). *Your research project 2nd edition*, Sage Publications, 1 Oliver's Yard, 55 City Road, London EC1Y 1Sp.

APPENDICES

Appendix A: Questionnaire for Teachers

This questionnaire aims at getting your opinion pertaining the marking, preparation and moderation of marking scheme. The information you give is for research purpose only and will be treated with confidentiality.

Part 1: General Information.

The response to the following items should be indicated by ticking (✓) the appropriate option or by filling in the missing information.

- a. Gender: Male [] Female []
- b. What is your academic qualification?
- | | |
|-----------------------|--------|
| S1 | [] |
| Diploma in Education | [] |
| Bachelor of Education | [] |
| Bachelor of Arts | [] |
| Master of Education | [] |
| Other(s) specify | |
- c. Are you a Kenya national examination council trained examiner? Yes [] No []
- d. Have you attended any workshop / in-service training in making examination?
Yes [] No []
- e. How long have you been teaching?
- | | |
|--------------------|--------|
| 1-5 | [] |
| 6-10 | [] |
| 11-15 | [] |
| 16-20 | [] |
| 21-25 | [] |
| 26 years and above | [] |

Part 2: Marking Scheme and consistency of marking.

Question one.

- a. What is your subject combination, A B
- b. How many streams are there in your school?
Form 1..... Form 2..... Form 3 Form 4.....
- c. Do you share a class with another teacher(s)? i.e. say you teaching Form 1R and another
teacher teach Form 1W
Subject A Yes [] No []
Subject B Yes [] No []
- d. Do you prepare marking scheme for all the exams you set?
Never []
Sometimes []
Often []
Very often []
- e. If you normally prepare marking scheme, when do you prepare them?
When setting exams []
After the exam is done before marking []
After marking some scripts []
I do not prepare at all []
- f. If you share a class with a colleague, do you meet to discuss the marking scheme?
Yes [] No []
- g. If when marking you realize that an answer to a certain question is wrong, do you meet your colleague to discuss such answer? Yes [] No []

Question two.

Indicate by ticking (√) in the appropriate box to show the extent of agreement using the words

Strongly agreed SA

Agree A

Uncertain U

Disagree D

Strongly disagreed SD

Do marking scheme(s) prepared in your department;

	Teachers opinion	SA	A	U	D	SD
1.	Include general instructions on marking					
2.	Are clear and designed so that they can be easily and consist applied					
3.	Allocate marks commensurate with the demand of question / tasks					
4.	Include the marks allocated for each question with a more detailed breakdown where necessary.					
5.	State the acceptable responses to each question, with details that allow marking in a standardized manner.					
6.	Two independent markers will arrive at the same score for a given response given the marking scheme.					
7.	Teachers always discuss any difference in interpretation after moderation and make adjustments accordingly					
8.	Unable to prepare marking criteria for some questions					

Question three.

Why do you think sometimes teachers do not prepare or use common marking schemes?

.....

Question four.

Does a well prepared marking scheme improve the consistency of scoring mathematics examination? Yes [] No []

Part 3: Moderation and consistency of marking

Question one.

Indicate by ticking (√) in the appropriate box to show the extent of agreement using the words;

- Strongly agree SA
- Agree A
- Uncertain U
- Disagree D
- Strongly disagreed SD

The following can improve consistency of marking among teachers marking same exams scripts

	Teachers opinion	SA	A	U	D	SD
1.	Use of common marking scheme.					
2.	Moderation of the marking scheme.					
3.	Discussing any difference in interpretation of marking scheme					
4	Setting specific standards before real scoring e.g. how to deal with bad handwriting, excessively short or long responses, candidate misunderstanding task.					
5.	Sampling by subject heads.					
6.	Training of teachers as examiners					

Question two

Which other ways can score variation be reduced among teachers marking same scripts.

.....
.....

Part 4: Other factors affecting consistency of scoring examinations.

Question one

Indicate by putting a tick (√) whether you agree or disagree with the following statement using the following words.

- Strongly agree SA
- Agree A
- Uncertain U
- Disagree D
- Strongly disagreed SD

The following are some of the causes of Score variation among teachers marking the same scripts.

