

**EFFECT OF ADJUSTING CUT SCORES ACROSS YEARS ON
STUDENTS' CLASSIFICATION ON GRADING CATEGORIES OF
MALAWI SCHOOL CERTIFICATE OF EDUCATION
EXAMINATIONS**

M.ED. (TESTING, MEASUREMENT AND EVALUATION) THESIS

By

BITMON KACHINGWE PINDANI

B. Ed (Tech) – University of Malawi

Submitted to the Department of Educational Foundations, Faculty of Education, in partial
fulfillment of the requirement for the degree of Master of Education
(Testing, Measurement and Evaluation)

**UNIVERSITY OF MALAWI
CHANCELLOR COLLEGE**

AUGUST, 2015

DECLARATION

I, Bitmon Kachingwe Pindani hereby declare that the text of this dissertation titled: *“Effect of adjusting cut scores across years on students’ classification on grading categories of Malawi School Certificate of Education Examination”* is substantially my own work and has not been used for any other award at the University of Malawi or any other University.

BITMON KACHINGWE PINDANI

Full Legal Name

Signature

Date

CERTIFICATE OF APPROVAL

The undersigned certify that they have read and recommend to the Postgraduate Studies and Research Committee and the Senate for acceptance of this thesis entitle *effect of adjusting cut scores across years on students' classification on grading categories of Malawi School Certificate of Education examinations*: submitted by Bitmon Kachingwe Pindani in partial fulfillment of the requirements for the degree of Master of Education (M Ed).

SIGNATURE: _____ **DATE:** _____

BOB W CHULU, PhD (SENIOR LECTURER)

MAIN SUPERVISOR

SIGNATURE: _____ **DATE:** _____

RICHARD NYIRONGO, PhD (SENIOR LECTURER)

MEMBER, SUPERVISORY TEAM

DEDICATION

I dedicate this work to my two sons, Witness and Thokozani who endured lonely times with my absence. Father, Mr. Kachingwe Pindani who guided me in the fear of the LORD and inspired me to like school, my wife Fenless for the encouragement. May the Good Lord bless you all? To you and the rest I say, “...*weeping may endure for a night but joy cometh in the morning.*” (*Psalms 30:5*).

ACKNOWLEDGEMENTS

I would like to acknowledge the help that a number of people rendered to me during the development of this work.

First, I would like to express my sincere gratitude to The Malawi National Examinations Board (MANEB), through the Executive Director, for funding the study. I would also like to thank Dr. Bob W. Chulu for kindly and ably supervising the production of this thesis. I am also indebted to A. C. Banda, C. Nzunga, S. Mafeni, W. Mgunda, D. Ayami (Mrs.), W. Lipenga, F. Lupiya, K. Luwe (Ms), C. Kansale, (Miss.) G. Kansale (Miss) and E. Manjombe for the various roles they played during data collection and data analysis stages of this study. May the good Lord bless them richly?

Lastly, but by no means the least, I want to appreciate the support and love I got from my lovely wife, Fenless, and our wonderful sons, Witness and Thoko. Above all, I thank God for His loving-kindness to us all.

ABSTRACT

The purpose of this study was to examine the Effect of adjusting cut scores across the years on grading categories of Malawi School Certificate Education Examination (MSCE) Mathematics paper I. The study was conducted in view of the need for comparability of standards across different administrations to ensure consistency in examination results and enhancement of public confidence in high stakes examinations such as the MSCE. It sought to find out how scores from different annual administrations hence different standard setting processes of the test relate to each other and to investigate the effect of changing the cut scores from one cohort to another on the grading categories. The study made use of quantitative data obtained from the three test (2005, 2006 and 2007) forms that were administered on equivalent groups of a sample of year 2012 form four (grade 12) students. It was found that pupils who wrote the 2005 and 2007 forms of the test performed significantly better than those who sat for the 2006 version, an indication that the 2006 form was more difficult than the 2005 and 2007 forms. The findings of the study, therefore supports the view that despite the best effort efforts by test developers to match the content and difficulty level of the forms across cohorts, it is impossible to construct strictly parallel forms of the test. This, then, suggest that it is fair to have different cut scores on different test forms since these forms are different in difficulty despite being similar in content.

TABLE OF CONTENTS

ABSTRACT.....	vi
TABLE OF CONTENTS.....	vii
LIST OF FIGURES	x
LIST OF TABLES.....	xi
LIST OF APPENDICES.....	xii
LIST OF ACRONYMS AND ABBREVIATIONS.....	xiii
CHAPTER ONE	1
INTRODUCTION	1
1.0 Chapter Overview	1
1.1Background.....	1
1.2 Statement of the Problem.....	3
1.2.1 Purpose of the Study	4
1.2.2 Specific Objectives	4
1.2.3 Significance of the Study	4
1.2.4 Research Questions.....	5
1.4 Definition of the terms	5
1.5 Delimitations and Limitations.....	6
1.5.1 Delimitations.....	6
1.5.2 Limitations	6
1.6 Outlook of the Thesis.....	6
CHAPTER TWO	8
LITERATURE REVIEW	8
2.0 Chapter Overview	8
2.1 Definition of Standard Setting	8
2.1.1 Content versus Performance Standards.....	10
2.1.2 Content Standards	10
2.1.3 Performance Standards	10
2.2 Standard Setting Models and Methods	11
2.2.1 Relative and Absolute Standards (Norm-Referenced Tests-NRT and Criterion-Referenced Tests- CRT)	12

2.3 Test-Centered and Examinee-Centered Tests	14
2.3.1 Test-Centered Models	14
2.3.2 Examinee – Centered Methods	18
2.3.3 Compromise Models	18
2.4 A Priori and Posteriori Methods	20
2.5 Standard Setting Procedure as Done by Malawi National Examinations Board (MANEB) ...	20
2.5.1 Standard Setting Assumptions	21
2.5.2 Performance Levels	21
2.5.3 The Standard Setting Procedure used by MANEB	21
2.5.4 Guidelines for Determining Cut Scores in Mathematics	22
2.6 Standard Setting Procedures by Other Examinations Boards	25
2.6.1 The General Certificate and School Certificate of Education (Grade 12) Examinations by Examinations Council of Zambia (ECZ)	25
2.6.2 The General Certificate of Education (GCE A- level Examinations).....	28
2.6.3 The Scottish Certificate of Education (SCE) Examinations	29
2.6.4 The International Baccalaureate (IB) Examinations	30
2.7 Review of Some Standard Setting Studies.....	31
2.7.1 The Need for Comparable Standards across Cohorts	31
2.7.2 Standard Setting Perspectives	32
2.7.3 Standards Set by Different Panels of Judges	33
2.7.4 Standards Set by Different Methods	34
CHAPTER THREE	36
METHODOLOGY	36
3.0 Chapter Overview	36
3.1 The Design	36
3.1.1 Targeted Population	37
3.1.2 Sample Size.....	37
3.1.3 Sampling Procedure	38
3.1.4 Instruments.....	40
3.1.5 Data Collection	41
3.2 Data Analysis	43
3.2.1 Comparison of Scores Obtained from Different Standard setting processes	43

3.2.2 Comparisons of Standards across the Years	43
3.2.3 Effect of Adjusting the Cut Scores (pass/fail) across the Years	44
3.3 Ethical Considerations	44
3.4 Validity and Reliability	44
3.4.1 Validity	44
CHAPTER FOUR.....	47
RESULTS AND DISCUSSION OF THE FINDINGS.....	47
4.0 Chapter Overview	47
4.1 Equivalency of the Three Test Forms in Difficulty	47
4.2 Comparison of Cut-Scores across the Years/Cohorts	50
4.3 Consequences of Changing Cut (Pass/Fail) Scores across the Years	54
4.3.1 When the 2005 Cut Scores are used across the Cohorts	54
4.3.2 When the 2006 Cut Scores are used across the Cohorts	57
4.3.3 When the 2007 Cut Scores are used across the Cohorts	59
CHAPTER FIVE	64
CONCLUSIONS, IMPLICATIONS AND RECOMMENDATIONS	64
5.0 Chapter Overview	64
5.1 Summary	64
5.2 Conclusions and Implications	65
5.3 Recommendations.....	67
5.4 Suggestions for Further Research	67
REFERENCES	69
APPENDICES	74

LIST OF FIGURES

Figure 1	Relative (Norm – Referenced) Standard setting Method.....	13
Figure 2	Absolute (Criterion – Referenced) Standard setting Method	14
Figure 3	Mean Score Plot.....	50
Figure 4	Percentage Comparison based on 2005 Cut Scores.....	56
Figure 5	Percentage Comparison based on 2006 Cut Scores.....	58
Figure 6	Percentage comparison based on 2007 Cut scores.....	60

LIST OF TABLES

Table 1	Descriptive Statistics of scores on Test Form.....	48
Table 2	ANOVA Table for Between and Within Group Differences	48
Table 3	Group Differences on Test Forms.....	49
Table 4	Comparison of Cut Scores across Test forms	51
Table 5	Standardized and Transformed Score Equivalents... ..	52
Table 6	Cumulative Frequencies for each test form.....	55
Table 7	2005 Cut Scores.....	56
Table 8	2006 Cut Scores.....	57
Table 9	2007 Cut scores.....	59

LIST OF APPENDICES

Appendix A	2005 MSCE Mathematics Paper I.....	74
Appendix B	2006 MSCE Mathematics Paper I.....	81
Appendix C	2007 MSCE Mathematics Paper I....	88
Appendix D	Request letter to MANEB for data	94
Appendix E	Request letter to EDMs for permission to conduct research.....	95
Appendix F	Request letter to schools to conduct research.....	96
Appendix G	Percentage cumulative frequencies for each test form.....	97
Appendix H	Pictures of candidates taking the test.....	99
Appendix I	MSCE Mathematics summary of content area.....	101

LIST OF ACRONYMS AND ABBREVIATIONS

A-level	Advanced Level
ANOVA	Analysis of Variance
CDSS	Community Day Secondary School
CRT	Criterion Referenced Tests
CTT	Classical Test Theory
ECZ	Examination Council of Zambia
EDM	Education Division Manager
GCE	General Certificate of Education
IB	International Baccalaureate
JCE	Junior Certificate of Education ²
MANEB	Malawi National Examinations Board
MSCE	Malawi School Certificate of Education
NRT	Norm Referenced Tests
PAEC	Public Accounts Examinations Council
PSLCE	Primary School Leaving Certificate Examination
SCE	Scottish Certificate Examinations
SEED	South East Education Division
SWED	South West Education Division
UEE	University Entrance Examinations
UNIMA	University of Malawi

CHAPTER ONE

INTRODUCTION

1.0 Chapter Overview

This chapter provides background information to the research problem, the statement of the problem, the purpose of the study and its research questions. The chapter also discusses the study's conceptual framework and its significance.

1.1 Background

Scores on tests are often used as one piece of information in making important decisions. Regardless of the type of the decision that is to be made, it should be based on the most accurate information possible (Kolen & Brennan, 2004, pp. 1-2).

One of the reasons for using examinations as a basis for making certification and admission decisions is that they are believed to be fair to all examinees (Khembo, 2004). "The need for fair and comparable performance standards in high stakes examinations cannot be overstated" (Khembo, 2004, p. vi). Fairness is required not within the same group only but also across cohorts as well. In other words, this year's certified candidates should not be held to a high or lower standard than the standard applied to the previous year's candidates. But if the standards applied to grade examinees are not comparable, then the meaning of the certificate is unclear.

Examination agencies all over the world that conduct high stakes examinations create new items every year because their previously used arsenals have been exposed, hence considered a source of construct-irrelevant variance (Chulu & Sireci, 2011). In Malawi organizations such as Public Accounts Examination Council (PAEC) and Malawi National Examinations Board (MANEB) are no exception to the practice. This means that there is need for a new test form to be administered on each test date (year). MANEB and PAEC allow for multiple sitting of the examinations for one to qualify for awarding of a full certificate. The University of Malawi (UNIMA), too, permits the accumulation of the required six credits over three years to qualify for University Entrance Examinations (UEE), which may lead to one being admitted into its constituent colleges. Chulu & Sireci, 2011 state:

“When public examination results are released, there is always a debate within the media about whether educational standards are changing. The percentage of students who have passed is always compared to the previous passing percentages and a determination is made regarding whether students are doing better or worse than previous cohorts and whether education standards are rising or not.”

The fact that the general public, the media and other stakeholders in the country and world over make comparisons of the scores/grades across cohorts, calls for a psychometric investigation and hence evidence of the comparability of these scores.

The increase in the number of remark requests by candidates who are dissatisfied with their grades, the time and the monetary costs involved makes it imperative that examination agencies adopt a robust assessment strategy. The concerns by the general

public, tertiary educational institutions, with respect to services offered by the students who have completed the secondary education are also an issue worth noting.

1.2 Statement of the Problem

Malawi National Examinations Board and most testing agencies all over the world that conduct high stakes tests create new items every year. The use of different test forms on different test dates suggests a potential problem (Kolen & Brennan, 2004). Even though test developers attempt to construct test forms that are as similar as possible to one another in content and statistical specifications, the forms will still differ in difficulty (Kolen & Brennan, 2004). One of the challenges facing examination institutions today is to ensure that standards remain the same over time (Chakwera, 2004). To go with the challenge of maintaining standards is the process of establishing the cut scores on psychometrically different test forms administered to different cohorts. Since in Malawi, (1) examinations are set every year, (2) that they are different in difficulty and (4) that they are not statistically equated, hence standards are set every year on each and every test form. Literature suggests that standards differ across occasions; methods used to set them and panels (George, Haque, & Oyebode, 2006; Koffler, 1980; Norcin, 2003; Khembo, 2004). The main objective of the current study is to establish the comparability of the cut scores set on different occasions by not exactly the same panels and to investigate the effect of changing the cut scores across cohorts.

1.2.1 Purpose of the Study

The purpose of the study is to investigate the effect of changing cut scores across years on students' classification into grading categories on Malawi School Certificate of Education (MSCE) Examinations.

1.2.2 Specific Objectives

The specific objectives of the study are to:

- Establish how scores from different annual administrations, hence different standard setting processes of the test, relate to each other,
- Investigate the effect of adjusting the cut score from one cohort to another on student classification.

1.2.3 Significance of the Study

Since decisions about examinations involve the future of individuals, the procedures used must be as accurate as possible and defensible. The focus must be on the standard setting process that would provide test score information that would correctly classify examinees into 'masters' and 'non-masters' or in terms of what examinee can and cannot do (Zoani, 1989).

Results of a number of studies on standard setting (Andrew & Hecht, 1976; Hambleton, On the Use of Cut-Off Scores With Criterion Referenced Tests in Instructional Settings, 1978; Koffler, 1980) have shown that there are procedural differences among the methods which in turn lead to different standards. Most of these studies dwelt on

comparisons between methods; finding out which method is more stringent than the other or which one is more reliable and valid (Hambleton R. K., 2001; Sireci, Hambleton, & Pitoniak, 2004; Pell & Roberts, 2006; George, Haque, & Oyeboode, 2006). Some studies still have focused on the application of the methods to actual public examinations such as Malawi School of Education-MSCE (Zoani, 1989), yet others looked at consistency and comparability of the standard setting processes (Khembo, 2004). Chulu and Sireci (2011) looked at the importance of equating high stakes examinations. It appears therefore that there is little if any detailed research on the effect of changing cut scores across years in public examinations.

1.2.4 Research Questions

The present study addressed the following questions:

1. How do the cut scores set on different occasions compare?
2. What are the consequences of adjusting cut scores across the cohorts?

1.4 Definition of the terms

Cut Score/ cut off score; a point on a score scale in which scores on or above that point are in different category or classification than the scores below the point.

Test form: examination paper

Cohort: a group of examiners who write the same examination form or paper.

Performance level: degree of mastery e.g. pass, credit or distinction

Passing Score: Is the lowest score that permits an examinee to be deemed competent, to receive a license or credential or gain admission.

National Secondary School: a secondary school where students are selected for admission from different district across Malawi.

Community Day Secondary School (CDSS): a secondary school where students are selected from the surrounding area

Grant Aided Secondary School: a secondary school established and run by the church but receives financial and material support from government.

1.5 Delimitations and Limitations

1.5.1 Delimitations

The study focused only on mathematics paper I only instead of the two papers, hence generalizability is limited. It was also not possible to involve more judges in the study and from different geographical positions within the country because of the limited scope of the study.

1.5.2 Limitations

Sampling all schools in the country would be appropriate for the study but due to time and financial constraints, only schools from two of the six education divisions were sampled. However the sampled schools are true representation of the schools in country.

1.6 Outlook of the Thesis

The first chapter introduces background to the study's problem. After presenting examination fairness and cut score issues, the chapter presents the study's purpose, research questions, limitations and delimitations

Chapter two reviews the related literature and research. The review focuses on general information on standard setting. The chapter then looks at some standard setting models and methods as well as standard setting procedures done by MANEB and other examinations organizations. The last section presents a review of research findings and general information on comparability of standard setting procedures.

Chapter three describes how the whole study was carried out. The list of questions to be answered in the study is presented first followed by the design of the study. The chapter then describes the sampling procedures, data collection instruments, administration of the instrument and data gathering, data analysis, ethical consideration, validity and limitation of the study.

Chapter four presents results and discussion of the study. First are the findings and discussions on the equivalency of the three groups of students who wrote the three test forms. Secondly, the findings and discussion on comparability standards (pass/fail) across the years are presented. Lastly, the chapter presents the consequences of adjusting the cut scores across the years.

In chapter five, implications and recommendations of the study based on the findings to the research questions are presented.

CHAPTER TWO

LITERATURE REVIEW

2.0 Chapter Overview

The literature review has seven sections. The first section gives the general information on standard setting. The second section discusses content and performance standards. The third section looks at some standard setting models and methods. Other categories of standard setting procedures are discussed in the forth. Standard setting procedures done by MANEB are discussed in the fifth. Sixth section discusses the standard setting procedures by other testing organizations related to this study. The last section presents a review of research findings, research reports as well as general information on comparability of standard setting procedures.

2.1 Definition of Standard Setting

Standard setting refers to the process of establishing one or more cut scores on a test that are used to divide a distribution of test scores into two or more categories of performance, representing distinct levels of knowledge, competency or proficiency in a given domain (Cizek, 1996; Eckes, 2012; Bejar, 2008; Pitoniak, 2010). Cusimano, (1996), cited by George, et al, (2006), defined “standard setting as the process of deciding what is good enough”. Norcin, (2003), refers to the cut score as a cut point or standard and defines it as

a special score that serves as the boundary between those who perform well and those who do not. Norcin, (2003), continues to say “standard setting is a systematic way of gathering value judgments, reaching consensus and expressing that consensus as a single score on a test” (Norcin, 2003, p. 464). Thus, examinees may be categorized as pass/fail, or may be placed into a greater number of ordered performance categories, of basic, proficient, and advanced or pass, credit and distinction (Eckes, 2012).

The practice of standard setting involves various approaches used to arrive on a cut score to differentiate the various levels of proficiency on a test. The outcome of any assessment is determined by the standard setting method used.