	Teachers opinion	SA	A	U	D	SD
1.	Organization of the answers on the scripts					
2.	Handwriting of the students					
3.	Personal feeling about particular students e.g. either he/she is very good or very poor					
4.	Decrease in Markers accuracy over time due to fatigue and lack of attention					
5.	Quality of scripts marked before it					

Question two.

The following student characteristics can influence markers judgment.

	Characteristics	SA	A	U	D	SD
1.	Gender					
2.	Race					
3.	Social class					
4.	Physical attractiveness					
5.	Knowledge of the student and previous performance					

Question three.

What do you think could be other sources of score variation among teachers marking the same school based examination paper in your school?

.....

Appendix B: Interview Schedule for deans of studies/deputy principals.

The interview aims at getting your opinion pertaining the marking, preparation and moderation of marking schemes. The information you give is for research purpose only and will be treated with confidentiality.

Questions

1. Is the school private or public?
2. How many streams are there in your school?
3. How many trained KNEC Examiners are there in your school?
4. Do they participate in marking national examination every year?
5. Do teachers in your school participate in workshop or seminars organized for setting marking and moderation of exams if any?
6. What is the school policy on setting and marking exams e.g.
 - When is the marking scheme prepared?
 - Who monitor the quality of the marking schemes?
 - Do teachers have meeting to discuss the marking scheme to ensure consistency in marking?
 - Who ensure that the teachers mark the exam scripts?

7. Are there cases when each teacher prepares his/her own marking scheme? If yes specify what could cause this?
8. Are there times when teachers set questions they cannot be able to prepare marking criteria?
9. Do student classify teachers based on how they mark? And does this affect how they select the elective subjects.
10. What other challenges are you experiencing as an administrator that can comprise the score given by teachers?
11. What measures are you taking to reduce score variation or improve consistency in marking among teachers?

Appendix C: Mathematic Examination Paper.

SECTION I (50 MARKS)

1. Use logarithms, correct to 4 decimal places, to evaluate: (4 marks)

$$\sqrt{\frac{1.794 \times 0.038}{1.243}}$$

2. Solve the following inequalities and represent the solutions on a single line. (2 marks)

$$3 - 2x < 5$$

$$4 - 3x \geq -8$$

3. A two digit number is 27 less than the value of the number formed by reversing the digits. If the sum of the digits is 15. Find the number. (4 marks)
4. Use reciprocal and square root tables to evaluate, to 4 significant figures, the expression. (3 marks)

$$\frac{5}{0.04796} + \sqrt{583.6}$$

5. During a certain period the exchange rate at Petamax Exchange Bureau was as follows.

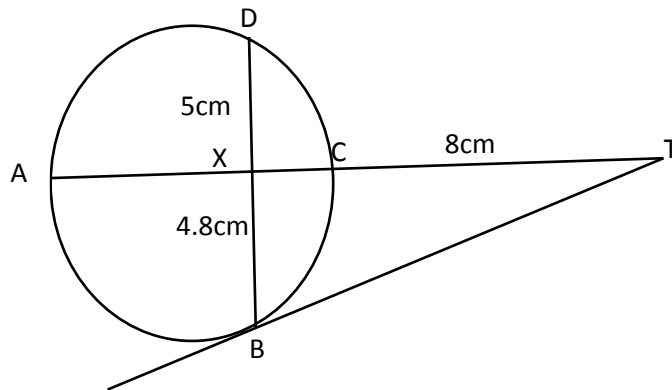
	Buying (Ksh)	Selling (Ksh)
Deutchmark (DM)	19.68	19.74

A tourist arrived in Kenya with 5840DM which he exchanged into Ksh. He spent $\frac{2}{3}$ of the money touring various sites and changed the balance to DM. Calculate his balance giving your answer to 4s.f. (4 marks)

6. Determine the value of x for which the matrix $\begin{pmatrix} 2x & x^2 \\ 2 & 1 \end{pmatrix}$ has no inverse. (2 marks)