Hambleton, (2001) says that most policy makers and the general public usually fail to distinguish between the two types of standards, namely: content standards and performance standards. It is also important that this study defines the two types of standards. Linn and Herman, 1997 in Hambleton (2001) defines content standards as referring to the curriculum and what learners (examinees) are expected to know and to be able to do. That is what examinees are expected to learn. Performance standards on the other hand refer to the level of performance that is expected of examinees to demonstrate. Thus how well examinees are expected to perform in relation to the content standards. Norcin, 2003 defines a standard as a special score that serves as a boundary between those who perform well enough and those who do not (Norcin, *Setting Standards on Educational Tests*, 2003). But Hambleton (2001), however, makes a further distinction between performance standards and cutoff scores by stating that cutoff scores are points

on the score scale that separate examinees into performance categories while performance standards as corresponding to the category descriptions of the cutoff scores.

2.1.1 Content versus Performance Standards

2.1.2 Content Standards

Phillips, 1994 describes Content Standards as the knowledge, skills and other understandings that schools should teach in order for other examinees to obtain high levels of competency in challenging subject matter. They represent what students are expected to learn. They often get operationalized in terms of curriculum guides, becomes the subject of the text books. Represent the goals and objectives of teaching and assessment (Phillips, 1994).

2.1.3 Performance Standards

Performance standards specify how good enough is “good”. They indicate how skillful or competent a student’s demonstration must be to indicate attainment of the content standards. They may also involve judgments of what distinguishes an adequate from outstanding level of performance (Hambleton R. K., 2001, p. 92). They indicate both the nature of the evidence such as an essay, mathematical proof, scientific experiment, project examination or combination of these, required to demonstrate that content standards have been met and quality of student performance that will be deemed acceptable,(i.e. what merits a passing or an A grade).

2.2 Standard Setting Models and Methods

Central to the role of any examinations board is the certification of examinees and the comparability of standards in examinations. In order to maintain comparability, a certain standard has to be achieved and maintained overtime. This ensures consistency in examination results and enhances public confidence.

Pitoniak (2010) defines standard setting as a process that determines performance levels on a test (Pitoniak, 2010). Standard setting involves various methods that are based on the methods postulated by Angoff, Ebel, Nedelsky and Bookmark. These methods call for either statistical (item-centered), rational judgment (person-centered) or both as approaches of arriving on cut score to differentiate various levels of proficiency on a test. Setting performance standards involves the identification of cut scores that divide examinees into pass or fail categories. They are also used to assign grades such as A, B, C, D and F. Generally, different standard setting procedures yield different cut scores and each has its own advantages and disadvantages (Pitoniak, 2010). All standard setting methods involve judgments; therefore they are subject to classification errors (Norcin, 2003; Hambleton, 1978; Pitoniak, 2010) implying that where the cut score is placed some people who should pass will fail (false negatives) and some who should fail will pass (false positives). It is therefore recommended that the method employed should fit the purpose of the score use and the structure of the test.

Standards can be classified as “ absolute or relative” generally referred to as criterion-referenced or norm-referenced” “a priori” or “a posteriori” and “test-centered” or “examinee-centered” (Cizek, 1996; Sanju, Haque, & Oyeboade, 2006; Norcin, 2003).

2.2.1 Relative and Absolute Standards (Norm-Referenced Tests-NRT and Criterion-Referenced Tests- CRT)

1. Relative (Norm Referenced) Standards

Relative (norm) Standards setting method compares how well the examinee has performed compared to others who took the test and the outcome (pass/fail) depends on the performance of the group (Sanju, Haque, & Oyeboade, 2006, p. 2). It is expressed as a number or percentage of examinees, so that the cut point is set for example at the score that will pass 60 best performers or separate the top 25% from the bottom 75% (Pell & Roberts, 2006, p. 93). Relative Standards are most appropriate for examinations where the purpose is to identify a certain number of examinees. Norm or relative standards are easy to use and understand. They can easily be explained to trainees and variations in test difficulty are automatically corrected as the pass mark is influenced by the performance of the examinee cohort.

Despite the advantages of the norm/relative methods stated earlier, some examinees will always fail regardless of their performance. Moreover examinees can influence the pass or cutoff score. The third disadvantage is that the pass/cut score is not known in advance (Figure 1). These include tests that are used to select the highest scores for selection or placement where limited number of students can be accommodated.

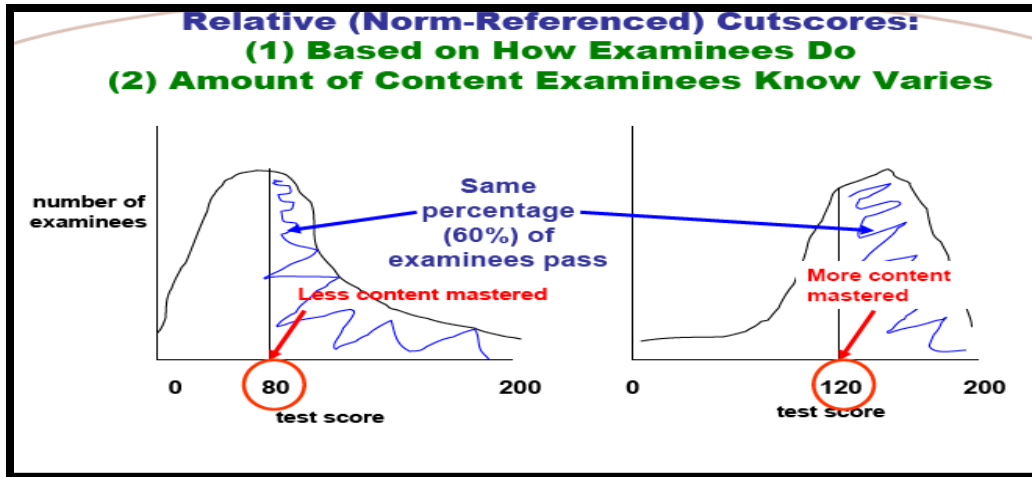


Figure 1 Relative (Norm-Referenced) Standard setting Method

Source: Pitoniak, 2010

2. Absolute (Criterion- Referenced) Standards

In absolute or criterion Standards on the other hand, the pass/fail score is determined by how well an examinee performs and is judged against an arbitrary set external standard (Sanju, Haque, & Oyeboode, 2006, p. 2). They are expressed as a number or percentage of the test questions, so that the cut point is set for example at 67% or two-thirds correct of the questions on the examinations (Figure 2). Absolute standards are most appropriate for examinations of competence, where the purpose is to establish that the examinees know enough for a particular purpose such as final or exit examinations and tests for certification and licensure.

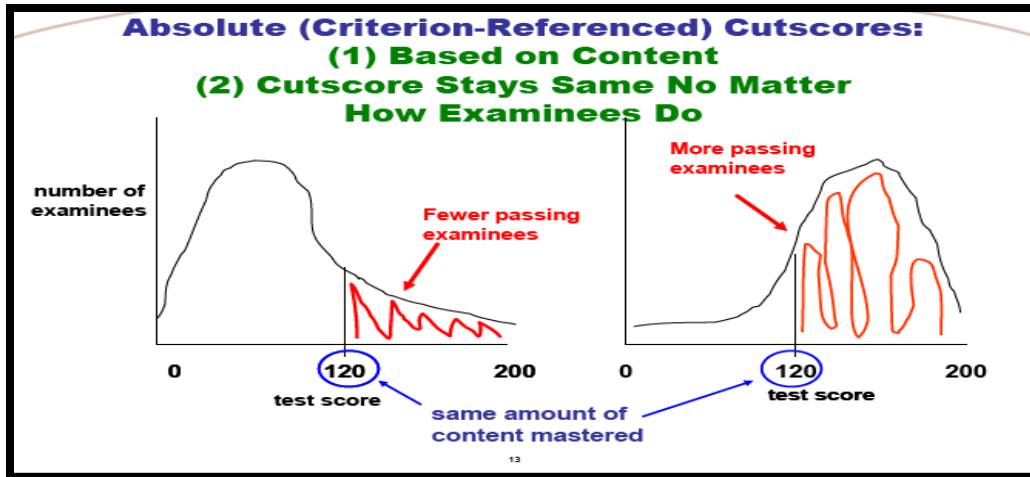


Figure 2 Absolute (criterion- referenced) standard setting method

Source: Pitoniak, 2010

2.3 Test-Centered and Examinee-Centered Tests

There are a variety of absolute standard-setting procedures that have been applied to student assessment (Norcin, 2003; Cizek, 1996). These models can be classified as either test centered or examinee centered. For test- centered methods, hypothetical decisions based on test content are used to derive a standard. For examinee-centered methods, judgments regarding actual examinee performance are used to determine appropriate cut points (Boulet, De Champlain, & McKinley, 2003).

2.3.1 Test-Centered Models

In test-centered models, subject matter experts, also known as participants or judges set the cutoff score by reviewing the items in the test and deciding on the level of performance on each item that is considered for the field under consideration. This leads to the determination of a minimum performance level (MPL) for the test quantified by

combining the judges' projections of the performance of a "minimally competent examinee" of all items. Examples of test-centered models are: Angoff method, Nedelsky's method and Ebel's method (Cizek, 1996).

1. Nedelsky's Method (1954)

Nedelsky's method is a test-centered, absolute or criterion-referenced. The method involves assigning values to multiple-choice test items based on the likelihood of the examinees being able to rule out incorrect options. In this method, participants carefully inspect test items and identify for each item an option which a hypothetical minimally incompetent examinee would rule out as incorrect. They then use reciprocal of the remaining number of options to compute Nedelsky's rating, i.e. the probability that the "F-D examinee" would answer the questions correctly. They finally sum the ratings to derive at a passing score. F students are failing students who would be attracted to options rejected by minimally competent students in a multiple choice test. D students are those with the lowest minimum competence, but are able to reject incorrect answer obvious to them. F-D students are students who possess just enough knowledge to reject F-responses and must choose among the remaining responses at random. To illustrate Nedelsky's method Cizek, 1996 states:

"... in a five option multiple-choice test with 50 items, the analysis could be as follows: if two options in each item were ruled out as incorrect, Nedelsky's rating would be: $5-2 = 3$ remaining options. As a probability this translates to 0.33. The sum of the options for the 50 items would be $0.33 * 50 = 16.5$ which becomes the cut score for the test." (1996, p.22)

Limitations of the Nedelsky's method as cited by Berk, 1984 and Shepard, 1980 in Cizek, 1996 are: the method can only be used with the multiple-choice format. The scale does not permit probabilities of 0.5 or higher. The judges tend not to assign probabilities of 1.0 (that is to assert that all examinees will answer an item correct). The last limitation is that the method often results in standards that are lower than those obtained using other methods.

2. Ebel's Method (1972)

This is also a test-centered, absolute and criterion based method. The methodology proposed by Ebel (1972) also requires participants to make judgments about the test items. The participants provide: Estimates of the difficulty of individual test items, judgment about the relevance of the test content areas. Predictions about the examinees' expected success on the combinations about the difficulty and relevance dimensions.

Participants are asked to categorise items according to three difficulty levels (easy, medium, and hard) and four relevance levels (essential, important, acceptable, and questionable), giving a 3 x 4 matrix. Judges read each item then consign it to one of the categories in the classification table. Judgments about the percentage of borderline candidates who would pass each category are made. A passing score is then calculated. Participants then make judgments about how minimally proficient examinees will perform on the test. This is usually in the form of expected percentage correct for each difficulty-by-relevance combination (Cizek, 1996; Pell & Roberts, 2006). Cizek states that there is one advantage of the Ebel's method which is that it can be used with item formats other than multiple choices. The disadvantages include the following: the method

reveals inadequacies in the test construction process by asking as to why questionable and irrelevant items should be included in the examination (Cizek, 1996).

3. Angoff's Method (1971)

In Angoff's method, judges are asked to first define the characteristics of the borderline group of examinees with a 50% chance of passing. They then consider the difficulty and importance of the first item on the test. Each judge estimates what percentage of hypothetical borderline examinee will respond correctly to the test. This judgment is often by data on the performance of the examinees. The judges discuss their estimates and are free to change them and then proceed in the same manner with the remainder of the items on the test. Finally, the judges' estimates are averaged for each item and the cut point is set at the sum of these averages (Norcin, 2003). The method is the most vigorously researched because it has been used widely and sometimes with modifications. In the modified Angoff procedures, two or more rounds of rating are included. Such modifications are desirable to provide for participants to see how their ratings compared with other participants' ratings before generating final ratings. It is also recommended that participants be provided with normative data such as the actual difficulty indices on one or more of the rounds. This step is desirable as a means of promoting reasonable conceptualisation of anticipated examinee performance.

Advantages of the Angoff Method are many. It is relatively easy to use. It also produces easy-to-obtain and acceptable results in many situations because it is not difficult to explain. Again, data collection and analysis are simpler than other methods in this category (Cizek, 1996). Cizek quoting Berk, (1986) states that Angoff Method appears to

offer the best balance between technical adequacy and practicability. The method can be applied to both multiple-choice and free response formats. (Norcin, Setting Standards on Educational Tests, 2003) says that disadvantages of the Angoff Method are: that judges find it difficult to concentrate on borderline (minimally competent) candidates, and will often consider the average examinee instead thereby inflating the pass mark; judges are uncomfortable pulling percentages out to the air; it can be tedious” (Norcin, 2003, p. 94).

2.3.2 Examinee – Centered Methods

In examinee-centered models, judges determine the cutoff scores and make pass/fail decisions about actual examinees after they have written the test, rather than a hypothetical minimum performance level (Cizek, 1996). Instead of providing judgments based on test materials, the judges are asked to review a series of examinee performances and make judgments about the demonstrated level of proficiency. In examinee –centered tests, descriptive statistics (distribution properties of data), such as mean, median, mode, standard deviation, skewness and kurtosis are examined and cutoff scores are set based on them. A decision may be made for example to fail all test takers who scored at less than one standard deviation below the mean. Examples of examinee-centered examinations include borderline group and contrasting group methods.

2.3.3 Compromise Models

There is also another category of setting standards known as the Compromise Method. These models were “...developed to strike a compromise between purely norm-referenced (relative) approaches and absolute methods. These methods can be used to derive passing scores outright or to adjust standards obtained using other

methods.”(Cizek, 1996, p. 25). Examples of compromise models according to Cizek (1996), include Beuk’s method and Hofstee method.

1. Beuk’s Method

In Beuk’s Method, each participant in the standard setting procedure is asked to make two judgments. First the participants are asked to estimate the minimum level of knowledge required to pass an examination, expressed as a percentage of the total raw score on the test. Secondly, to estimate the passing rate expected, expressed as a percentage of the examinee population. When the examination has been administered, these expectations are compared with reality. If the expectations differ from reality, a compromise between the two can be struck using the information provided by the participants’ judgments (Cizek, 1996). A graph of passing rate against percent correct is then plotted. A straight line is then drawn on the intersection of points of the means of both the expected percentage correct against expected passing and the adjusted percentage correct against the adjusted passing rate. The point, at which the straight line passes the curve showing functional relationship between percentages of successful examinees and possible cutting scores, is used to derive the consequent passing rate. The recommended passing score can be obtained by multiplying the adjusted percent correct by the number of items in the animation.

2. Hofstee’s Method

The method is implemented by asking each participant to respond to four questions: What is the maximum acceptable passing score? What is the minimum acceptable

passing score? What is the minimum acceptable fail rate? What is the maximum acceptable fail rate?

The four points are then plotted on a graph to form a box. The top left and bottom right corners of the box are joined, where this line crosses the candidates' performance line gives the pass score and failure rate. The advantages of the Hofstee's Method are that it is easy to implement. The educators are comfortable with the decisions and it can be used with other methods. One main disadvantage is that there is a possibility that the candidates' performance line may not pass through the box defined by the participants.

2.4 A Priori and Posteriori Methods

In a priori methods of standard setting, judges select the cut score prior to the test administration. It is generally based on judgments about the difficulty of test items for a certain group of individuals. Examples include: Angoff, Modified Angoff, Ebel, Nedesky and Jaeger. In contrast to a priori, the posteriori methods apply only after the test data have been collected. A cut score is set based on the actual rather than hypothesized test performance. Examples are the contrasting and the borderline group.

2.5 Standard Setting Procedure as Done by Malawi National Examinations Board (MANEB)

During standard setting as done by Malawi National Examinations Board (MANEB), the cut scores are determined paper by paper and the aggregate is used to cut subject grade boundaries for the various grades (Chiunda, 2010).

2.5.1 Standard Setting Assumptions

The following assumptions are made during standard setting (fixing) process by MANEB:

- That the current examinations are comparable to those set in the previous years in terms of levels of difficulty, content coverage, and skills tested, quality of questions and the language used.
- That the examinations are administered under similar conditions at every centre.
- That the cohort taking the examination is comparable in terms of preparedness with the cohorts which took the examinations in the previous years.

2.5.2 Performance Levels

The performance levels of the Malawi School Certificate of Education (MSCE) examinations are divided into the following grade categories (Yadidi D. C., 2010):

- 1-2 signifying pass with distinctions.
- 3-6 signifying pass with credit.
- 7-8 signifying an ordinary pass and
- 9 signifying a fail.

2.5.3 The Standard Setting Procedure used by MANEB

The standard setting procedure involves both norm- and criterion-referencing models in which the quantitative information presents largely the relative performance of examinees where as the chief examiner's report contains qualitative information.

The chairperson of the panel invites the chief examiner to give an account of the examination paper in terms of content validity, accuracy, level of difficulty of the question paper, clarity and relevance of questions, difficulty level of questions, grammatical and typographical errors. The chief examiner also comments on comparability of the examination to past papers, and how the candidates' scripts were marked, highlighting the challenges encountered during marking. He/she then presents suggested cut scores for consideration.

The chairperson invites members of the awards panel to comment on the issues raised by the chief examiner. The panel then reviews the main statistical parameters of the current examination year i.e. Mean score, Mode, Median and Standard Deviation. It compares these to those of the previous three years and percentages of the examinees captured at a given cut-score are considered (Chiunda, 2010, p. 113). The size of the mean score among other factors influences whether or not the current year's cut score should be adjusted.

2.5.4 Guidelines for Determining Cut Scores in Mathematics

Following are the guidelines for determining where to award distinction, credit or pass as used by MANEB.

1. Distinction:

To be awarded a distinction, the examinee has to:

- Demonstrate accuracy in all numerical and non-numerical operations.

- Make logical presentation of solutions showing clear mode of thinking and layout, drawing conclusions such as proficiency in deductive processes, extrapolation from graphical presentations, interpretations of trends and interpretation of statistical data.
- Manifest ingenuity and originality.

2. Credit

To be awarded a credit the examinee must:

- Show good knowledge of Mathematics and must be able to manipulate decimals and irrational numbers,
- Show arithmetical competence in conversion of units in calculations.
- Be able to use decimals and irrational numbers, show arithmetical competency in conversion of units, in calculation of numerical question from various types.
- Be able to use graphical representation for deductions and interpretations using four-figure tables.
- Apply operations to variables, solving linear, simultaneous and quadratic equations.
- Be able to handle algebraic fractions with both numeric and algebraic denominators.
- Have good knowledge of basic/numerical construction and deductive geometry.