7. Simplify $\left(x + \frac{1}{y}\right)^2 - \left(x - \frac{1}{y}\right)^2$ (3 marks)
8. Without using calculators and mathematical tables, evaluate and simplify completely the logarithms $2\text{Log}_{10} 3 + \frac{1}{2}\text{Log}_{10} 64 - \text{Log}_{10} 12$. (2 marks)
9. The coordinates of points O, P, Q and R are (0,0), (3,4), (11,6) and (8,2) respectively. A point T is such that the vector **OT**, **QP** and **QR** satisfy the vector equation $\mathbf{OT} = \mathbf{QP} + \frac{1}{2}\mathbf{QR}$. Find the coordinates of T. (3 marks)
10. Solve the simultaneous equation $xy - 5x = 32$ and $x + y = 17$ (4 marks)
11. The length and breadth of a rectangular floor were measured and found to be 4.1m and 2.2m respectively. If absolute error of 0.01m was made in each of the measurements. Find the:
- Maximum and minimum possible area of the floor. (2 marks)
 - Maximum possible wastage in a carpet ordered to cover the whole floor. (1 mark)
12. Simplify $\frac{4}{\sqrt{5} + \sqrt{2}} - \frac{3}{\sqrt{5} - \sqrt{2}}$ (3 marks)
13. Given that $\sin(x+20)^\circ = -0.7660$, find x, to the nearest degree, for $0^\circ \leq x \leq 360$. (3 marks)
14. Make S the subject of the formula in
$$V^2 = \sqrt{\frac{1+S^2}{U^2}} + \frac{U}{3}$$
 (3 marks)
15. (a) Expand $\left(x + \frac{2}{x}\right)^4$ fully and state the constant term of the expansion. (2 marks)
- (b) By getting a suitable substitute for x, use your expansion in (a) above to evaluate $(10.2)^4$ correct to 4 decimal places. (2 marks)

16. In the figure below, BT is a tangent to the circle at B. AXCT and BXD are straight lines AX=6cm, CT = 8cm, BX = 4.8cm and XD = 5cm.



Find the length of XC and BT. (3 marks)

SECTION II (50 MARKS)

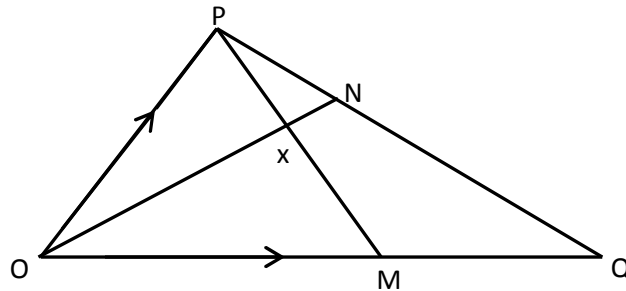
17. Three ships X, Y and Z are approaching a harbour H. X is 16km from the harbour on a bearing of 090° . Y is 14km from the harbor on a bearing of 130° , and Z is 26.31km to the west of Y and on a bearing of 240° from the harbor. Calculate.
- the distance between X and Y. (4 marks)
 - the distance of z from harbor. (3 marks)
 - the distance between X and Z. (3 marks)
18. The distance between two towns A and B is 460km. A minibus left town A at 8.45am and travelled towards B at an average speed of 65km/h. A matatu left B at 10.55am on the same day and travelled towards A at an average speed of 80km/hr.
- How far from town B did they meet. (4 marks)
 - At what time did the two vehicles meet? (2 marks)
 - A motorist started from his home at 9.15am on the same day and travelled to B at an average speed of 120km/h. He arrived at the same time as the minibus. Calculate the distance from B to his home. (4 marks)

19. (a) Find the inverse of the matrix $\begin{pmatrix} 9 & 8 \\ 7 & 6 \end{pmatrix}$ (2 marks)

(b) In a certain week a businessman bought 36 bicycles and 32 radios for a total of Ksh.227 280. In the following week, he bought 28 bicycles and 24 radios for a total of Ksh.174 960. Using matrix method, find the price of each bicycle and each radio that he bought. (4 marks)

(c) In the third week the price of each bicycle was reduced by 10% while the price of each radio was raised by 10%. The businessman bought as many bicycles and as many radios as they had bought in the first two weeks. Find by matrix method, the total cost of the bicycles and radios that the businessman bought in the third week. (4 marks)

20. The diagram below shows triangle OPQ in which M and N are points on OQ and PQ respectively, such that $OM = \frac{2}{3} OQ$ and $PN = \frac{1}{4} PQ$. Lines PM and ON meet at x.



(a) Given that $\mathbf{OP} = \mathbf{p}$ and $\mathbf{OQ} = \mathbf{q}$, express in terms of \mathbf{p} and \mathbf{q} the vectors.

(i) \mathbf{PQ} (1 mark)

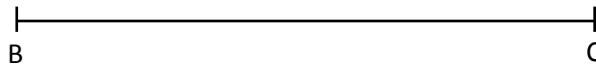
(ii) \mathbf{PM} (1 mark)

(iii) \mathbf{ON} (2 marks)

- (b) Given further that $\mathbf{OX} = k\mathbf{ON}$ and $\mathbf{PX} = h\mathbf{PM}$ where k and h are constants.
- (i) Express \mathbf{OX} in two different ways. (2 marks)
- (ii) Hence determine the value of the constants h and K . (3 marks)
- (iii) Find the ratio $PX: XM$. (1 mark)

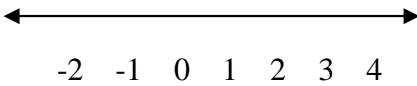
21. Use a ruler and a pair of compasses only for all constructions in the question.

- (a) On the line BC given below, construct triangle ABC such that $\angle ABC = 30^\circ$ and $BA = 12\text{cm}$. (2 marks)



- (b) Construct a perpendicular from A to meet BC produced at D . Measure CD . (2 marks)
- (c) Construct triangle A^1BC such that the area of the triangle A^1BC is three quarters of the area of triangle ABC and on the same side of BC as triangle ABC . Describe the locus of A^1 . (6 marks)

Appendix D: Mathematic Examination Paper Marking Scheme.

1.	<table border="1"> <thead> <tr> <th data-bbox="386 226 581 268"><i>No.</i></th> <th data-bbox="581 226 782 268"><i>Log</i></th> </tr> </thead> <tbody> <tr> <td data-bbox="386 268 581 331">1.974</td> <td data-bbox="581 268 782 331">0.2539</td> </tr> <tr> <td data-bbox="386 331 581 401">0.038</td> <td data-bbox="581 331 782 401">$\bar{2}.5798$</td> </tr> <tr> <td data-bbox="386 401 581 573">1.243</td> <td data-bbox="581 401 782 573">$\bar{2}.8337$</td> </tr> <tr> <td data-bbox="386 573 581 663">$\sqrt{\quad}$</td> <td data-bbox="581 573 782 663">0.0945</td> </tr> <tr> <td data-bbox="386 663 581 747">2.342×10^{-1}</td> <td data-bbox="581 663 782 747">$\bar{2}.7392 \div 2$</td> </tr> <tr> <td data-bbox="386 747 581 810"></td> <td data-bbox="581 747 782 810">$\bar{1}.3696$</td> </tr> </tbody> </table>	<i>No.</i>	<i>Log</i>	1.974	0.2539	0.038	$\bar{2}.5798$	1.243	$\bar{2}.8337$	$\sqrt{\quad}$	0.0945	2.342×10^{-1}	$\bar{2}.7392 \div 2$		$\bar{1}.3696$	
<i>No.</i>	<i>Log</i>															
1.974	0.2539															
0.038	$\bar{2}.5798$															
1.243	$\bar{2}.8337$															
$\sqrt{\quad}$	0.0945															
2.342×10^{-1}	$\bar{2}.7392 \div 2$															
	$\bar{1}.3696$															
2.	$\begin{array}{l} -2x < 5 - 3 \quad -3x \geq -8 - 4 \\ -2x < 2 \quad -3x \geq -12 \\ X > -1 \quad x \leq 4 \end{array}$ 															
3.	<p>Let the two digit number be xy</p> $10x + y + 27 = 10y + x$ $10x - x + y - 10y = -27$ $9x - 9y = -27 \text{ dividing by } 3$ $x - y = -3 \dots \dots (i)$ $x + y = 15 \dots \dots (ii)$ $2x = 12 \quad x = 6 \quad y = 9 \quad \text{Therefore the no. is } 69.$															
4.	$\frac{5}{0.04796} + \sqrt{583.6} = \frac{5}{4.796 \times 10^{-2}} + \sqrt{5.836 \times 10^2}$ $= (5 \times 0.2086 \times 10^2) + (2.416 \times 10)$ $= (5 \times 20.86) + 24.13 = 128.46$															
5.	<p>Amount in ksh = $5848 \times 19.68 = \text{Ksh } 114,931.2$</p> <p>Remaining amount = $\frac{1}{3} \times \text{ksh } 114931.2 = \text{Ksh } 38,310.4$</p> <p>Remaining amount in DM = $\frac{38310.4}{19.74} = \text{DM } 1940.75 \approx 1941 \text{ (4.s.f)}$</p>															