- Be able to use trigonometric ratios of right angled triangles including solutions of scalene triangles.
- Demonstrate some knowledge and application of three dimensional geometric figures of spherical geometry.
- Be able to draw trigonometric, algebraic, statistical and linear programming graphs.
- Show that he/she is familiar with vectors, matrices, simple statistics, estimator's errors and accuracy, probability and elementary calculus.

3. Pass

To be awarded a pass the examinee:

- Acquire and/or attain a level of comprehension of mathematical statements given in words and is able to represent them in symbolic form, perform simple mathematical operations such as subtraction, division, multiplication and addition on all types of numbers (integers, fractions and decimals).
- Show to be familiar with types of weights and measures.
- Be able to extract information from tables, schedules, graphs including pie charts and bar graphs.
- Use algebraic variables to the extent of understanding formulas and deals with substitution and solving of simple linear equations, demonstrate competency with the use of geometrical instruments to construct and/or describe triangles, other polygons, measure and calculate lengths and angles, use co-ordinates to plot points on graphs and join the points together, is able to calculate

profits/loss, percentages usage, simple interest in business arithmetic etc.
(Yadidi D. C., 2010).

2.6 Standard Setting Procedures by Other Examinations Boards

The use of experienced judges to apply common standards of performance across different years occurs in many curriculum-based examination programmes conducted at the end of secondary education in a number of countries. In such programmes the challenge is to have judges internalize the standards of student performance that have been established, and then apply them to different forms of the examinations administered in different years. Norcin (1990) reports on the use of judges to produce cut-off score equivalences across different forms of an examination. This and many other studies show that such procedures can be made sufficiently accurate.

2.6.1 The General Certificate and School Certificate of Education (Grade 12) Examinations by Examinations Council of Zambia (ECZ)

Examinations Council of Zambia (ECZ) inherited most of its guidelines and standard of practice from the University of Cambridge Examination Syndicate which include the process of setting performance standards. The procedure used by ECZ for determining cut scores can be characterized as a combination of various methods. Typically, the standards used are relative in nature the council ensures on a balance by collecting relevant evidences concerning the content to make professional judgments on the cut scores.

In this regard, the process of setting performance standards for Grade 12 at ECZ called Standard Fixing and Awards involves various sources of evidence which are taken into account when arriving at cut scores. There are basically five sources of evidence:

1. Chief Examiner's report on the quality of candidate performance and the recommended cut scores.
2. Judgmental exercise involving council's staff (i.e. management, subject officers, research officers, examination officers) and external examiners. The personnel form the awards committee and professional judgments to decide on the final grade boundaries.
3. Test raw score distributions displaying the actual and cumulative frequency distributions. These include the mean, standard deviation, minimum and maximum score, the number of candidates that entered, sat and were absent.
4. A record of performance of the candidates as compared to the past years.
5. The circumstance of the examination year and subject content.

The chief examiner's report on the quality of candidates' performance provides an important source of information used in determining the cut scores. The report is presented to the committee to serve as a guide when setting the grade boundaries.

The committee decides on the grades based on: a record of preliminary cut scores proposed by the chief examiner, the circumstances of the examination year and subject content, test statistics – measures of central tendency and dispersion, a record of performance of the candidates as compared to the previous years and the actual versus the desired (normal) performance of the candidates.

The ‘awarding’ committee considers the chief examiners’ proposed cut scores for grades 8E (pass), 6C (credit) and 2A (distinction) then with iterate discussions on circumstances of the examination year and subject content, test statistics, previous years adopted grades, a basis to fix the final grade boundary is achieved.

The standard setting procedures at ECZ are also concerned with the passing percentage in order to maintain the proportions of candidates at specific grade levels. In this regard, cumulative percentages of mark distributions and other statistics are compared to the previous years’ performances to determine the grade boundaries that allow maintenance of set standards unless judged by the circumstances of the year. Basically, the committee fixes the grade boundaries for distinction; credit and pass (grade 2A, 6C and 8E respectively, while the five intermediary grade boundaries are statistically done using the following formulas:

- Grade 1A (upper distinction) = $\frac{(9\alpha-2C)}{7}$ (I)
- Grade 3B (upper merit) = $\frac{4\alpha+3C}{7}$ (II)
- Grade 4B (lower merit) = $\frac{3\alpha+4C}{7}$ (III)
- Grade 5C (upper credit) = $\frac{2\alpha+5C}{7}$ (IV)
- Grade 7 (pass) = $\frac{C+E}{2}$ (V)

Where: α = grade 2A (Distinction)

C = grade 6C (Credit)

E = grade 8E (Pass)

In total 8 grades are used to describe proficiency in all subjects for the School Certificate and General Certificate of Education. The 8 levels also serve to minimize the classification errors and ensure an accurate interpretation of the examinees' level of proficiency.

The impact of this procedure for setting performance standards for School Certificate at grade 12 is that the pass rate has been averaging 60 percent in the last five years in Zambia.

2.6.2 The General Certificate of Education (GCE A- level Examinations)

Bennett (1998), states that in the General Certificate of Education (GCE) A- levels examinations conducted in England and Wales, the process of determining cut- off scores relating to the various grades awarded involves a team of highly experienced judges who have been involved in the setting and scoring of the examination. Prior to meeting to set the cut-off scores, the judges ensure they are fully conversant with the overall standard of work associated with the cut-off scores established in the previous years. As the main objectives are to maintain the grade standards over time and across different subjects, question papers, scoring keys and student responses defining grade boundaries for previous examinations are reviewed in the context of relevant statistics. The examining board maintains an archive covering a number of years and containing responses awarded each cut-off score. Evidence from the first year of examination, when the performance standards were originally set, is also retained to guide the judges in setting their cut-off scores.

The establishment of cut-off scores relating to the different grades awarded requires the judges to work as a group and take account of a variety of factors. These include examination papers and the scoring keys, samples of student responses to the examination items, technical information relating to the examination and the items (such as facility values for multiple-choice items and mark distributions for papers), statistical information from previous years, grade descriptions, archived examinations scripts, question papers, and details of significant background changes in entry patterns and choice of options (Bennett, 1998).

2.6.3 The Scottish Certificate of Education (SCE) Examinations

In the Scottish Certificate of Education (SCE) examinations, cut-off scores corresponding to the grades awarded are set by subject experts using professional judgment and supported by statistical evidence. The statistical evidence provided includes cut-off scores and distributions of grades awarded in the previous three examinations, and the frequency distribution of students' scores on the current examination.

In order to set the cut-off scores on the examination in each course so that the same standard of performance receive the same grade every year, a meeting is held between senior officers of the Scottish Examinations Board, the Principal Examiner and other subject experts. At this meeting, agreement is reached on the cut-off scores to be applied (Bennett, 1998).

2.6.4 The International Baccalaureate (IB) Examinations

For the International Baccalaureate (IB) examinations, the determination of grade boundaries follows structured process which entails using the professional judgment of a number of examiners supported by statistical data and the examination papers and samples of student responses from previous years. It is common for different teams of judges (examiners) to consider different components of the examination.

The judges responsible for setting the grade boundaries are required to become familiar with the examination paper and consider feedback provided by those who had scored the students' work and those who had prepared the students to sit for the examination. Key points are noted and taken into consideration when samples of students' responses are reviewed.

Histograms that show the score distribution for the various components of the examination are also provided. While these are important, the judges are reminded that they should not be used as the sole basis for determining grade boundaries.

Cut-off scores are established by considering a number of student scripts that scored at and around a set of initial cut-off scores suggested by a senior examiner. Once the members of the team have settled on the cut-off scores, they are given the grade distribution percentages from previous examinations. The judges are able to make further adjustments to the cut-off scores, if they feel changes are warranted (Bennett, 1998).

2.7 Review of Some Standard Setting Studies

In this section, literature on studies that have a direct bearing on the present study has been presented. First is on the need for comparable standards across cohorts, followed by standard setting perspectives. Third to be presented are studies on standards set by different panels of judges (i.e. on different occasions) and finally standards set by methods.

2.7.1 The Need for Comparable Standards across Cohorts

For all examinations, pass/fail decisions must be made and such decisions must be the same over time and for all different forms of the test (Norcin, 1990). This is one of the challenges facing the examinations institutions today: to ensure that standards remain the same over time. Examining institutions need to ensure that the cut scores established each year represent the same level of proficiency. When this is achieved then fairness between cohorts and comparability of inter-year grade will also have been achieved.

Over the years psychometricians have devised and used various methods for maintaining standards. One method has been to develop examinations of equivalent difficulty and maintain the same cut score from one year to the next. But different forms of the test are rarely equal in difficulty (Kolen & Brennan, 2004). Another approach would be to use the same test to successive cohorts of examinees and use the same cut scores. The danger with this approach is that in the long run, the items in the test will have different relevance with repeated administrations (Khembo, 2004). Further, the repeated administration poses a security risk as some examinees may memorize the items or their

content and reveal them to the next cohort of examinees (Chulu & Sireci, 2011). The third and most common way practitioners ensure comparability of standards is by test equating. This is conducted to establish equivalence between test scores. Where test items are disclosed following administration, as is the case in Malawi, it is not possible to do statistical equating of test scores using anchor items. In such situation, judgmental equating, also known as social moderation championed by Waltman (1997) cited by Khembo, 2004 or linking becomes necessary.

2.7.2 Standard Setting Perspectives

It has generally been established that different standard setting methods produce different cut scores (Andrew & Hecht, 1976), resulting indifferent classification of examinees (Khembo, 2004). Because of lack of agreement in standards set by different methods, some authors have written against standard setting itself. Shepard (1979) quoted by Khembo (2004), advised people to avoid setting standards whenever possible. Glass went further by saying that setting performance standards on tests and exercises by known methods is a waste of time because all the methods were arbitrary (Glass, 1978). But Pophan (1978), counter argued that, while performance standards were set judgmentally, it was incorrect to equate human judgment with arbitrariness in this negative sense. Mehrens and Cizek, 2001 quoted by Khembo (2004) also defended standard setting: saying“... to argue against standard setting is to, in effect, argue against making categorical decisions”, (p.49). This is echoed by Hambleton, (1978) who asserted that instructional decisions cannot be made without cut scores.

Whatever one's position is regarding standard setting, the need for agreed guidelines is pertinent. This will ensure that standards or cut scores established by different panels of judges or on different occasions represent the same level of proficiency in performance. When this is achieved, then fairness and comparability of grades between cohorts will also be achieved.

2.7.3 Standards Set by Different Panels of Judges

Different standard setting procedures generally produce different cut scores when applied to the same or parallel test by two or more panels of judges. Khembo (2004), comparing the cut scores set by two panels using the same performance descriptors observed that cut scores set by one panel was higher than those set by another panel and that about 2.7 percent of the examinees would be classified differently if scores of one panel were replaced with those of the other panel (Khembo, 2004). Khembo (2004) also cites findings reported by Jaeger (1980) and Good & Cresswell (1998) where up to 71 percent and 30 percent of examinees respectively could be classified differently if the cut scores set by one panel were replaced with those set by another. In their study, results of a study by Jaeger et al (1980) cited by Khembo (2004) confirm that different panels of judges will produce different cut scores. They independently set passing scores on one of the North Carolina school achievements tests. There were wide variations in the cut scores set. On reading, the proportion of examinees who would have failed if one awarding team was substituted for another ranged from a low of nine percent to a high of 30 percent. The situation was worse in mathematics where failure rate ranged from 14.4 percent to a high

of 71.1 percent. In general, different groups of judges will set different standards when using the same methods, especially if the judges represent different interest groups.

2.7.4 Standards Set by Different Methods

As it has been pointed out, one of the problems with standard setting studies is that different standard setting methods produce different results. Results of study conducted by George, Haque and Oyeboode (2006), found out that there were significant differences in the outcomes of standard setting methods between the modified Angoff and Norm-referenced methods. This was shown by the proportion of candidates who passed and failed the test. The Angoff method produced a 100% pass rate while the norm-referenced i.e. Mean minus 1.0 SD was 85%. Koffler (1980), in his study, compared two methods for setting proficiency standards between Nedelsky and Contrasting group's methods. His findings were that there was no agreement between the cut-off scores developed by two methods and recommended that no one standard setting procedure should be relied upon, but rather a number of procedures should be used. This was echoed by Zoani (1989), in his study in which five standard setting methods of Angoff, Hofstee, Borderline, Contrasting groups and norms approach. Among the findings, were that the Angoff's method produced high and stringent cut scores while Hofstee's method tended to be less stringent. This led to the conclusion that each standard setting method is likely to misclassify some candidates and no two methods will ever produce similar results. Skakun and Kling investigated whether the Nedelsky, the modified Ebel and the normative approaches generate similar passing scores on a national certifying examinations in general surgery (Skakun & Kling, 1980). They also investigated the

effects of different passing scores on overall pass rates. Their results indicated that judges using the Nedelsky procedure set the passing score at 66.7%, while the two Ebel approaches yielded passing scores of 69.7% and 71.7%. The normative approach of setting performance standards established the passing score at 70.6%. The results of this study further indicated that if Nedelsky approach had been employed, then 22.5% of the examinees would have failed the test. While if the Ebel I and Ebel II methods had been employed then 35% and 45.6% would have failed the test respectively. But if the normative method had been employed, the 41.3% of the examinees would have failed the test. This, once again, reinforces the conclusion that different approaches for establishing passing scores on an examination produce different standards.

CHAPTER THREE

METHODOLOGY

3.0 Chapter Overview

This chapter describes how the whole study was carried out. The list of questions to be answered by the study is given first. This is followed by the design of the study, population and sample, data collection instruments, administration of the instrument and data gathering, data analysis, ethical considerations, validity and reliability. In the final section, a narrative of delimitations and limitations of the study is present.

3.1 The Design

The study used the experimental design as its main strategy and specifically, it used a modified post-test only with equivalent groups. The experimental design was chosen to be the study's main research strategy because for one to determine the extent to which the adjusting of cut scores across the years affect the grading categories requires manipulating the variables in the study. This allows the researcher to carry out the most rigorous test of causal hypotheses as the researcher keeps constant all other extraneous variables so that differences to be noted on the dependent variable should be explained only as a result of the manipulated independent variable (Cohen, Manion, & Morrison, 2000). The study featured random assignment of the 2012

MSCE candidates into three equivalent groups. A spiraling process was used during the administration of the tests to create three independent samples of examinees (Group A had 574 students, while Group B had 574 students and Group C had 576 students). In this process, when the test booklets are being handed out, the first student receives the 2005 test form, while the second receives the 2006 form and the third receives the 2007 form. The cycle is repeated with the fourth student receiving the 2005 form and the fifth receiving the 2006 form and so on.

3.1.1 Targeted Population

The study targeted the two education divisions in the southern part of Malawi namely South East Education Division (SEED) and South West Education Division (SWED) and ten educational districts.

3.1.2 Sample Size

The sample size of the study comprised:

- 1724 Malawi School Certificate of Education (MSCE) Examinations candidates in 15 secondary schools from six Educational Districts in the South East Education Division (SEED) and South West Education Division (SWED). The districts are: Balaka, Mangochi, Machinga, Zomba Rural, Zomba Urban in SEED and Blantyre Urban in SWED. Of these, 12 schools were from the SEED and three from the SWED.

- Four (three males and one female) MSCE Mathematics Paper I Markers or raters also known as examiners including chief examiner and senior examiner in the five districts. These scored the students' scripts.
- Ten (10) - (three females and seven males) MSCE Mathematics Paper I standard setting participants (judges).
- 15 head teachers and 18 Mathematics teachers from the sampled schools who motivated the students to write the examinations and helped in the administration of the test papers respectively.

3.1.3 Sampling Procedure

There are two main methods of sampling; probability sampling, also known as random and non-probability sampling also known as non-random or purposive sampling. In probability sampling, all members of the wider population have equal chance of being selected for the sample while in non-probability sampling; members of the wider population do not have equal chances of being selected for the sample (Cohen, Manion, & Morrison, 2000). The study used the multi-stage/phase sampling methodology (Elder, 2009). The first step was to use convenient sampling to identify the two education divisions of SEED and SWED these two divisions were chosen because they are easily accessible to the researcher. This was followed by another convenient sampling to identify the six educational districts from which the sampled secondary schools are located according to the geographical positions of the schools in the study. The third stage used purposive sampling to identify the 15 schools. Of these: two were single sex

schools (one being boys' only and the other girls' only) and 13 were co-education schools enrolling boys and girls, eight were rural schools while seven were urban schools.

The study involved; a total of 1724 students, 1068 of which were boys representing 62% and 656 were girls representing 38%. It also involved 10 subject matter experts who were standard setting judges, three of which were females and seven were males. These set the cutting scores. Of these 10 experts, six officials from the Malawi National Examinations Board (MANEB), two were the teachers who scored the students' scripts (one female and one male) and two were experienced senior examiners well-qualified teachers with more than ten years of teaching experience at secondary school. They had also been involved in assessment activities with MANEB. These activities include item writing, moderation of test items and scoring national examinations. Their experience with MANEB activities was a vital in the standard setting process.

The standard setting Panel comprised three females and seven males. The panelists' qualifications ranged from Diploma in Education to Master's Degree in Education. Stratified random sampling involves dividing the population into homogeneous groups (strata), each group containing subjects with similar characteristics (Cohen, Manion, & Morrison, 2000).

The schools were also sampled according to three performance categories of high, medium and low performance in the 2011 MSCE results. Of the 15 sampled schools, four were high performance, seven were medium performance and four were low performance

schools according to 2011 MANEB examination results. In stratified sampling, "... the population is divided into strata and the strata are represented in the sample through proportional allocation. The researcher does not only define the strata but also how many members of the strata to include in the sample."(Hinkle, Wiersman, & Jurs, 1994, p. 161). Though Fraenkel & Wallen (1996) and Cohen, Manion & Morrison (2000) state that members in the stratified sampling should be allocated proportionately, Hinkle, Wiersman & Jurs, (1994) suggest that equal members can be selected from the strata regardless of their size. Selecting schools according to their performance helped to have a true representation of the national sample.

The third step used random sampling to identify which of the 1724 students to write which paper using spiraling procedure (Kolen & Brennan, 2004).According to Fraenkel and Wallen (1996), a sample size can be any number as large as the researcher can obtain to with reasonable expenditure of time and energy (p.104). Manion, Cohen and Morrison (2000) state that:

There is no clear cut answer to correct size of the sample, for the correct sample size but depends on the purpose of the study and the nature of the population under scrutiny i.e. a sample size of 30 is held by many to be minimum of cases if researchers plan to use some form of statistical analysis on their data. Survey research usually requires a large sample particularly if inferential statistics are to be calculated (Cohen, Manion, & Morrison, 2000, pp. 93-4).

3.1.4 Instruments

The instruments that were used in the study were the 2005, 2006 and 2007 Malawi School Certificate of Education Examination (MSCE) Mathematics Paper I question papers (see appendices A, B and C). MSCE examinations are developed by MANEB to

assess students' mathematical proficiency at the end of secondary education. The papers were purposively chosen because at least five years had gone since the three papers were written to minimize students from giving correct responses due to practice since access to the three papers was very minimal according to the experience judgment of the researcher. The paper I was also chosen because it has many questions unlike paper II which has 12 questions to ensure wide coverage of the syllabus. The three papers were administered randomly to the students using a spiraling procedure to obtain three equivalent groups.

3.1.5 Data Collection

1. Administering of the Test Papers

The test papers were administered to a group of 1724 (62% males and 38% females) MSCE (grade 12) candidates between three to five weeks prior to the start of the 2012 Examinations. This was done to ensure that the students had covered the whole syllabus and that they were fully prepared for the 2012 examinations. This was the time when the schools had already written and revised their "mock" examinations that are normally given to students to assess their preparedness for the forthcoming national examinations and to polish the areas deemed not mastered by the teachers. The design was that three papers were administered at the same time to the students by random assignment in each class using the spiraling procedure. The students were told that they would be writing different forms but not necessarily different papers to avoid some of them to have their preferences hence defeating randomization process. The reason for the randomization was to create three equivalent groups. The spiraling procedure was done to

reduce sources of bias and threat to validity due to selection (Kolen M. J., 2007). After the tests were written, the scripts were scored by a group of four mathematics teachers; comprising one female and three males. These teachers were trained by MANEB and had vast experience in scoring MSCE (grade 12) examinations.

2. Setting Cut Score

After scoring the scripts, a standard setting (awards) meeting was convened to set the cut scores using both qualitative and quantitative information. The qualitative data was provided by the chief examiner who supervised the scoring process. The chief examiner gave an account of the test papers in terms of syllabus coverage, accuracy, level of difficulty of the papers, (this is expressed in terms of average, below or above average without actually making calculations). Then clarity and relevance of questions, difficulty level of questions, and any presence of typographical errors were also reported. He reported that all the three papers were similar in terms of content but noted that the 2006 paper was slightly more difficult than the other two according to the qualitative observation by those who scored the test papers on the general performance of the examinees. The chief examiner finally presented the suggested cut-off scores at each grade.

The panel then reviewed the statistical parameters that were provided in the folders which included: frequency distributions of students on each test mean scores, mode, median and the standard deviations. This type of standard setting is a form of compromise method and is a standard practice for MANEB examinations in which new

cut scores are set every year on the new test form. This allows a balance between normative and absolute information. This is in line with the recommendations by Koffler (1980) who said "... no one standard setting procedure should be relied upon to determine cut scores, but rather a number of procedures should be used" (p. 177). Shephard (1983) also said that compromise methods allow a balance between normative and absolute information in order to set reasonable cut scores.

3.2 Data Analysis

The study generated quantitative data. First, the students' scores were entered into the computer and an analysis was done by researcher. Descriptive statistics were generated using SPSS version 16 software. The graphs and tables were plotted using excel computer software.

3.2.1 Comparison of Scores Obtained from Different Standard setting processes

Equivalence of three groups taking the 2005, 2006 and 2007 test forms was checked by testing the difference in mean scores using One-way between groups Analysis of Variance (ANOVA) with post-hoc tests.

3.2.2 Comparisons of Standards across the Years

To compare the standards across the years, cut scores at each category (pass, credit and distinction) were compared for all the three years using the Z-scores and T- scores.

3.2.3 Effect of Adjusting the Cut Scores (pass/fail) across the Years

Using the descriptive data that were generated from the raw scores obtained from the three cohorts of examinees who wrote the three Mathematics Paper I forms, percentages of students at each category, such as pass, credit and distinction, were compared for the three cohorts to determine the effect of changing the cut scores across the cohorts. The percentages of pass, credit and distinction for the 2005 were compared with the pass, credit and distinction percentages for the 2006 and 2007. This comparison continued with the 2006 percentages of students at pass, credit and distinction levels compared with the 2005 and 2007. Overall pass rate was also compared for 2005, 2006 and 2007 test forms to investigate the effect of different cut scores.

3.3 Ethical Considerations

Written permissions were obtained from the Education Divisional Managers (EDMs) from SEED and SWED, District Education Managers (DEMs), head teachers, proprietors of the private schools teachers who would be employed as raters (examiners) and judges as well as the students themselves to conduct the study in schools under their jurisdiction. The subjects (participants) were further assured of the confidentiality with which the data collected would be held. As a result, no names of individual subjects were used.

3.4 Validity and Reliability

3.4.1 Validity

Validity refers to the degree to which evidence supports any references a researcher makes based on the data he or she collects using a particular instrument (Fraenkel &

Wallen, 1996). Validity therefore depends on the amount and type of evidence there is to support the interpretations researchers wish to make concerning data they collected. There are many kinds of validity evidence and listed are just some of them:

1. Content-Related Evidence of Validity

On content validity, MANEB item writers also known as setters developed the instrument that was used in this study. These item writers are well-trained personnel with vast experience in teaching and scoring MSCE mathematics. The test items are then critiqued and moderated by yet other groups of experts who usually are chief examiners and senior examiners as well as subject matter experts from the university and colleges and mathematics curriculum specialists. During item writing, critiquing and moderation stages, specification tables and/or blue prints are used as a guide on the content and the level of cognitive demands. The specification table helps to maintain consistency of difficulty of the tests over years. The test papers used, therefore, possessed the required validity. The test papers were also administered at least three weeks prior to the start of the 2012 MSCE examinations to ensure that the examinees had completed the syllabus and that they were fully prepared for the 2012 examinations. Hence their responses were taken as true mathematics knowledge and skills.

2. Criterion – Related Evidence of Validity

The results of the original tests by MANEB were used for certification and selection into the university. Most of those who were selected into the university performed well and

have now graduated. Hence the results of the test forms have successfully predicted the performance of those who wrote the tests.

3. Construct – Related Evidence of Validity

On validity of the instruments that were used in the study, the instruments were developed by MANEB. The items are later on critiqued by another group of experts. Furthermore, the items are moderated by yet another group of experts who are experienced in item writing and marking. Since the papers were already administered by MANEB in the past, there is no doubt that they possess the required construct-related validity evidence.

4. Reliability

Reliability refers to the consistency of the scores obtained. How consistent they are from one administration of instrument to another and one set of items to another or from one rater to another. Provisional marking schemes were provided by MANEB to ensure that all the markers mark to the same standard. The provisional marking schemes were also standardized to include all the possible answers. Conveyer system of marking was also employed to reduce inter-rater variability.

CHAPTER FOUR

RESULTS AND DISCUSSION OF THE FINDINGS

4.0 Chapter Overview

In this chapter, results and discussions of the study are presented in three main sections. The sections are formulated according to the research questions already stated earlier in the study. First are the findings and discussions on the equivalency of the three groups of students who wrote the three test forms which answers research question one. The second section answers research question two, presents the findings and discussions on comparability of standards (pass/fail) across the years. The third section presents the consequences of changing the cut scores across the years. Finally the chapter summary is presented.

4.1 Equivalency of the Three Test Forms in Difficulty

Descriptive statistics for all the three test form scores are presented in Table 1. the table gives information about each group's mean, standard deviation, median, mode skewness, kurtosis, range, minimum and maximum score.

One-way between-groups analysis of variance (ANOVA) with post-hoc tests was conducted to explore the difference in difficulty between the three test forms by testing the difference in mean scores. There was statistically significant difference at $p < .05$ level of the three groups of students [$F_{(2,1721)} = 16.11, p = .000$]. Post-hoc comparisons using

the Tukey HSD indicated that the mean score for 2006 test form ($M=32.95$, $SD=22.16$), was significantly different from 2005 form ($M=38.86$, $SD=26.42$) and 2007 form ($M=40.98$, $SD=25.72$). There was no statistically significant difference between the 2005 and 2007 test forms (refer Tables 4.2A and 4.2B).

Table 1: Descriptive Statistics of Students' Performance on 2005, 2006 and 2007 Test Forms

Statistic	TEST FORM		
	2005	2006	2007
N	574	574	576
Mean	38.86	32.95	40.98
Median	37.00	29	40.00
Mode	5	6	40
Std. Deviation	26.42	22.16	25.72
Skewness	.28	.64	.19
Kurtosis	-1.15	-0.38	-1.07
Range	96	95	96
Minimum	0	1	0
Maximum	96	96	96

Table 3: Summary of ANOVA Test of Group Differences on Test Forms

ANALYSIS	Mean Difference	Sig.	95% CI	
			Lower Bound	Upper Bound
Group A (2005) vs Group B (2006)	5.91	.000	2.47	9.35
Group A (2005) vs Group C (2007)	-2.12	.317	-5.56	1.32
Group B (2006) vs Group C (2007)	-8.03	.000	-11.46	-4.59

Test forms pairs of 2005 and 2006, 2006 and 2007 have large mean score differences (5.91 and 8.03 respectively with significance values of .000 but 2005 and 2007 test form pair have a small mean score difference of 2.12. Given that the three were randomly equivalent, this finding suggests that there was difference in difficult level across the 2005, 2006 and 2007 test forms.

Also the mean plot (see Figure1) shows that 2005 and 2007 test forms had mean score differences close to each other but 2006 form was well below the other two forms.

This is in line with literature and other study findings previously done (Kolen & Brennan, 2004; Chulu & Sireci, 2011; Khembo, 2004). The results also suggest that the differences in the mean, mode, median, kurtosis and skewness was not due to the differences in Mathematics proficiency by the cohorts of examinees who wrote the different test forms, but due to differences in the difficulty of the forms. This is in agreement with literature and other previous research findings that despite the best efforts by test developers to match the content and difficulty levels of the test forms across cohorts, it is impossible to construct strictly parallel forms of the test (Crocker & Algina, 1986; Kolen & Brennan, 2004; Khembo, 2004; Chulu & Sireci, 2011).

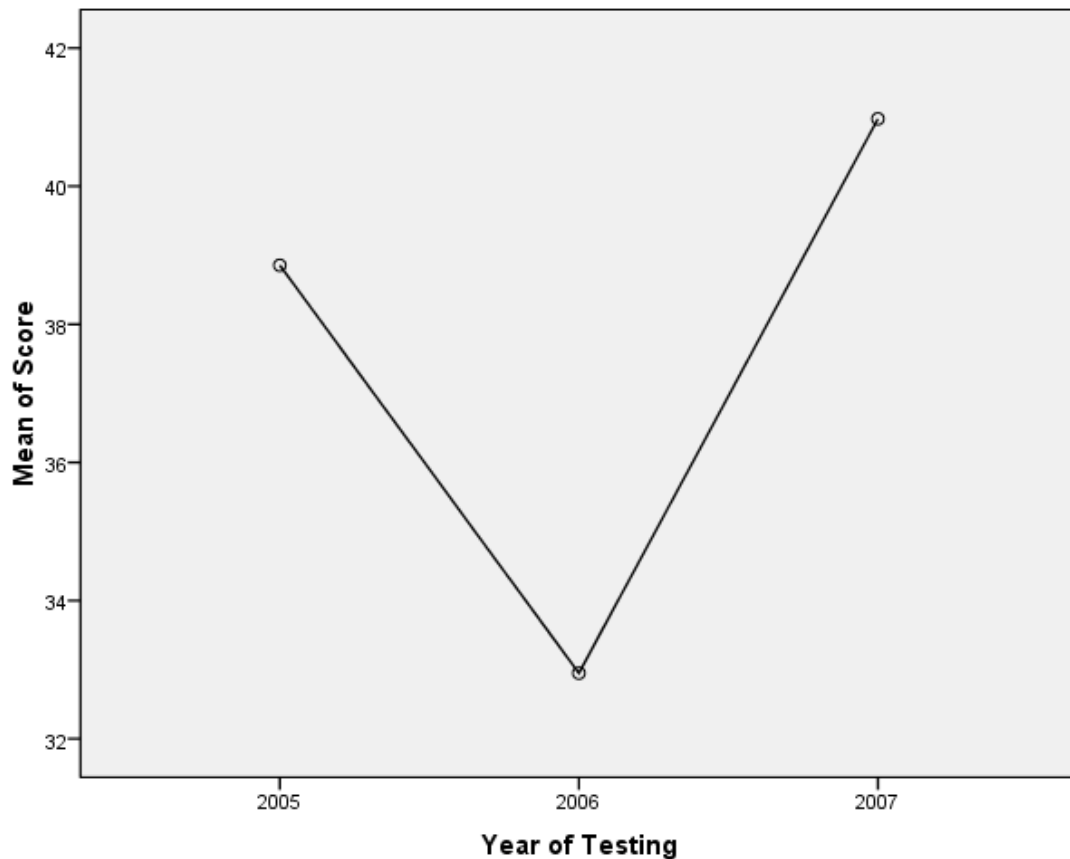


Figure 3: Mean Score Plot

4.2 Comparison of Cut-Scores across the Years/Cohorts - Research Question One

To answer research question one and in order to determine if cut scores obtained from different standard setting processes can yield comparable results on the three forms of the test. The comparison involved computing numbers and percentages of examinees falling in each grade category of pass, credit and distinction. The following two assumptions were made; that the three groups of examinees were equivalent in terms of mathematics ability and that the test forms were similar or comparable in terms of content and syllabus coverage. This was done after controlling the randomization of the groups by spiraling. It

was expected that the cut scores would produce approximately equal proportions of examinees at each category. Table 4 present all the results for all the examinees.

Table 4: Comparison of Cut Scores across Test Forms

Grade Category	TEST FORM								
	2005			2006			2007		
	Cut Score	N	%	Cut Score	N	%	Cut Score	N	%
Distinction	70	102	17.77	64	62	10.80	70	93	16.15
Credit	44	144	25.09	38	149	25.96	47	155	26.91
Pass	26	108	18.82	22	150	26.13	28	119	20.66
Fail	25	220	38.33	21	213	37.11	27	209	36.28

Pass Rates for 2005, 2006 and 2007

Form	N	Pass Rate
2005	574	61.67
2006	574	62.89
2007	576	63.72

The results in Table 4 show that raw cut scores for the three test forms were somewhat different. The 2007 seems to have the highest raw cut scores at all grade categories. The pass/fail raw cut score for 2007 was 28% (28 points out of 100) was greater than the 2005 raw cut score 26% by two points and the 2006 cut score 22% by six points. While the credit/pass cut score for 2007 (47%) was greater than the 2005 cut score (44%) by three points and 2006 cut score (38%) by nine points. However on the distinction/credit raw

cut score, there are similar cut scores for 2005 and 2007 (70%) but they are still greater than the 2006 cut score of 64 by six points.

The results suggest that the raw cut score differences are almost negligible when converted to standardized (Z-score) or transformed scores (T- scores), (Tables 5 and 6). For example, the pass/fail Z-scores for 2005, 2006 and 2007 were 0.487, 0.494 and 0.504 standard deviations below the mean respectively (Table 5). This approximately means that all the pass/fail cut scores are set at one-half standard deviation below the mean. And when the Z-scores are translated into T- scores, the pass/fail cut scores are 45.13%, 45.06% and 44.96% for 2005, 2006 and 2007 test forms respectively. This again translates to a transformed pass/fail cut score of 45 % for all the three test forms.

Table 5: Summary of Standardized (Z-Scores) and Transformed (T-Scores)

Category	2005			2006			2007		
	Raw Score	Z- Score	T- Score	Raw Score	Z- Score	T- Score	Raw Score	Z- Score	T- Score
Pass	26	-0.487	45.13	22	-0.494	45.06	28	-0.504	44.96
Credit	44	0.195	51.95	38	0.228	52.28	45	0.156	51.56
Distinction	70	1.179	61.79	64	1.401	64.04	70	1.127	61.27
Descriptive	Mean = 38.86			Mean = 32.95			Mean = 40.98		
	SD = 26.42			SD = 22.16			SD = 25.74		

At credit/pass cut scores, the z-scores were 0.195 for 2005, 0.228 for 2006 and 0.156 for 2007. This means that for all the three test forms, the credit/pass fail was set at approximately one-fifth of standard deviations above the mean. But when T-scores are

considered, then the credit/pass cut scores were 51.95%, 52.28% and 51.56% for 2005, 2006 and 2007 respectively. All these T-scores round off to 52% as a credit/pass cut scores for all the three test forms. For distinction/credit cut scores, the Z-score equivalents of the raw scores were 1.179, 1.401 and 1.127 for 2005, 2006 and 2007 respectively. This is the only category where there were notable differences. Even when T-scores are considered, cut scores for 2005 and 2006 are set at 64% while for 2007 is set at approximately 61%, giving a difference of three percentage points. A possible explanation for this is that examinees who wrote the 2006 form benefited from a lower pass cut score than those who wrote the 2005 and 2007 forms. This is in agreement with findings by (Khembo, 2004). The credit category seems to follow the trend at pass rates in which there was an increase of approximately one percent from 2005 to 2006 and finally to 2007.

At credit category, the proportions of examinees followed the general performance of pass rates with all the three forms producing approximately similar numbers of 25%, 26% and 27% for 2005, 2006 and 2007 respectively. This means that the standard setting process used, produced similar results at credit category.

At distinction, the 2006 form produced fewer examinees than the 2005 and 2007 forms the lower cut score of 64% compared to the 70% cut score for both 2005 and 2007 forms. This still is proof that the 2006 test form was more difficult than the other two forms, hence further confirming the well-known fact in literature and findings by other researchers that no two or more test forms are similar despite effort by experts to match

the content and difficulty levels of the test forms across occasions (Crocker & Algina, 1986; Kolen & Brennan, 2004; Chulu & Sireci, 2011; Khembo, 2004). This means that to have the same cut score across years to grade examinees who practically write different test forms in terms of difficulty, may be unfair on some examinees that are misclassified by the constant cut score despite having the same level of proficiency to their colleagues.

4.3 Consequences of Changing Cut (Pass/Fail) Scores across the Years - Research

Question Two

Tables 5, 6 and 7, and Figures 4, 5 and 6 show the cumulative percentages of examinees in each test form, the standardized (Z) scores as well as transformed (T) scores for easy comparison on the effect of changing the cut scores across the years and percentages of examinees at each grade category if the cut scores were changed vice versa.

4.3.1 When the 2005 Cut Scores are used across the Cohorts

If the cut scores were not changed, for example if the 2005 cut scores were used across the years: (i.e. if the pass score was pegged at 26 percent, then 61.67% of examinees would pass the 2005 test but only 56.62% would pass the 2006 test as opposed to 62.89% while 66.32% would pass the 2007 test against 63.72% realized when the pass cut scores were changed for each year (Tables 6 and 7, and Figure 4).