6.	$\det(M) = 0$ $2x - 2x^2 = 0$ $2x(1 - x) = 0$ $x = 0 \text{ or } 1$
7.	$\left(x + \frac{1}{y}\right)^2 - \left(x - \frac{1}{y}\right)^2 = \left(x^2 + \frac{2x}{y} + \frac{1}{y^2}\right) - \left(x^2 + \frac{2x}{y} + \frac{1}{y^2}\right)$ $= x^2 - x^2 + \frac{2x}{y} + \frac{2x}{y} + \frac{1}{y^2} - \frac{1}{y^2}$ $= \frac{2x}{y} + \frac{2x}{y} = \frac{4x}{y}$
8.	$\log 9 + \log \sqrt{64} - \log 12$ $\log 9 + \log 8 - \log 12 = \log \left(\frac{9 \times 8}{12}\right) = \log 6$
9.	$\mathbf{QP} = \begin{pmatrix} 3 \\ 4 \end{pmatrix} - \begin{pmatrix} 11 \\ 6 \end{pmatrix} = \begin{pmatrix} -9 \\ -2 \end{pmatrix}$ $\mathbf{QR} = \begin{pmatrix} 8 \\ 2 \end{pmatrix} - \begin{pmatrix} 11 \\ 6 \end{pmatrix} = \begin{pmatrix} -3 \\ -4 \end{pmatrix}$ $\mathbf{OT} = \begin{pmatrix} -8 \\ -2 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} -3 \\ -4 \end{pmatrix}$ $= \begin{pmatrix} -9.5 \\ -4 \end{pmatrix}$ the coordinates of T is (-9.5,-4)
10.	$x + y = 17$ (i) $xy - 5x = 32$ (ii) From (i) $y = 17 - x$ Sub.in (ii) $x(17 - x) - 5x = 32$ $17x - x^2 - 5x = 32$. $(x - 4)(x - 8) = 0$. $x = 4 \text{ or } 8$ $y = 13 \text{ or } 9$
11.	(a) Maximum possible area = $4.11 \times 2.11 = 9.0831 \text{ m}^2$ Minimum possible area = $4.09 \times 2.19 = 8.9571 \text{ m}^2$
	(b) Max.wastage = Max.Area - Min.Area $= 9.081 - 8.9571 \text{ m}^2 = 0.126 \text{ m}^2$