Table 6: Cumulative Frequencies for Each Test Form

Raw Score	Cumulative Frequency			Raw Score	Cumulative Frequencies			Raw Score	Cumulative Frequencies		
	2005	2006	2007		2005	2006	2007		2005	2006	2007
0	0.9	0	1	34	48.1	59.8	43.8	68	81.7	92	82.8
1	1.9	1	1.7	35	48.8	61	45.5	69	82.2	92.7	83.9
2	3.1	1.7	3.5	36	49.7	62.4	46.5	70	83.4	93	84.9
3	5.6	4.2	4.9	37	50.5	63.2	47.4	71	84.1	93.4	85.6
4	7	5.1	6.4	38	51.7	64.5	47.9	72	85.2	94.1	86.1
5	9.8	6.3	8.7	39	52.8	65.7	49	73	86.6	94.1	87.2
6	12.4	9.6	9.7	40	53.7	66.7	51.9	74	87.5	94.6	87.3
7	13.9	11.5	11.5	41	54.2	67.9	53.1	75	88.2	95.3	87.8
8	16.4	13.6	13.2	42	55.9	69	54.2	76	88.5	95.3	88.7
9	17.1	15.9	13.7	43	57.1	69.5	54.7	77	89.2	95.8	89.6
10	18.6	17.9	14.6	44	58.7	70.6	55.7	78	90.1	96	91
11	20.6	19	15.6	45	60.1	72	56.6	79	91.1	96	91.8
12	22	21.1	17	46	60.8	74	56.9	80	92.2	96.5	92
13	23.3	22.6	17.9	47	61.8	74.6	58.3	81	93.6	97	92.7
14	24.9	24.7	19.1	48	62.7	75.6	59.9	82	94.4	97.4	94.1
15	26.8	26.7	20.8	49	63.2	76.3	60.9	83	95.8	97.6	94.6
16	28.4	28.6	22.2	50	64.5	76.8	62.2	84	96.2	98.1	95.3
17	29.4	30.7	23.6	51	65.5	78.2	64.2	85	96.5	98.1	95.7
18	30.3	32.4	25.2	52	67.1	78.7	65.3	86	97.6	98.6	96.4
19	30.8	34.3	26	53	68.6	79.6	65.8	87	97.7	98.6	96.5
20	32.9	35.5	27.4	54	69.9	80.7	67.2	88	98.4	99	97.4
21	33.8	37.1	29	55	70.6	82.4	68.1	89	98.6	99.1	98.3
22	35.2	38.5	30.4	56	71.3	83.8	68.9	90	99	99.3	98.6
23	36.1	39.7	32.3	57	72.5	84.1	70	91	99.3	99.3	98.6
24	37.5	41.6	33	58	73.7	85	70.3	92	99.7	99.7	98.8
25	38.3	43.4	33.7	59	75.3	85.4	71.5	93	99.7	99.7	99.3
26	40.2	45.6	35.2	60	76.1	86.2	72.7	94	99.7	99.7	99.5
27	40.8	47	36.3	61	76.8	87.1	74.5	95	99.8	99.8	99.8
28	42.2	48.4	37.3	62	76.8	88.3	75.9	96	100	100	100
29	43.2	50.3	38	63	77.7	89.2	76.2	97	100	100	100
30	43.9	51.7	39.2	64	78	89.9	77.4	98	100	100	100
31	44.4	54	40.8	65	79.1	90.8	79.3	99	100	100	100

Table 7: Percentage Comparison Based on 2005 cut off Scores

Test Form	2005 cut score								
	Pass and Above			Credit and Above			Distinction		
	Old	New	Δ	Old	New	Δ	Old	New	Δ
2005	61.67	61.67	0	42.9	42.9	0	17.8	17.8	0
2006	62.89	56.62	-6.27	36.8	30.5	-6.3	10.8	7.3	-3.5
2007	63.72	66.32	2.6	44.3	45.3	1	16.1	16.1	0

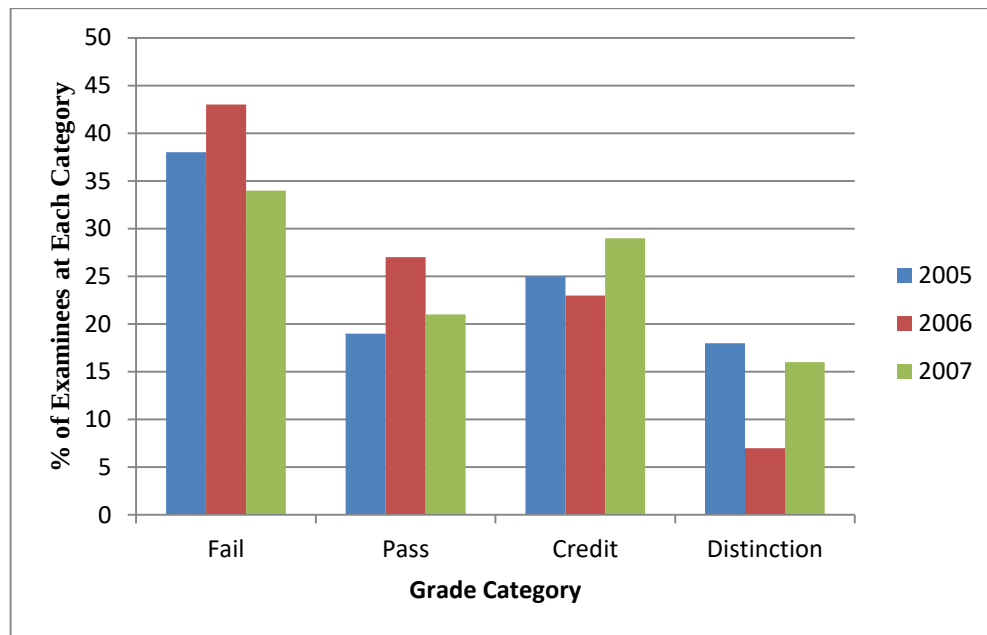


Figure 4: Percentage Comparison Based on 2005 Cut Scores

Results shown in Tables 6, 7 and 8, and Figure 4 suggest that if the cut score of 44 (out of 100) for 2005 was used across all the three test forms, then 30.5% would pass the 2006 test with credit and above as opposed to 36.8% which is a reduction of 6.3%. While those who wrote the 2007 test, 45.3% would pass with credit and above compared to 44.3% representing an increase of only 1%.

For distinction category, the cut scores were 70 (out of 100) for both 2005 and 2007 test forms but 64% for the 2006 test form. If the cut score of 70% which was for 2005 and 2007 was used, only 7.5% would pass the 2006 form with distinction as opposed to 10.8% which is a decrease of 3.5%.

4.3.2 When the 2006 Cut Scores are used across the Cohorts

But if the cut score for 2006 was used across the years (see Table 4.6B and Figure 4.3), which was 22 (out of 100), then 66.2% would pass the 2005 test, an increase of 4.53% and 71% would pass the 2007 test representing an increase of 7.28% while the 2006 pass rate would still be at 62.89%.

Table 8: Percentage Comparison Based on 2006 Cut Scores

Test Form	Pass and Above			Credit and Above			Distinction		
	Old	New	Δ	Old	New	Δ	Old	New	Δ
2005	61.67	66.2	4.53	42.9	49.5	6.6	17.8	22.3	4.5
2006	62.89	62.89	0	36.8	36.8	0	10.8	10.8	0
2007	63.72	71	7.28	44.3	52.6	8.3	16.1	23.8	7.7

But if the credit cut score of 38 (out of 100) for 2006 test was used, then 49.5% of the examinees who wrote the 2005 would pass with credit and above compared to 42.9% which is an increase of 6.6%. While those who wrote the 2007 test, 52.6% of the examinees would pass with credit and above representing an increase of 8.3%.

But if the cut score for 2006 which was 64 (out of 100) was used across the years for establishing distinction category, then 22.3% of the examinees would have passed 2005 test with distinction as opposed to 17.8% representing a 4.5% increase and 23.8% would have passed the 2007 test with distinction compared to 16.1% which is an increase of 7.7% (refer Table 8 and Figure 5).

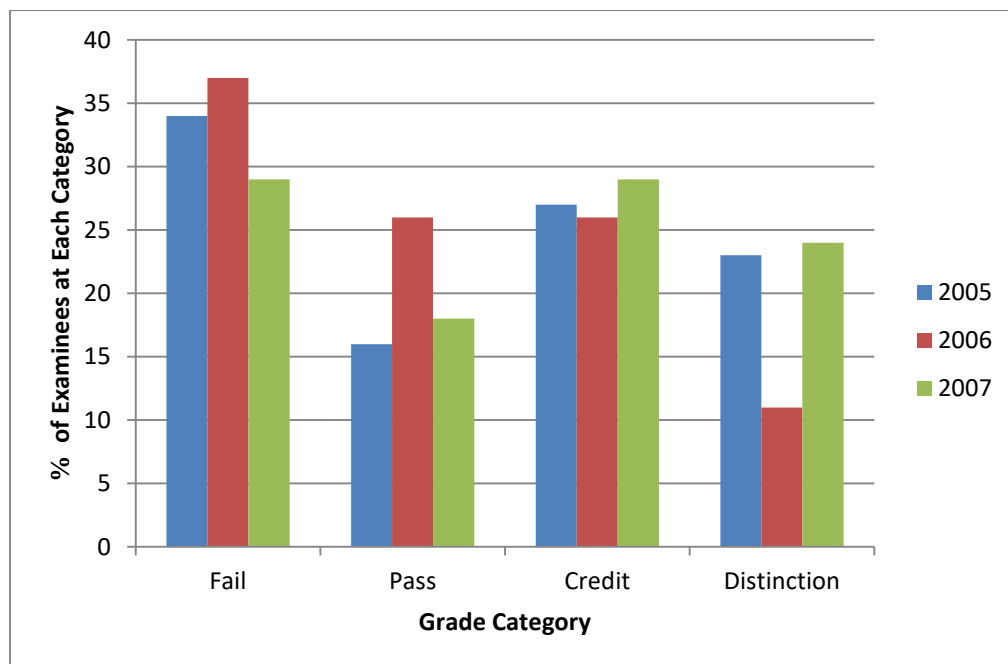


Figure 5: Percentage Comparison Based on 2006 Cut Scores

4.3.3 When the 2007 Cut Scores are used across the Cohorts

On the other hand, if the pass cut score of 28 which was for the 2007 test, was used across the years, then only 59.23% would pass the 2005 test. A reduction of 2.44% but 53% would pass the 2006 test representing a reduction of 9.89% while the 2007 would remain at 63.72% (Refer Tables 6 and 9 and Figure 6).

While if the credit cut score of 45 was used, then 41.3% of those who wrote the 2005 test would pass with credit and above. This is a decrease of 1.6% from 42.9%. But for those who wrote the 2006 test, then 29.4% would pass with credit and above representing a reduction of 7.4% from 36.8% (Refer to Table 9 and Figure 6).

Table 9: Percentage Comparison Based on 2007 Cut Scores

Test Form	Pass and Above			Credit and Above			Distinction		
	Old	New	Δ	Old	New	Δ	Old	New	Δ
2005	61.67	59.23	-2.44	42.9	41.3	-1.6	17.8	17.8	0
2006	62.89	53	-9.89	36.8	29.4	-7.4	10.8	7.3	-3.5
2007	63.72	63.72	0	44.3	44.3	0	16.1	16.1	0

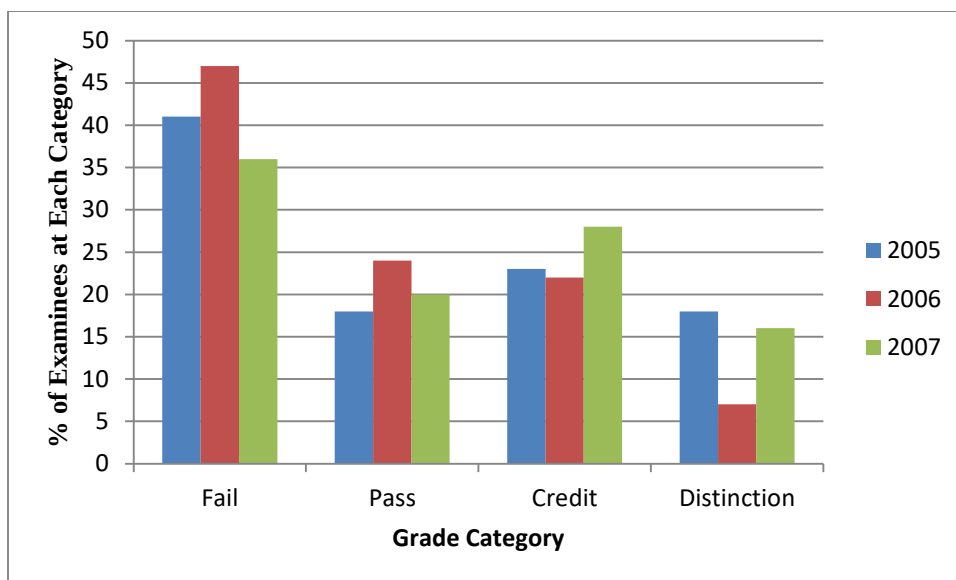


Figure 6: Percentage Comparison Based on 2007 Cut Scores

When proportions or percentages of examinees falling at each performance category are compared, there seem to be similarity in total pass rates and credit performance category. The pass rate for 2005 was 61.67% while pass rate for 2006 was 62.89% and that of 2007 was 63.72% with a difference of between one and two percentage points only (Table 4). In pass category, 2006 had the largest proportion of examinees (26.13%) followed by 2007 (20.66%) and finally 2005 with 18.82% of examinees. The 2006 form had the smallest percentage of examinees at distinction category of 10.8 thus approximately seven percent lower than 2005 which had 17.77% and approximately five percent lower than 2007 which had 16.15% of the examinees. Similar results produced in the pass rate and at credit category of all the three test forms, suggest the same level of mathematics ability and proficiency among the different groups (cohorts) of examinees who wrote the test forms despite the numerically different cut scores, assuming equivalence of the cohorts.

. This is in agreement with the conclusion made by Zoani (1989) who stated that such drastic changes in pass rates would be unacceptable and would bring about a number of questions about the validity of the examinations and quality of teaching (Zoani, 1989). Although the cut scores for the three forms of the examinations were different, they appeared to represent the same level of proficiency, hence grade.

Results of the study suggest that depending on which year's cut scores are used, different number of examinees fail. For example, had the 2005 passing cut scores been used across the cohorts, the 220 out of 574 candidates representing 38.33% would fail the 2005 test while 43.38% (249 out of 574) of the candidates would fail the 2006 test and 33.68% (192 out of 576) of the candidates would fail the 2007 test. But when the 2006 passing cut score is used, then 33.8% (194 out of 574) of the candidates would fail the 2005 test while 37.11% (213 out of 574) of the candidates would fail the 2006 test and 36.28% (209 out of 576) of the candidates would fail the 2007 test. Finally, had the 2007 passing cut score been used, then 40.71% (234 out of 574) of the candidates would fail the 2005 test and 47.04% (270 out of 574) of the candidates would fail the 2006 test while 36.28 (209 out of 576) of the candidates would fail the 2007 test.

The differences in failure rates ranged from 4.65% to 9.7% when 2005 cut scores are used but 3.31% to 8.11% when 2006 passing cut score was used and between 4.47% and 10.72% when passing cut score for 2007 form was used as opposed to a range of .83% to 2.05% realized when the cut scores were changed. A pass rate change of up to 10% from one year to another especially a negative one would create a public outcry as opposed to a

change of 2% which was realized when the cut scores were changed. This supports the fact that since two or more test forms are always different in difficulty even though they are developed following the same specifications, testing organization such as MANEB should vary the cut scores according to cohorts especially if the tests are not equated. This in line with recommendation by Norcin (1990) stated that differences in cut scores that appropriately reflect form difficulty are desirable because they ensure that all examinees face the same challenges.

In conclusion, a cut score for the same grade could be numerically different due to differences in test difficulty, but reflecting the same level of ability. In terms of fairness and comparability, most of the examinees across the years are misclassified when raw scores are not changed from year to year. When raw scores were converted to Z- scores and T- scores cut score changes were very small if not negligible. This again means that although the cut scores for the three test forms of the examinations were different, they appeared to represent the same level of proficiency and hence grade. This is in agreement with another study by Khembo, 2004 who concluded that the same level of proficiency can be represented by different cut scores in different forms of the test or the same cut score can represent different levels of proficiency in different forms of examinations. What is needed is just to make sure that: the same test blue print or specification table is used in constructing the different test forms to satisfy the similarity in their tasks and cognitive demands, the same conditions of administration are applied and the same pattern of scoring is used. In this study, all the three test forms were developed by The Malawi National Examinations Board (MANEB) and all the steps were followed such as

the use of specification table, moderation and the administration was supervised by the researcher personally.

The major contribution of the study is that it has succeeded in making the case for changing cut scores in high stakes examinations especially in the absence of equating which currently may not be possible due its statistical and psychometric technicalities since most of the stakeholders are not conversant with psychometric issues in the country

CHAPTER FIVE

CONCLUSIONS, IMPLICATIONS AND RECOMMENDATIONS

5.0 Chapter Overview

In this chapter, summary, conclusions, implications and recommendations of the study based on the findings to the research questions are made.

5.1 Summary

The major purpose of the study was to investigate the effect of adjusting cut scores across years on students' classification into grading categories at Malawi School Certificate of Education (MSCE) examinations. Specifically the study sought to find out whether adjusting cut scores in grading MSCE examinees in Mathematics across the years has any serious negative effects comparability of grades that examinees get in different forms of the examinations. The study also wanted to examine the comparability of scores and/or standards across different administrations of the examinations.

The finding that the 2006 test form was the most difficulty and had substantially lower cut scores than the other two forms had implications for Malawi's test developers. The different cut scores were expected because the two test forms cannot be exactly equal in

difficulty. But the differences between 2006 on one hand and 2005 and 2007 on the other were larger than had been expected.

The results of the study have shown that increase or decrease of cut scores across the years does not necessarily signify changes in performance of examinees or differences in ability levels of the cohorts but differences in the difficulty levels of the test forms. The results also illustrate that even though test developers attempt to construct test forms that are as similar as possible to one another in content and statistical specifications, the forms will still differ in difficulty. In short, the results of the study are consistent with previous researches (Chulu & Sireci, 2011; Khembo, 2004; Zoani, 1989).

5.2 Conclusions and Implications

The main purpose of the study was to investigate the effect of adjusting cut scores across the years (cohorts) on students' classification into grading categories at Malawi School Certificate of Education (MSCE) Examinations.

- It can be concluded that: the spiraling procedure which was employed in identifying the three cohorts worked effectively in coming up with randomly equivalent groups of examinees who wrote the test forms.
- The results showed different cut scores at all the grade categories of pass, credit and distinction for all the three forms of 2005, 2006 and 2007 except at distinction where the 2005 and 2007 forms produced similar cut scores of 70%.

- The differences in the cut scores showed some consistency at each grade category. This once again suggests that the judges were consistent and used the same criteria in coming up with the cut scores for all forms of the tests. It was observed that cut scores set for 2006 were lower than those set for 2005 and 2007.
- The different cut-off scores were expected because the three forms cannot be constructed to be exactly equal in difficulty. If two forms of the examinations are not identical in difficulty their scores are not equivalent nor are the decisions such as pass/fail that emanate from them. To ensure equivalence of such different tests, some form of statistical adjustment or equating should be applied.
- When raw cut scores are considered, the negative effects of keeping the cut scores unchanged across the years seem to be more than when they are changed. This means that as many as 9.89% of the examinees would have failed the 2006 while in fact they passed, if the cut scores were not changed from year to year. This is in agreement with the conclusion made by Zoani (1989) who stated that drastic changes in pass rates would be unacceptable and would bring about a number of questions about the validity of the examinations and quality of teaching (Zoani, 1989). Although the cut scores for the three forms of the examinations were different, they appeared to represent the same level of proficiency, hence grade.