12.	$\frac{4}{\sqrt{5}+\sqrt{2}} - \frac{3}{\sqrt{5}-\sqrt{2}} = \frac{4(\sqrt{5}-\sqrt{2})-3(\sqrt{5}+\sqrt{2})}{(\sqrt{5}+\sqrt{2})(\sqrt{5}-\sqrt{2})}$ $= \frac{4\sqrt{5}-4\sqrt{2}-3\sqrt{5}-3\sqrt{2}}{5-2} = \frac{\sqrt{5}-7\sqrt{2}}{3}$
13.	<p>Acute angle = $\cos^{-1} 0.7660$ $(x + 20) = 230^{\circ}, 310^{\circ}, 570^{\circ}$ $= 50^{\circ}$ $x = 210^{\circ}, 290^{\circ}$</p>
14.	<p>$v^2 - \frac{u}{3} = \sqrt{\frac{1+s^2}{u^2}}$ Squaring both sides. $u^2 \left(v^2 - \frac{u}{3}\right)^2 = 1 + s^2$</p> $\left(v^2 - \frac{u}{3}\right)^2 = \frac{1+s^2}{u^2} \quad s = \pm \sqrt{u^2 \left(v^2 - \frac{u}{3}\right)^2 - 1}.$
15.	<p>(a)</p> $\left(x + \frac{2}{x}\right)^4 = 1(x)^4 \left(\frac{2}{x}\right)^0 + 4(x)^3 \left(\frac{2}{x}\right)^1 + 6(x)^2 \left(\frac{2}{x}\right)^2 + 4(x)^1 \left(\frac{2}{x}\right)^3 + 1(x)^0 \left(\frac{2}{x}\right)^4$ $= x^4 + 8x^2 + 24 + \frac{32}{x^2} + \frac{16}{x^4}$ <p>(b)</p> <p>$x = 10$ $(10.2)^4 = 10^4 + 8(10)^2 + 24 + \frac{32}{(10)^2} + \frac{16}{(10)^4}$ $= 10000 + 800 + 24 + 0.32 + 0.0016$ $= 10824.3216 \approx 10820.$</p>
16.	<p>$6 \times XC = 5 \times 4.8$ $BT^2 = 18 \times 8$ $XC = \frac{5 \times 4.8}{6}$ $BT^2 = 18 \times 8$ $= 4 \text{ cm}$ $BT = \sqrt{144} = 12 \text{ cm}$</p>
SECTION II (50 MARKS)	
17.	<p>a) $XY^2 = 14^2 + 16^2 - 2 \times 14 \times 16 \cos 40^{\circ}$ (b) $\frac{HZ}{\sin 40^{\circ}} = \frac{14}{\sin 30^{\circ}}$ $= 196 + 256 - 343.187911$ $HZ = \frac{14 \times \sin 40^{\circ}}{\sin 30^{\circ}}$ $= 108.812089$ $= 18.00 \text{ km}$ $XY = \sqrt{108.812089}$ $= 10.43 \text{ km.}$</p> <p>c) $XZ^2 = 16^2 + 18^2 - 2 \times 16 \times 18 \times \cos 150^{\circ}$ $= 580 - 576 \times \cos 150^{\circ}$ $= 1078.83$ $XZ = \sqrt{1078.83}$</p>

$$= 32.85 \text{ km.}$$

18.

(a)

Distance covered by minibus by 10.55 a.m.

$$\begin{aligned} D &= S \times T && 1055 \text{ hrs} \\ &= 56 \text{ km/h} \times 2 \frac{10}{30} \text{ hrs} && \underline{0845 \text{ hrs}} \\ &= 140 \frac{5}{6} \text{ km} && \underline{0210 \text{ hrs}} \end{aligned}$$

Distance apart = $460 - 140.83 \text{ km} = 319.17 \text{ km}$

Relative speed = $65 + 80 \text{ km} = 145 \text{ km.}$

$$\text{Time taken before meeting} = \frac{319.17}{145} = 2.201 \text{ hours}$$

$$\text{Distance from B} = 80 \text{ km/h} \times 2.201 = 176.08 \text{ km.}$$

(b)

$$\text{Time taken before meeting} = \frac{319.17}{145} = 2 \text{ hrs } 12 \text{ min.}$$

1055 hrs

0212 hrs

1307 hrs

The two vehicles meet at 1.07p.m

(c)

$$\text{Time taken by minibus from A to B} = \frac{460}{65} \text{ h} = 7 \text{ hrs } 5 \text{ minutes.}$$