5.3 Recommendations

From the findings of this study, a number of recommendations can be made when one considers the effect of changing cut scores across years on students' classification into grading categories. This can be done at all the stages of the examinations such as setting, moderation, administration, scoring and standard setting or award meeting.

1. Since the differences in the test difficulty has been found to be a critical aspect in determining the cut scores, test developers these are item writers and moderators should make sure that test items that they develop are similar in their tasks, cognitive demands and level of difficulty in order to minimize the difficulty differences.
2. MANEB can continue using the current system of varying the cut scores across the cohorts but strive to venture into equating and standardization as the current system lack psychometric backing.
3. Different cut scores on different test forms can mean the same performance level provided the grades represented by the cut scores are the same.

5.4 Suggestions for Further Research

In order to answer some related questions on the effect of changing cut scores across years and to contribute more knowledge to the topic, the following recommendations are made.

1. Doing a similar study but to include mathematics paper II also and not paper I only.

2. Comparing the cut scores changes from different subjects such as mathematics and biology or English.
3. Doing a similar study with more resources to cover a larger sample, across the country and to use more participants as judges.
4. Carrying out a similar study at other levels like the Primary School Leaving Certificate Examinations (PSLCE) and Junior Certificate Examinations (JCE).

REFERENCES

- Andew, B., & Hecht, J. (1976). Preliminary Investigation of Two Procedures for Setting examination Standards. *Education and Psychological Measurement* , 45-50.
- Bejar, I. I. (2008, October). Standard Setting: What is it? Why is it Important? *R & D Connections*. New Jersey, Princeton, United States of America: Education Testing Services.
- Bennett, J. (1998, November). *Board of Studies NSW 1998*. Retrieved September 16, 2010, from Board of Studies NSW : <http://www.boardofstudies.nsw.edu.au>
- Boulet, J. R., De Champlain, A., & McKinley, D. (2003). Setting Defensible Performance Standards on OSCEs and Standardized Patient Examinations. *Medical Teacher*, 25, 245-249.
- Chakwera, E. K. (2004, June 28). High-Stakes Testing in The Warm Heart Of Africa. The Challenges and Successes Of The Malawi National Examinations Analysis Archives. *Educational Policy Analysis Archives* , pp. 1-21.
- Chiunda, G. (2010). Standard Fixing Procedures as Done by The Malawi National Examinations Board: *Benchmarking Assessment Procedures* (pp. 111-114). Mangochi: unpublished Paper.
- Chulu, B. W., & Sireci, S. G. (2011). Importance Of Equating High- Stakes Educational Measurements. *International Journal Of Testing* , 38-52.
- Cizek, G. J. (1996). Setting Passing Scores. *Education Measurement Issues and Practice* , 20-31.

- Cohen, L., Manion, L., & Morrison, K. (2000). *Research Methods in Education*. London: Routledge Falmer.
- Cresswell, J. (2003). *Research Design: Qualitative, Quantitative and Mixed Methods Approaches* (2nd Edition ed.). London: Sage Publications.
- Creswell, J. W. (1994). *research Design: Qualitative, Quantitative Approaches*. Thousand Oaks, CA: Sage Publications.
- Crocker, L., & Algina, J. (1986). *Introduction To Classical Test Theory And Modern Test Theory*. New York: Holt, Rinehart, and Winston.
- Crocker, L., & Algina, L. (1986). *Introduction to Classical and Modern Test theory*. New York: Holt, Rinehart and Winston.
- Eckes, T. (2012). Examinee-Centered standard setting for large-scale assessments: The prototype group method. *Psychological Test and Assessment Modeling* , 54 (3), 257-283.
- Elder, S. (2009). *International Labour Organisation School-to-Work Transition Survey: A Methodological Guide; Module3*. Geneva: International Labour Organisation.
- Fraenkel, J., & Wallen, N. (1996). *How to Design and Evaluate Research* (2rd Edition ed.). New York: McGraw Hill, Inc.
- George, S., Haque, S. M., & Oyeboode, F. (2006, September 14). Standard Setting: A Comparison of Two Methods. 6 . Birmingham, United Kingdom. Retrieved October 13, 2011, from BioMed Central Ltd.: <http://creativecommons.org/licences/by/2.0>
- Glass, G. V. (1978). Standards and Criteria. *Journal of Educational Measurement* , 15 (4), 237-261.

Hambleton, R. K. (1978). On the Use of Cut-Off Scores With Criterion Referenced Tests in Instructional Settings. *Journal of Educational Measurement* , 277-290.

Hambleton, R. K. (2001). Setting Performance Standards on Educational Assessments and Criteria for Evaluating the Process. In G. Cizek, (Ed.). *Setting Performance Standards: Concepts, Methods and Perspectives* (pp. 89-116). Mahwah NJ: Erlbaum Publishers.

Hambleton, R. K., & Joness, R. W. (n.d). Comparison of Classical Test Theory and Item Response Theory and Their Applications to Test Development. In A. N. Module, (Ed.). *Instructional Topics in Educational Measurement (ITEMS)* (pp. 253-262). National Council on Education and Measurement in Education.

Hinkle, D. E., Wiersman, W., & Jurs, S. G. (1994). *Applied Statistics for the Behavioural Sciences*. Boston: Houghton Mifflin Company.

Khembo, D. J. (2004). *Using Performance Level Descriptors to Ensure Consistency and Comparability in Standard Setting*. Retrieved January 21, 2012, from www.fma.org/FMAOnline/Certifications.pdf:
<http://www.fma.org/FMAOnline/certifications>

Kline, Q. (2005, January 10). *Classical Test Theory*. Retrieved July 17, 2010, from http://www.sagepub.com/upm-data/4869_kline

Koffler, S. L. (1980). *A Comparison of Approaches for Setting Proficiency Standards*. Retrieved October 7, 2011, from JSTOR: <http://www.jstor.org/stable/143832>

- Kolen, M. J. (2007). Data Collection Designs and Linking Procedures. In N. j. Dorans, M. Pommerich, & P. W. Holland (Eds.), *Linking and Aligning Scores and Scales* (pp. 31-54). New York: Springer.
- Kolen, M., & Brennan, R. (2004). *Test Equating, Scaling and Linking: Methods and Practices* (2nd Edition ed.). New York: Springer Verlag.
- Livingston, S. A. (2004). *Equating Test Scores (Without IRT)*. Princetown NJ: Educational Testing Services, ETS.
- Norcin, J. J. (1990). Equivalent Pass/Fail Decisions. *Journal of Educational Measurement* , 59-66.
- Norcin, J. J. (2003). Setting Standards on Educational Tests. *Medical Education* , 464-469.
- Pell, G., & Roberts, T. E. (2006). Setting Standards for Student Assessment. *International Journal of Research & Methods in Education* , 91-103.
- Phillips, G. (1994). Methods and Issues in Setting Performance Standards.
In A.C. Tuijnman, & N. T. Postlethwaite (Eds.), *Monitoring The Standards of Education* (pp. 191-210). Washington: Pergamon.
- Pitoniak, M. J. (2010). Fundermentals of Standard Setting. *International Association of Educational Assessment*. Bangkok: Educational Testing Service.
- Pophan, J. B. (1978). As Always Provocative. *Journal of Educational Measurement*, 297-300.

- Sanju, G., Haque, S. M., & Oyeboade, F. (2006, September 14). *BMC Med Educ*. Retrieved October 13, 2011, from Creative commons Attribution License: <http://creativecommons.org/licenses/by/2.0>
- Sireci, S. G., Hambleton, R. K., & Pitoniak, M. J. (2004). *Setting Passing Scores on Licensure Examinations Using the Direct Consensus*. Center for Educational Assessment Research .
- Skakun, E. N., & Kling, s. (1980). Comparability of Methods for Setting Standards. *Journal of Educational Measurement* , 229-235.
- Yadidi, D. C. (2010, November 1). Enhancing Use of Qualitative Judgement to Determine Cut Scores. *A paper presented at the Joint MANEB and Chief/Senior examiners' Retreat* . Magochi: Unpublished paper.
- Yadidi, D., & Banda, A. C. (2009). Assessment in Multi-Cultural Society. *Association of Examination Associations in Africa (AEAA)*. Younde: Unpublished Paper.
- Zoani, A. (1989, October). *Application of Standard Setting Methods in Public*, Unpublished doctoral thesis, Monash University.

APPENDICES

APPENDIX A

2005 MALAWI SCHOOL CERTIFICATE OF EDUCATION EXAMINATION

MATHEMATICS

Subject Number: M131/I

Time Allowed: 2 hours

Tuesday, 18 October

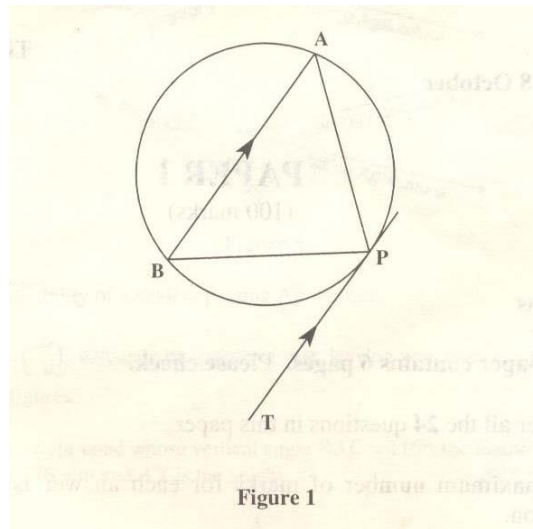
8:30 – 10:30 am

PAPER I (60 marks)

Instructions

1. **This paper contains 6 pages. Please check.**
2. Answer **all** the **24** questions in this paper.
3. The maximum number of marks for each answer is indicated against each question.
4. Mathematical tables and answer books are provided.
5. Used supplementary sheets must be handed in together with the answer book.
6. All working must be clearly shown; it should be done on the same sheet as the rest of the answers.
7. Use of electronic calculators is **not** allowed.
8. Write your **Examination Number** on top of each page of your Answer Book.

1. Factorise completely $3t^2 - 4t + 1$. (3 marks)
2. Without using a calculator or four-figure tables, simplify $\sqrt{125} + \sqrt{5} - \sqrt{45}$, leaving your answer in surd form. (3 marks)
3. In **Figure 1**, **TP** is a tangent to the circle **APB** at **P** and **AB** is parallel to **PT**.



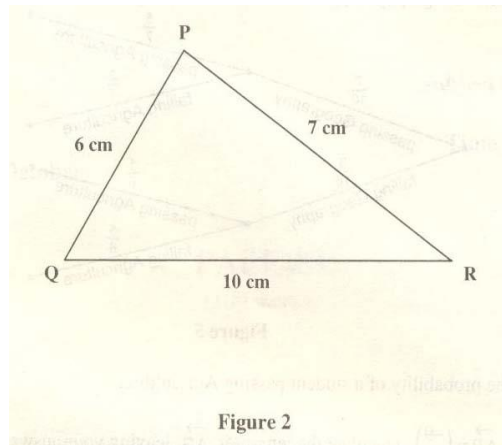
Prove that **AP = BP**.

(4 marks)

4. Simplify $\frac{x^2 - y^2}{x^2 - xy}$. (3 marks)
5. Given that $\{a, c, e, h, i, l, m, s, t, w\}$ is a universal set and $Y = \{a, c, e, h, i, m, s\}$, find $n(Y')$. (3 marks)
6. The function $y = 2 + x$ has the range $\{3, 6\}$. Find its domain. (3 marks)
7. A point $P(-2, 4)$ is translated to a point P' . If P' is 5 units down and 3 units to the right of point P , find the coordinates of P' . (3 marks)
8. Given that $\log_m 27 = 3$, find m . (3 marks)

Continued/...

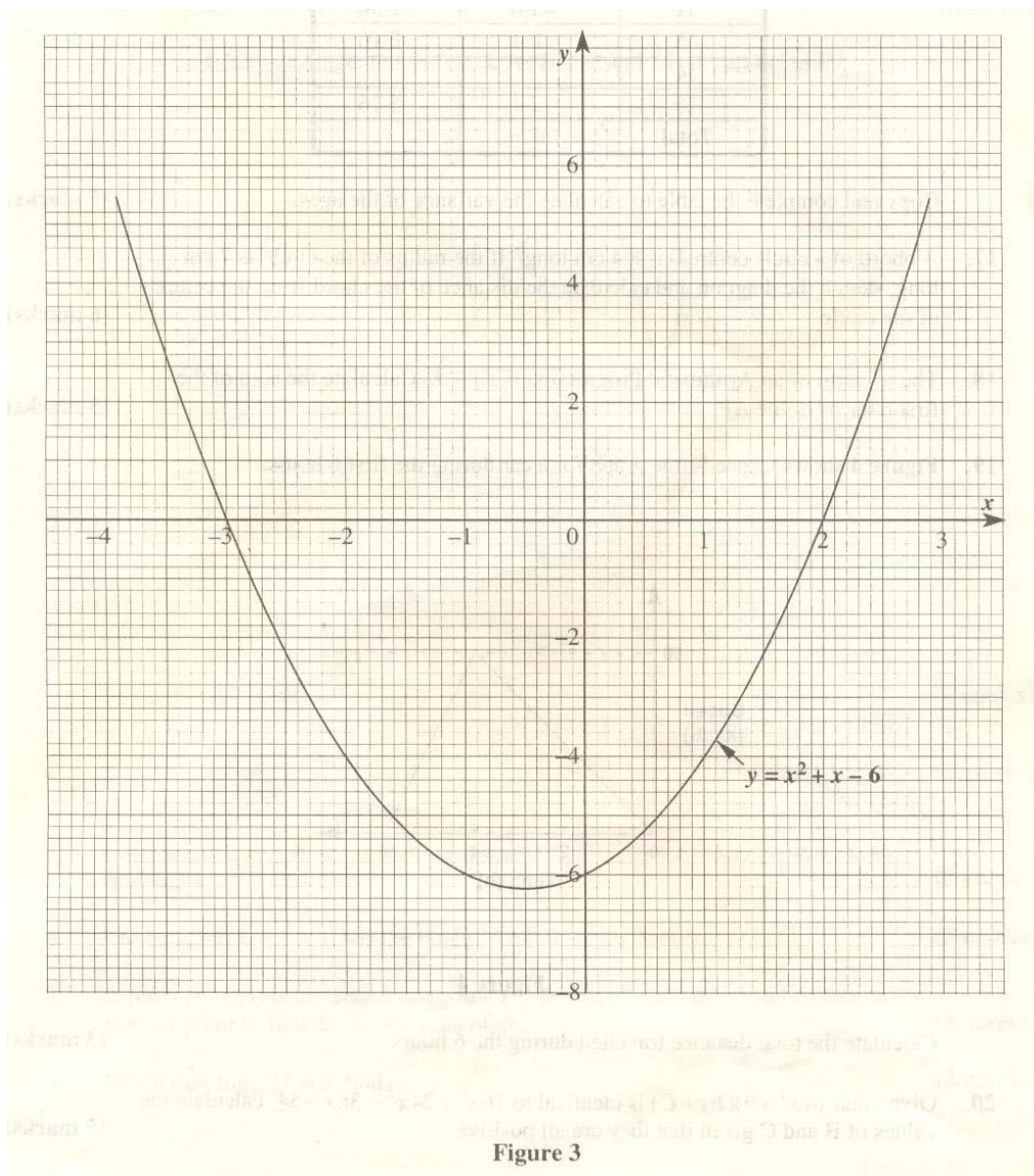
9. Make r the subject of the formula $S = \pi(2r)^2$. (4 marks)
10. In **Figure 2**, PQR is a triangle such that $PQ = 6$ cm, $QR = 10$ cm and $RP = 7$ cm.



- Calculate angle PRQ to the nearest degree. (5 marks)
11. A trapezium has a height of 3 cm and its area is 6 cm^2 . Calculate the area of a similar trapezium with a height of 12 cm. (4 marks)
12. Find the equation of a straight line passing through the point (0.7) and parallel to the line $y = 2x + 5$. (4 marks)
13. Given that P varies as a product of q and r^2 , and that $P = 50$ when $q = 1$ and $r = 5$, find P when $q = 3$ and $r = 8$. (5 marks)
14. A farmer is selling at most 70 chickens out of which less than 30 are hens. Using x to represent the number of hens and y to represent the number of cocks, write down **four** inequalities involving x and y . (6 marks)

Continued/...

15. **Figure 3** shows a graph of $y = x^2 + x - 6$.



Use the graph to solve the equation $x^2 = 5 - x$.

(4 marks)

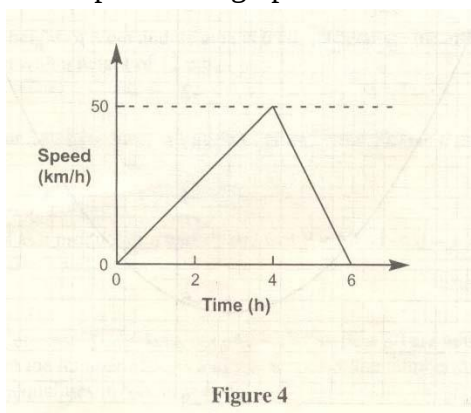
16. The **table below** shows ages of 5 pupils with the mean age of 12.6 years.

Age(yrs)	Deviation from mean	Square of deviation
10	-2.6	6.76
11	-1.6	2.56
13	0.4	0.16
14		
15	2.4	5.76
Total	0	

Copy and complete the table to calculate the variance of the ages. **(5 marks)**

17. A chord of a circle centre O is 8.4 cm long. If the radius of the circle is 7 cm long, sketch the diagram and calculate the distance of the chord from the centre of the circle. **(6 marks)**
18. The n^{th} term of an Arithmetic Progression is $5n - 3$. Calculate the sum of the first 6 terms of the AP. **(5 marks)**

19. **Figure 4** shows a speed-time graph for a car during the first 6 hours.



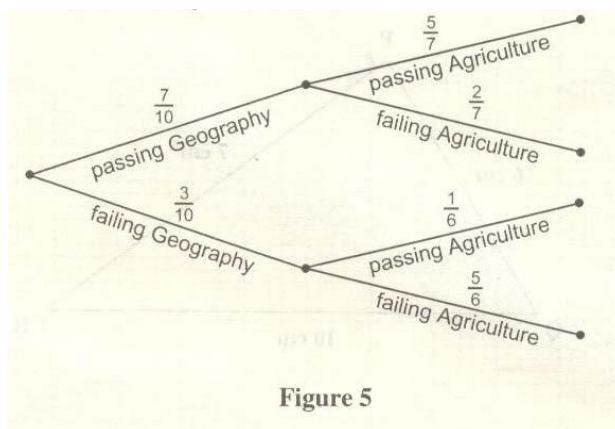
Calculate the total distance travelled during the 6 hours. **(3 marks)**

2005

M131/I

20. Given that $(4x^2 - 9)(Bx + C)$ is identical to $16x^3 + 24x^2 - 36x - 54$, calculate the values of B and C given that they are all positive. **(5 marks)**

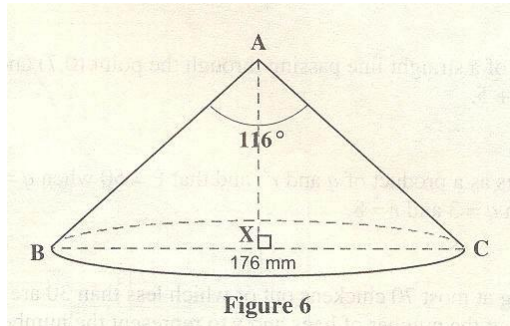
21. **Figure 5** is a tree diagram illustrating the probability of a student passing Agriculture and Geography in an examination. The probability of passing Geography in the examination is $\frac{7}{10}$ and the probability of passing Agriculture after one has passed Geography is $\frac{5}{7}$. The probability of passing Agriculture after one failed Geography is $\frac{1}{6}$.



Calculate the probability of a student passing Agriculture. **(5 marks)**

22. Given that $\vec{AB} = \begin{pmatrix} -9 \\ 4 \end{pmatrix}$ calculate the length of $\vec{AB}$, leaving your answer correct to 3 significant figures. **(4 marks)**

23. **Figure 6** shows a right-cone whose vertical angle $BAC = 116^\circ$, the diameter of its base $BC = 176$ mm and AX is the height.



Calculate the length of AX . **(5 marks)**

24. Solve the simultaneous equations

$$xy = -9,$$

$$y = x + 6.$$

(5 marks)

END OF QUESTION PAPER

NB: This paper contains 7 pages.

APPENDIX B

2006 MALAWI SCHOOL CERTIFICATE OF EDUCATION EXAMINATION

MATHEMATICS

Subject Number: M131/I

Time Allowed: 2 hours

Tuesday, 17 October

8:30 – 10:30 am

PAPER I

(100 marks)

Instructions

1. **This paper contains 6 pages. Please check.**
2. Answer **all** the **24** questions in this paper.
3. The maximum number of marks for each answer is indicated against each question.
4. Mathematical tables, graph paper and answer books are provided.
5. Calculators may be used.
6. Used graph paper and/or supplementary sheets must be handed in together with the answer book.
7. **All working must be clearly shown;** it should be done on the same sheet as the rest of the answers.
8. Write your **Examination Number** on top of each page of your Answer Book.

Answer **all** the **twenty-four** questions.

1. Simplify the following fraction:

$$\frac{3x+6}{(x-1)(x+2)}.$$

(3 marks)

2. Factorise completely $2x^2 + 4xy - 30y^2$.

(3 marks)

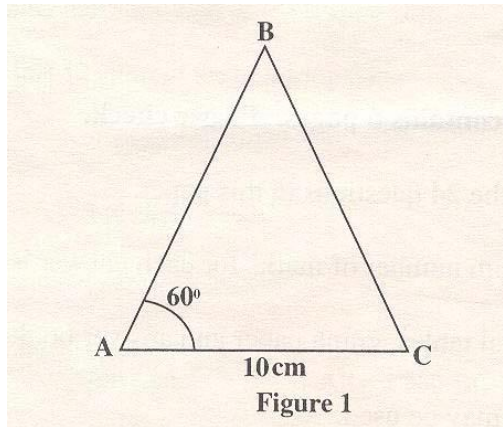
3. A cuboid is 76 cm long, 50 cm wide and 40 cm high. Calculate the volume of the cuboid.

(3 marks)

4. If $f(x) = 8^x - 6$, find $f\left(\frac{2}{3}\right)$

(4 marks)

5. **Figure 1** is a triangle **ABC** in which angle **BAC** = 60° , **AC** = 10 cm and the area of the triangle is $15\sqrt{6}$ cm².



Calculate the length of **AB** leaving the answer in its simplest surd form. (5 marks)

6. When the polynomial $ax^2 + bx + c$ is divided by $(x + 1)$ and $(x + 3)$ gives a remainder of 2 in each case. Find the polynomial.

(4 marks)

Continued/ ...

7. **Table 1** shows marks that student **A** and student **B** got from tests. Student **A**

sat for 5 tests while student **B** sat for 4 tests. Student **B** has mark x missing.

Table 1

STUDENT A	55	70	80	30	65
STUDENT B	67	60	x	53	

Given that the mean mark of student **A** is the same as the mean mark of student **B**, calculate the value of x . (5 marks)

8. John is twice as old as Mary. If the sum of the squares of their ages is 125, how old is Mary? (6 marks)
9. Without using a calculator or four-figure tables, find $\cos B$ if $\sin B = 0.8$. (6 marks)
10. Triangle **ABC** has vertices **A**(-1,2), **B**(3,3) and **C**(2,-1). Prove that angle **BAC** = angle **ACB**. (5 marks)
11. Find the values of x and y in the following matrix equation:

$$\begin{pmatrix} 6 & 3 \\ 2 & -3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 12 \\ -4 \end{pmatrix}$$
(5 marks)
12. **Table 2** shows the speed of a train recorded every 10 seconds.

Table 2

SPEED (m/s)	200	400	600	800
TIME (s)	10	20	30	40

Using a scale of 2 cm to represent 100 m/s on the vertical axis and 2 cm to represent 10 seconds on the horizontal axis, draw a speed-time graph and use

it to calculate the acceleration of the train.

(6 marks)

13. Given that $x \propto \frac{y}{z}$. When $x = 10$, $y = 2$ and $z = 4$. Find value of x when $y = 1$

and $z = 5$.

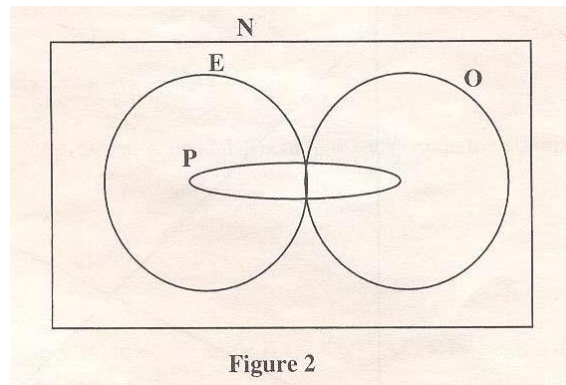
(5 marks)

14. Two lines **G** and **H** intersect at a point **P**. **G** passes through the points $(-4, 0)$ and $(0, 6)$. Given that **H** has the equation: $y = 4x - 4$, find, by calculation,

the coordinates of **P**.

(3 marks)

15. **Figure 2** shows a venn diagram representing set of all numbers (**N**), set of even numbers(**E**), set of odd numbers (**O**) and set of prime numbers (**P**).



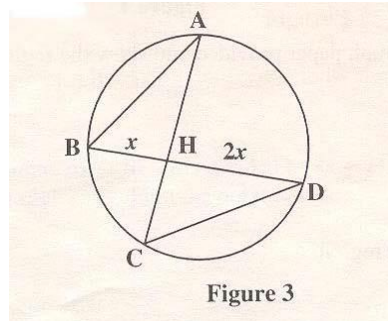
Copy the venn diagram and place the numbers $\frac{1}{2}$, 2, 6, 9 and 13 in the right places.

(6 marks)

16. Using a ruler and pair of compasses only, draw a line **AB** = 10 cm and construct a circle with the line **AB** as a diameter. Mark a point **C** on the circle such that **AC** = 6 cm. Join **AC** and **BC**. Construct a tangent **CP** such that angle **BCP** is acute.

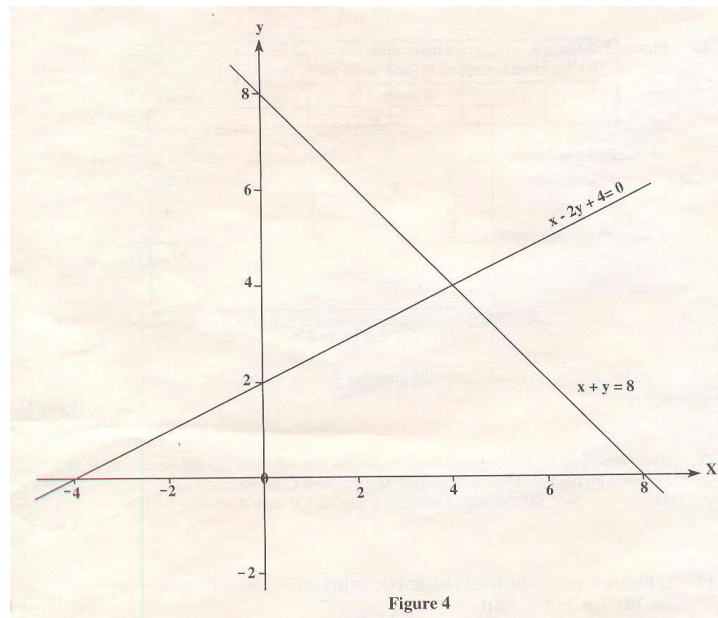
(5 marks)

17. In **Figure 3**, circle **ABCD** has chords **AC** and **BD** intersecting at **H**. Length of line **HD** is twice that of **BH**.



Find the ratio of areas of triangle **ABH** to triangle **CDH**. (4 marks)

18. **Figure 4** shows graphs of $x + y = 8$ and $x - 2y + 4 = 0$.



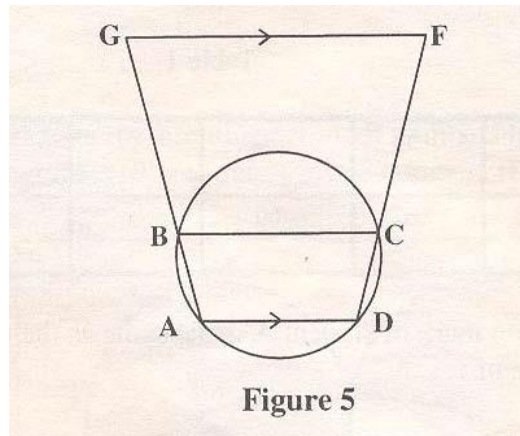
Copy the figure on the graph paper provided and show the region bounded by the following inequalities;

$$\begin{aligned} x &\geq 0, \\ x + y &\leq 8, \\ x - 2y + 4 &\geq 0, \end{aligned}$$

by shading the unwanted region.

(6 marks)

19. **Figure 5** shows a circle **ABCD** in which **AB** and **DC** are produced to **G** and **F** respectively such that **GF** is parallel to **AD**.



Prove that quadrilateral **GBCF** is cyclic. (4 marks)

20. Given that $\log_{10} n - \log_{10} m = 2 \log_{10} h$, show that $n = mh^2$. (4 marks)

21. Coin **A** is tossed followed by coin **B**. The probability that coin **A** shows head is $\frac{1}{2}$ while the probability that coin **B** shows head is $\frac{1}{4}$. Using a tree diagram, calculate the probability that both coins **A** and **B** show tails. (4 marks)

22. Find the sum of the first 12 terms of the following **GP**:

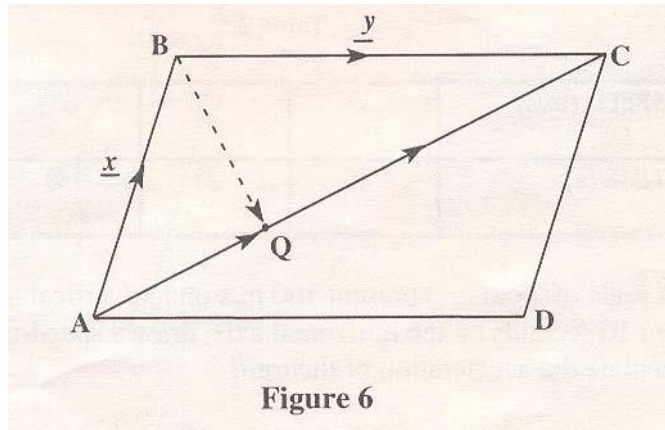
$$\frac{1}{2187}, \frac{1}{729}, \frac{1}{243}, \dots$$

Give your answer correct to 2 decimal places. (5 marks)

23. Simplify $\frac{x^2 - 2}{x - \sqrt{2}}$. (3 marks)

Continued/...

24. **Figure 6** shows a parallelogram **ABCD** in which $\vec{AB} = x$ and $\vec{BC} = y$.



If $\vec{AQ} = \frac{1}{4} \vec{AC}$, find $\vec{BQ}$ in terms of $\underline{x}$ and $\underline{y}$.
(5 marks)

END OF QUESTION PAPER

NB: This paper contains 6 pages.

APPENDIX C

2007 MALAWI SCHOOL CERTIFICATE OF EDUCATION EXAMINATION

MATHEMATICS

Wednesday, 15 October

Subject Number: M131/I

**Time Allowed: 2 hours
8:30 – 10:30 am**

PAPER I

(100 marks)

Instructions

- 1. This paper contains 6 pages. Please check.**
- 2. Answer all the 24 questions in this paper.**
- 3. The maximum number of marks for each answer is indicated against each question.**
- 4. Mathematical tables and answer books are provided.**
- 5. Calculators may be used.**
- 6. Used supplementary sheets must be handed in together with the answer book.**
- 7. All working must be clearly shown;** it should be done on the same sheet as the rest of the answer.
- 8. Read the instruction(s) on the Answer Book carefully.**
- 9. Write your Examination Number at the top of each page of your Answer Book.**

Answer **all** the **twenty four** questions in this **paper**.

1. Express $\frac{\sqrt{2}}{\sqrt{3}}$ with a rational denominator. (3 marks)
2. Given that $a = \begin{pmatrix} 2 \\ -4 \end{pmatrix}$ and $b = \begin{pmatrix} -4 \\ 0 \end{pmatrix}$ find $\frac{1}{2}(b - a)$. (4 marks)
3. Factorise completely $2x^2 - 4x - 126$. (3 marks)
4. The function $f(x) = \frac{1}{3x-1}$. Given that $\{-1, 0, 2\}$ is the domain, find the range. (4 marks)
5. Express b in terms of a and c in the formula $c = ab - \frac{b}{a}$. (4 marks)
6. **Figure 1** shows a tangent to a circle **ABC** with centre **O**. Line **CT** is parallel

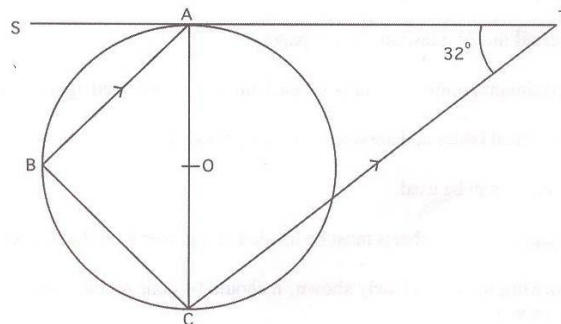


Figure 1

to **BA** and angle **ATC** = 32° .

Calculate angle **ACB**.

(3 marks)

7. The gradient of a straight line passing through point $P(-2, 5)$ is $-\frac{1}{2}$. Find the equation of the line in the form $y = mx + c$. (4 marks)

Continued/...

8. **Figure 2** is a venn diagram showing the number of elements in sets **R**, **S** and universal set ξ

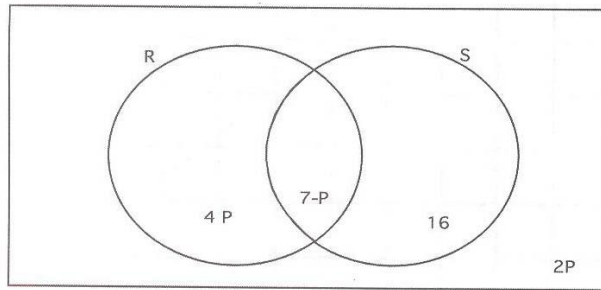


Figure 2

- If $n(R \cup S) = 29$, calculate the value of p . (4 marks)
9. Given that $A = \begin{bmatrix} 2 & 6 \\ 1 & -4 \end{bmatrix}$, $B = \begin{bmatrix} 3 & 2 \\ 0 & -5 \end{bmatrix}$ and $C = \begin{bmatrix} 1 & 8 \\ -5 & 7 \end{bmatrix}$ simplify $\frac{1}{4}(A - B + C)$ (6 marks)
10. Simplify $\frac{1}{a-b} + \frac{1}{a+b}$. (4 marks)
11. Given that $\log_a 2 = 0.668$ and $\log_a 3 = 0.884$, evaluate $\log_a 12$. (5 marks)
12. P varies directly as x^3 and inversely as y . When $x = 2$ and $y = 4$, $P = 3$.
Find the value of x when $P = 12$ and $y = 4$. (6 marks)
13. When the polynomial $x^3 + 5x^2 + Kx + 3$ is divided by $(x + 2)$ it gives a remainder of 1. Find the value of K . (5 marks)
14. The fourth term of an Arithmetic progression is 11 and the seventh term is 20.
Calculate the first term. (5 marks)
15. Solve the simultaneous equations:
$$y = x + 2$$
$$x^2 - xy = 4.$$
 (5 marks)

16. **Figure 3** shows the image P' of P after translation.

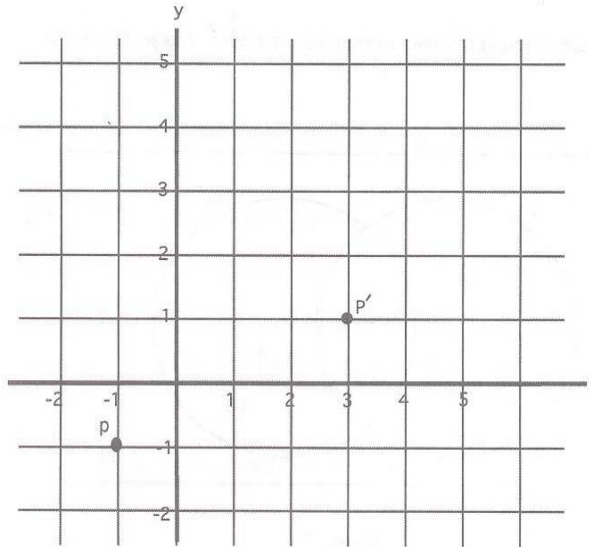


Figure 3

Find the translation vector that maps P into P' .

(3 marks)

17. **Figure 4**, shows the region R bounded by three inequalities.

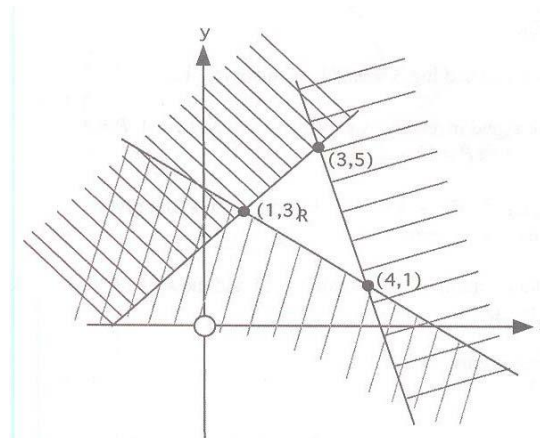


Figure 4

Calculate the maximum value of $5x - 4y + 8$ in this region.