Time taken by motorist 1550 hrs

0915 hrs

0635 hrs

Distance of the motorists' home from B

$$= 120 \text{ km/h} \times 6 \frac{35}{60} \text{ h}$$

$$= 790 \text{ km.}$$

<p>19.</p>	<p>(a)</p> $\text{Det (M)} = (9 \times 6) - (7 \times 8) = -2 \quad M^{-1} = \frac{1}{-2} \begin{pmatrix} 6 & -8 \\ -7 & 9 \end{pmatrix}$ <p>(b)</p> $36b + 32r = 227\,280$ $28b + 24r = 174\,960 \text{ Dividing by 4.}$ $9b + 8r = 56\,820 \quad \begin{pmatrix} 9 & 8 \\ 7 & 6 \end{pmatrix} \begin{pmatrix} b \\ r \end{pmatrix} = \begin{pmatrix} 56\,820 \\ 43\,740 \end{pmatrix}$ $7b + 6r = 43\,740$ $\frac{1}{-2} \begin{pmatrix} 6 & -8 \\ -7 & 9 \end{pmatrix} \begin{pmatrix} 9 & 8 \\ 7 & 6 \end{pmatrix} \begin{pmatrix} b \\ r \end{pmatrix} = \frac{1}{-2} \begin{pmatrix} 6 & -8 \\ -7 & 9 \end{pmatrix} \begin{pmatrix} 56\,820 \\ 43\,740 \end{pmatrix}$ $\begin{pmatrix} b \\ r \end{pmatrix} = \frac{1}{-2} \begin{pmatrix} -9000 \\ -4080 \end{pmatrix}$ $\begin{pmatrix} b \\ r \end{pmatrix} = \begin{pmatrix} 4500 \\ 2040 \end{pmatrix}$ <p>The price of the bicycle was Ksh 4500 and that of radio Ksh 2040.</p> <p>(c)</p> <p>New bicycle price = $\frac{90}{100} \times \text{ksh } 4500 = \text{ksh } 4050$.</p> <p>New radio price = $\frac{110}{100} \times \text{ksh } 2040 = \text{Ksh } 2244$.</p> $\begin{pmatrix} 36 & 32 \\ 28 & 24 \end{pmatrix} \begin{pmatrix} 4050 \\ 2244 \end{pmatrix} = \begin{pmatrix} 217\,608 \\ 167\,256 \end{pmatrix}$ <p>Total Cost = ksh 217 608 + 167 256 = ksh 384 864.</p>
<p>20.</p>	<p>(a)(i) $PQ = -p + q$</p> <p>(ii) $PM = -p + \frac{2}{3}q$</p> <p>(iii) $ON = P + \frac{1}{4}PQ$</p> $= p + \frac{1}{4}(-p + q)$ $= p - \frac{1}{4}p + \frac{1}{4}q$ $= \frac{3}{4}p + \frac{1}{4}q$ <p>(b) $OX = kON$</p> $= k \left(\frac{3}{4}p + \frac{1}{4}q \right)$ $= \frac{3}{4}kp + \frac{1}{4}kq \dots\dots\dots(i)$

	$OX = OP + PX$ $= p + hPM$ $= p + h\left(-p + \frac{2}{3}q\right)$ $= p - hp + \frac{2}{3}hq$ $= (1 - h)p + \frac{2}{3}hq \dots\dots(ii)$	$(c) \frac{3}{4}k = 1 - h \dots\dots(i)$ $\frac{1}{4}k = \frac{2}{3}h \dots\dots(ii)$ <p>From (ii) $k = \frac{8}{3}h$ substituting this in (i)</p> $\frac{3}{4} \times \frac{8}{3}h = 1 - h$ $2h + h = 1 \quad 3h = 1$ $h = \frac{1}{3} \text{ and } k = \frac{8}{3} \times \frac{1}{3} = \frac{8}{9}$
21.	<p>21. (a) Constructing $30^\circ \dots\dots M_1$ Locating A and Completing triangle.....M_1</p> <p>(b) Constructing perpendicular.....M_1 Measuring $CD = 5.2 \text{ cm} \pm 0.1 \text{ cm}$.</p> <p>(c) Area of $ABC = \frac{1}{2} \times 12 \times 7.5 \times \sin 30^\circ$ $= 22.5 \text{ cm}^2 \dots\dots A_1$</p> <p>Area of $A^1BC = \frac{3}{4} \times 22.5 \text{ cm}^2$ $= 16.875 \text{ cm}^2 \dots\dots A_1$</p> <p>Position of $A^1 = \frac{16.875 \times 2}{7.5} = 4.5 \text{ cm} \dots\dots A_1$ i.e. $\frac{1}{2} \times 7.5 \times h = 16.875 \text{ cm}^2$</p> <p>Construction of locus of A.....M_2 Describing locus of A.....M_1 i.e. A is 4.5 cm away from B and also parallel to line BC.</p>	