(3 marks)

18. Calculate the total surface area of a solid hemisphere of radius 21 cm.

(Area of a sphere = $4\pi r^2$; Take $\pi = \frac{22}{7}$)

(5 marks)

19. Figure 5 shows the speed time graph of a moving object.

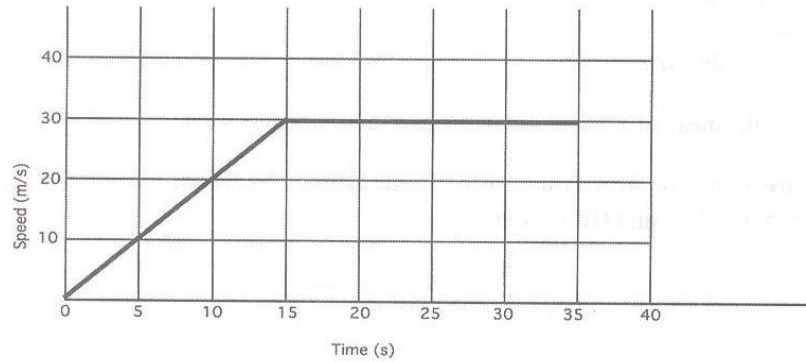


Figure 5

Use the graph to find the total distance travelled by the object in the first 35 seconds.

(3 marks)

20. Figure 6 shows region G bounded by inequalities.

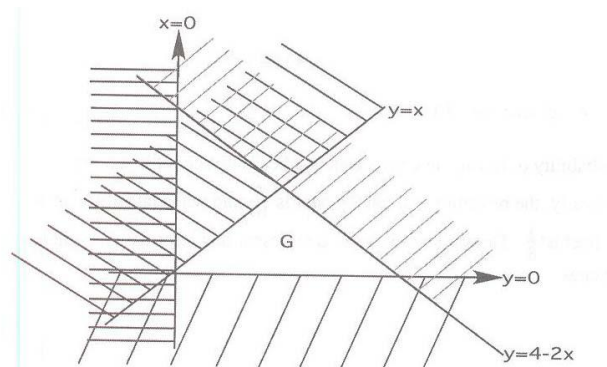


Figure 6

Write the **three** inequalities that describe the region G. (4 marks)

Continued/...

21. The angle of depression of a car from the top of a pole is 35° . If the top of the pole is 25 m from the ground, calculate the distance of the car from the pole. **(6 marks)**
22. Three data values x , y , and z have the following relationship:
- $$x = a^2 - a$$
- $$y = 2 - a$$
- $$z = 7 + 5a - a^2.$$

Calculate the mean of x , y and z in terms of a in its simplest form. **(3 marks)**

23. In **Figure 7**, triangle **ABC** is similar to triangle **DBA**. The area of triangle **DBA** is 24 cm^2 , **AB** = 8 cm and **DB** = 4 cm.

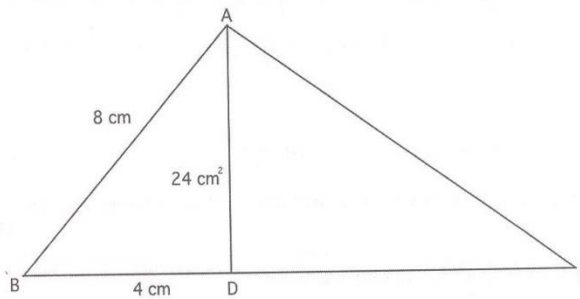


Figure 7

Calculate the area of triangle ABC. **(4 marks)**

24. The probability of having an early lunch at a boarding school is $\frac{2}{3}$. When the lunch is early, the probability of having beef is $\frac{7}{10}$ and when late, the probability of having beef is $\frac{1}{8}$. Draw a tree diagram to represent this information completing all branches. **(4 marks)**

END OF QUESTION PAPER

NB: This paper contains 6 pages.

APPENDIX D

REQUEST LETTER TO MALAWI NATIONAL EXAMINATION BOARD (MANEB)

University of Malawi
Chancellor College
Faculty of Education
Department of Education Foundation
P.O. Box 280
Zomba.

The Acting Executive Director
Malawi National Examinations Board
P. O. Box 191
Zomba

Dear Sir,

REQUEST FOR DATA AND PERMISSION TO USE PAST PAPERS

I write to request for examination data that I need to use for my dissertation which is:
Effect of changing cut scores across cohorts in standard setting.

The data that is needed is as follows:

- MSCE Mathematics Paper I and marking schemes for 2005, 2006 and 2007.
- Percentages of candidates at pass, credit and distinction categories for Mathematics paper I for 2005, 2006 and 2007.

I would also like to seek your permission, sir to re – administer the 2005, 2006 and 2007 MSCE Mathematics paper I to current form fours in selected secondary schools in South East Education Division (SEED) and South West Education Division SWED.

I am a student currently pursuing a Master's Degree course in Testing, Measurement and Evaluation at Chancellor College.

The research is for academic purposes only and any information will be treated with utmost confidentiality. Ethically, no names will be required.

May I thank you in advance.

Yours faithfully,

B. K. Pindani

APPENDIX E

REQUEST LETTER TO EDUCATION DIVISION MANAGERS (EDM)

University of Malawi
Chancellor College
Faculty of Education
Department of Education Foundation
P.O. Box 280
Zomba.
13th May, 2012

The Division Manager
South Eastern Education Division
Private Bag 48
Zomba

Dear Sir/Madam,

REQUEST TO CONDUCT RESEARCH

I am a student currently pursuing a Master's Degree course in Testing, Measurement and Evaluation at Chancellor College.

I therefore write to request for permission to conduct an academic research in some schools in your Division that is planned to take place sometime in May. The research is on the Effect of Changing Cut Scores across Cohorts at Malawi School Certificate of Education Examinations (MSCE). The targeted schools are Malosa, Mangochi, St. Monica Girls, Lisumbwi, Balaka, Bakhita Girls, Likangala, Domasi Demonstration, Chilunga CDSS, Sacred Heart CDSS and Mulunguzi Secondary Schools.

The research is for academic purposes only and any information will be treated with utmost confidentiality. Ethically, no names will be required or used unless seriously required in which case consent from the owner shall be sought.

May I thank you in advance.

Yours faithfully,

B.K. Pindani

APPENDIX F
REQUEST LETTER TO SCHOOLS

University of Malawi
Chancellor College
Faculty of Education
Department of Education Foundation
P.O. Box 280
Zomba.

10th May, 2012

The Head teacher

.....
.....
.....

Dear Sir,

REQUEST FOR PERMISSION TO ADMINISTER A TEST

I write to request for permission from your office to administer a mathematics test in your secondary school.

I am a student currently pursuing a Masters Degree course in Testing, Measurement and Evaluation at Chancellor College. The test is the main instrument for collecting data for my study which is Effect of Changing Cut Scores in standard setting at Malawi School Certificate of Education (MSCE) examinations.

The research is for academic purposes only and any information will be treated with utmost confidentiality. Ethically, no names will be required or used.

I will be grateful for your assistance in my request.

Yours faithfully,

Bitmon Kachingwe Pindani

APPENDIX G

Table 4.3

Percentage Cumulative Frequencies for Each Test Form

Raw Score	Cumulative Frequency %			Raw Score	Cumulative Frequencies %			Raw Score	Cumulative Frequencies %		
	2005	2006	2007		2005	2006	2007		2005	2006	2007
0	0.9	0	1	34	48.1	59.8	43.8	68	81.7	92	82.8
1	1.9	1	1.7	35	48.8	61	45.5	69	82.2	92.7	83.9
2	3.1	1.7	3.5	36	49.7	62.4	46.5	70	83.4	93	84.9
3	5.6	4.2	4.9	37	50.5	63.2	47.4	71	84.1	93.4	85.6
4	7	5.1	6.4	38	51.7	64.5	47.9	72	85.2	94.1	86.1
5	9.8	6.3	8.7	39	52.8	65.7	49	73	86.6	94.1	87.2
6	12.4	9.6	9.7	40	53.7	66.7	51.9	74	87.5	94.6	87.3
7	13.9	11.5	11.5	41	54.2	67.9	53.1	75	88.2	95.3	87.8
8	16.4	13.6	13.2	42	55.9	69	54.2	76	88.5	95.3	88.7
9	17.1	15.9	13.7	43	57.1	69.5	54.7	77	89.2	95.8	89.6
10	18.6	17.9	14.6	44	58.7	70.6	55.7	78	90.1	96	91
11	20.6	19	15.6	45	60.1	72	56.6	79	91.1	96	91.8
12	22	21.1	17	46	60.8	74	56.9	80	92.2	96.5	92
13	23.3	22.6	17.9	47	61.8	74.6	58.3	81	93.6	97	92.7
14	24.9	24.7	19.1	48	62.7	75.6	59.9	82	94.4	97.4	94.1
15	26.8	26.7	20.8	49	63.2	76.3	60.9	83	95.8	97.6	94.6
16	28.4	28.6	22.2	50	64.5	76.8	62.2	84	96.2	98.1	95.3

17	29.4	30.7	23.6	51	65.5	78.2	64.2	85	96.5	98.1	95.7
18	30.3	32.4	25.2	52	67.1	78.7	65.3	86	97.6	98.6	96.4
19	30.8	34.3	26	53	68.6	79.6	65.8	87	97.7	98.6	96.5
20	32.9	35.5	27.4	54	69.9	80.7	67.2	88	98.4	99	97.4
21	33.8	37.1	29	55	70.6	82.4	68.1	89	98.6	99.1	98.3
22	35.2	38.5	30.4	56	71.3	83.8	68.9	90	99	99.3	98.6
23	36.1	39.7	32.3	57	72.5	84.1	70	91	99.3	99.3	98.6
24	37.5	41.6	33	58	73.7	85	70.3	92	99.7	99.7	98.8
25	38.3	43.4	33.7	59	75.3	85.4	71.5	93	99.7	99.7	99.3
26	40.2	45.6	35.2	60	76.1	86.2	72.7	94	99.7	99.7	99.5
27	40.8	47	36.3	61	76.8	87.1	74.5	95	99.8	99.8	99.8
28	42.2	48.4	37.3	62	76.8	88.3	75.9	96	100	100	100
29	43.2	50.3	38	63	77.7	89.2	76.2	97			100
30	43.9	51.7	39.2	64	78	89.9	77.4	98			100
31	44.4	54	40.8	65	79.1	90.8	79.3	99			100
32	45.8	55.7	41.8	66	80.1	91.5	80.6	100			100
33	46.7	57.7	42.9	67	81	91.6	81.9				

APPENDIX H

PICTURES OF CANDIDATES TAKING THE TEST





APPENDIX I

MSCE MATHEMATICS SUMMARY OF CONTENT AREA

1. A. Quadratic Expressions and Equations

- Factorise quadratic expressions
- Calculate roots of quadratic expressions
- Complete the square of a quadratic expression
- Calculate roots of quadratic equations by completing the square
- Calculate roots of quadratic equations by formula
- Formulate quadratic equations given roots
- Formulate quadratic equations from word problems
- Solve quadratic equations involving word problems

B. Simultaneous Linear and Quadratic Equations

- Calculate the solutions of simultaneous linear and quadratic equations by substitution

C. Exponential and Logarithmic Equations (Functions)

- Solve exponential equations
- Formulate an exponential equation from logarithmic equation and vice versa
- Evaluate logarithms of numbers to given base
- State the rules of logarithms
- Solve logarithmic equations
- Apply the rules of logarithms in computations

D. Change of the subject of the Formula

- Change subject of a formula (i.e literal equation, equations involving powers roots or logarithms)

E. Polynomials

- State the degree of any polynomial
- Divide a polynomial of higher degree by a polynomial of lower degree
- Find the remainder using the remainder theorem
- Factorize polynomials of third degree
- Find the roots of polynomial equations of third degree
- Find polynomial co-efficients in identical polynomials

F. Mapping and Functions

- Write functions in different forms calculate the range given the domain

- Calculate the domain given the range
- Draw arrow diagrams

2. A. Chord Properties of a Circle

- Identify chords, arcs segments and sectors
- State chord properties of a circle
- Illustrate the chord properties
- Solve problems by applying chord properties

B. Angle Properties of a Circle

- Identify angles subtended at the centre and at the circumference of a circle by the same arc or chord
- State angle properties of a circle
- Illustrate angle properties of a circle
- Solve problems by applying angle properties of a circle
- Describe a cyclic quadrilateral
- Show that the quadrilateral is cyclic
- Show that points are concyclic using angle properties of a circle

C. Tangent to a Circle

- Define a tangent to a circle
- Deduce from measurement that a tangent is perpendicular to the radius at the point of contact
- Show that tangents to a circle from an external point are equal
- Illustrate that if two circles touch externally or internally, the point of contact lies on the straight line through the centres
- Identify angles in alternate segments
- Illustrate that angles in alternate segments are equal
- Apply the principle in solving the problems
- Construct a tangent to a circle
- Construct tangents from an external point

3. A. Irrational Numbers

- Describe an irrational number
- Simplify surds
- Add surds
- Subtract surds
- Multiply surds
- Divide surds
- Write conjugate surds
- Rationalize surd denominators

B. Algebraic Fractions

- Simplify fractions to lowest terms
- Add, subtract, multiply and divide fractions
- Solve equations involving fractions
- Formulate and solve fractional equations from word problems

4. A. Inequalities

- Sketch the graphs of regions described by the inequalities in one variable
- Find inequalities in one variable that describe a given region
- Illustrate graphically the solution of simultaneous inequalities in one variable
- Find simultaneous linear inequalities in one variable that describe a region
- Sketch the graph of a linear inequality in two variables
- Illustrate graphically the solution of a simultaneous linear inequalities in two variables

B. Linear Programming

- Identify variables
- Formulate inequalities
- Formulate objective functions
- Illustrate graphically the region described by the inequalities (shading the unwanted region)
- Find solutions of linear programming problems using the graph and the objective function

C. Travel graphs

- Draw speed-time graphs
- Use speed-time graphs to
 - Find acceleration by calculating the gradient of a line
 - Find deceleration by calculating the gradient of a line
 - Calculate the area under speed-time graph
 - Interpret that the area under speed-time graph is equal to the distance covered
 - Calculate the distance traveled by using average speed $\times$ time

D. Graphs of Functions – Linear, Quadratic and Cubic

- Construct tables of values of a quadratic function
- Draw graphs of quadratic functions

- Find maximum or minimum value of a quadratic function
- Find the equation of the line of symmetry
- Solve quadratic equations graphically
- Formulate the graphic equation give a quadratic graph which cuts the $x - \text{axis}$
- Solve simultaneous, linear and quadratic equations given graphically
- Construct tables of values of cubic functions
- Draw graphs of cubic functions
- Solve cubic functions graphically
- Solve the simultaneous linear and cubic functions graphically

5. A. Statistics

- Calculate the mean of data
- Calculate variance of data
- Calculate the standard deviation of ungrouped data
- Determine class boundaries
- Group data
- Find frequency of a class interval
- Draw a histogram
- Calculate the class centres of class intervals
- Draw frequency polygons

B. Probability

- Define probability space
- Construct a probability space table
- Solve probability problems using probability space
- Determine experimental probability events
- Construct a tree diagram
- Calculate probability of an event using at tree diagram

C. Sets

- Describe a universal set
- Identify elements of a complement of a set
- List elements of a complement of a set
- List down elements in a union of two or three sets
- Find the number of elements in a union of two or three sets
- List down elements in intersection of two or three sets
- Find the number of elements in intersection of two or three sets
- Illustrate union of sets in a venn diagram
- Illustrate intersection of sets in a venn diagram
- Solve problems using venn diagrams

6. A. Vectors

- Present magnitude in different notations
- Calculate the magnitude of a vector
- Describe zero / null vector
- Describe a position vector
- Find a position vector
- Identify parallel vectors
- Find the mid- point of a vector
- Show that points are collinear using vector method
- Add vectors using parallelogram law
- Solve problems by applying a parallelogram law

B. Matrices

- Present information in a matrix form
- Identify the order of matrix
- Identify square matrix
- Identify row matrix
- Identify column matrix
- Identify zero matrix
- Locate elements of a matrix
- Add matrices together when possible
- Subtract matrices when possible
- Multiply a matrix by a scalar number
- Multiply a matrix by another matrix of order 2

C. Coordinate Geometry

- Calculate the length of a straight segment
- Find gradient of a straight line given two points
- Calculate the gradient of a straight line from the graph
- Relate gradient to a tangent of an angle
- Finding the equation of a line with a given gradient through a given point
- Write down an equation of a line passing through two given points
- Find the equation of a line from the graph
- State the relationship between the gradient of parallel lines
- Find the equation of a line through a give point and parallel to a given line
-

7. A. Transformation

- Describe a transformation
- Give examples of transformations
- Describe a translation

- Draw a translation
- Write down coordinates of a translation in column vector
- Describe an enlargement
- Find the centre of an enlargement

B. Similarity

- Calculate area of similar shapes
- Calculate ratios of areas of similar shapes (area factor)
- State the principle of area factor
- Find lengths of sides of similar shapes using area factor

8. A. Trigonometry

- Define sine of an angle
- Define cosine of an angle
- Define tangent of an angle
- Calculate the three ratios of given angles within the range of 0 to 360 degrees
- Find angles when given ratios
- Solving right-angled triangles using trigonometric ratios
- Derive the fractional trigonometric ratios of 30, 45, 60 and 90 degrees
- Define an angle of elevation
- Define an angle of depression
- Sketch angles of elevation and depression
- Calculate an angle of depression
- Calculate an angle of elevation
- Calculate bearing of a point relative to a given point
- Calculate a side of a triangle using cosine rule
- Calculate an angle of a triangle using a sine rule
- calculate a side of a triangle using a sine rule
- calculate an angle of a triangle using a cosine rule
- solve problems using sine / cosine rules
- sketch bearing of a point
- calculate bearing of a point using sine /cosine rule
- calculate area of a triangle using area rule
- Calculate the angles of a triangle using area rule.

B. 3 – Dimensional Figures

- Sketch three dimensional figures
- Find surface area of 3-D figures
- Find volume of 3-D figures
- Identify vertical, horizontal and slanting lines (edges) and planes
- Identify angles between two planes

- Identify angles between a plane and a line
- Calculate lengths of sides in 3-D figures
- Calculate angles in 3-d figures

9. A. Variation

- Formulate mathematical equations involving joint variation
- Calculate constant of a joint variation
- Solve problems involving joint variation
- Formulate mathematical equations involving partial variation
- Calculate constant of a partial variation
- Solve problems involving partial variation
- Sketch a graph of partial variation

B. Progression

- Recognize an arithmetic progression (AP)
- Calculate common difference of an AP
- Calculate the n^{th} term of an AP
- Recognize the general form of an AP
- Use the formula for the n^{th} term to calculate the common difference and the number of terms
- Calculate the sum of terms of an AP using the formula
- Recognize the Geometrical Progression (GP)
- Use the formula for the n^{th} term to calculate the common ratio and the number of terms
- Use the formula to calculate the sum of terms of a GP
- Solve real life problems involving GPs.