

**ASSESSING THE ITEM RESPONSE THEORY KEY PARAMETERS OF 2020
MALAWI SCHOOL CERTIFICATE OF EDUCATION (MSCE) ENGLISH
PAPER I**

**MASTER OF EDUCATION (TESTING, MEASUREMENT AND EVALUATION)
THESIS**

CHRISTINA MUKOKO

UNIVERSITY OF MALAWI

OCTOBER, 2024



**ASSESSING THE ITEM RESPONSE THEORY KEY PARAMETERS OF 2020
MALAWI SCHOOL CERTIFICATE OF EDUCATION (MSCE) ENGLISH
PAPER I**

**MASTER OF EDUCATION (TESTING, MEASUREMENT AND EVALUATION)
THESIS**

CHRISTINA MUKOKO

Bachelor of Education (Agriculture) - University of Malawi, Bunda College

Submitted to the Department of Education Foundations, School of Education, in Partial
fulfillment of the requirements for the degree of Master of Education (Testing,
Measurement and Evaluation)

UNIVERSITY OF MALAWI

OCTOBER, 2024

DECLARATION

I, the undersigned declare, that this thesis is solely my original work and has never been submitted to any other institution for similar purposes. Where other people's work has been used, acknowledgements have been made accordingly. I hold the responsibility for the contents of this paper.

CHISTINA MUKOKO

Full Legal Name

Signature

Date

CERTIFICATE OF APPROVAL

The undersigned certifies that the thesis represents the student's own work and effort and has been submitted with our approval.

Signature: _____ **Date:** _____

Winiko Symon, PhD (Senior Lecturer)

Supervisor

Signature: _____ **Date:** _____

Yohane Chakasika, (Lecturer)

Postgraduate Coordinator (Education Foundations)

DEDICATION

I dedicate this work to Almighty God who has been my source of strength, grace and wisdom, to my husband Felix Kakowa, my daughter Thandie and my son Sean for the tireless support rendered during my academic sojourn.

ACKNOWLEDGEMENTS

First and foremost, I have to thank my research supervisor, Dr. S. Winiko. Without his assistance and involvement in every step throughout the process, this paper would have never been accomplished.

Getting through my thesis required more academic support and I would like to thank my six classmates for their support, for listening to and, at times, having to tolerate me over the past years of my study.

To my family, my sisters, I say thank you for your prayers, encouragement and understanding. Remain blessed.

I also thank everyone who rendered a hand to make this study a success, I appreciate you all.

ABSTRACT

The study examined the quality of 2020 MSCE English Paper I, specifically in relation to the three key parameters of Item Response Theory (IRT) by computing and analyzing the item-difficulty parameter, item-discrimination parameter and pseudo-guessing parameter of the test items using the 3 Parametric Logistic Model. In 2020 MSCE examinations, candidates did not perform well in English and subsequently the overall passing rate was also very poor. Social economic factors were attributed to the poor pass rate. Scholars contend that items of poor difficult level, as well as those with a low discrimination index also affects examination results. It was therefore, worthy conducting this study, since these studies help to flag poor quality items. In this study, a descriptive research design and quantitative research approach were employed. The study systematically and randomly sampled Form 4 students. The instrument had dichotomously scored items only. Subject matter experts scored the scripts. The findings indicate that 23% of the items had all three key parameters of IRT while 77% of the items were of poor quality and to some extent contributed to the poor pass rate in 2020. The study recommends the practice of item analysis as a way to enhance examination quality.

Keywords: Item discrimination, Item difficulty level, pseudo-guessing, 3 Parametric Logistic Model, Item response theory

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 TO STUDY	1
1.1 Chapter overview	1
1.2 Introduction	1
1.3 Problem statement.....	3
1.4 Justification for the choice of English.....	6
1.5 Purpose of study	6
1.6 Research Objectives	6
1.6.1 Main Objective	6
1.6.2 Specific objectives	7
1.7 Research questions	7
1.7.1 Main Research Question.....	7
1.7.2 Specific Research Questions	7
1.8 Significance of the Study	7
1.9 Operational Definition of terms	8
1.10 Study Limitations	9
1.11 Chapter Summary.....	10
CHAPTER TWO	11
REVIEW OF RELATED LITERATURE	11
2.1 Chapter Overview	11

2.2 English as a Subject.....	11
2.3 A Test	11
2.4 Key parameters of IRT	12
2.4.1 Item discrimination.....	12
2.4.2 Item difficulty level	13
2.4.3 Pseudo guessing.....	18
2.5. The three-parameter logistic model (3 PLM).....	19
2.6. Item Characteristics Curve (ICC).....	19
2.7 Item analysis.....	22
2.8 Theoretical Framework	24
2.9 Chapter Summary.....	26
CHAPTER 3	27
METHODOLOGY	27
3.1 Chapter overview	27
3.2 Research Paradigm.....	27
3.3 Research design and methodology	27
3.3 Sample size.....	28
3.4 Sampling strategy	30
3.5 Data Collection Techniques	30
3.6 Data Analysis	31
3.7 Ethical Consideration	32
3.8 Reliability and validity of data collection instrument	32
3.9 Data Management Plan	33
3.10 Research dissemination strategy	34
3.11 Research risks.....	34
3.12 Foreseeable and unforeseeable circumstances	35
CHAPTER 4.....	36
RESULTS AND DISCUSSION OF THE STUDY.....	36
4.1 Chapter overview	36
4.2 Results	36
4.2.1 Checking for IRT assumptions	36
4.2.2 Assessing IRT key parameters.	40

4.2.3. Interpretation of the indices in relation to b-parameter	44
4.3 Interpretation of c-parameter in relation to Item Characteristic Curves	48
4.3.1 Interpretation of the Pseudo Guessing Parameter in Relation to the Table 14. 51	
4.4 General discussion of the results	52
4.4.1 Items with all the key parameters of IRT	54
4.5 Chapter summary	55
CHAPTER 5	56
SUMMARY, CONCLUSIONS AND IMPLICATIONS.	56
5.1. Chapter overview	56
5.2 Summary of results.....	56
5.2.1 To what extent in terms of a parameters does the MSCE 2020 English paper 1 had? 56	
5.2.2 To what extent in terms of b parameters does the MSCE 2020 English paper 1 had?	57
5.2.3 To what extent in terms of c parameters does the MSCE 2020 English paper 1 had? 57	
5.3. Implication of findings	58
5.4 Conclusion and Future Directions for Research	59
5.5 Chapter summary	61
REFERENCES.....	62
APPENDICES	66

LIST OF FIGURES

Figure 1: Graphical representation of English examination pass rate	4
Figure 2: ICC for an average item	21
Figure 3: ICC for an easy item.....	21
Figure 4: ICC for a difficult item.....	22
Figure 5: Scree Plot.....	39
Figure 6: Graphical representation of a- parameters values in percentages. <i>Source:</i> <i>researcher</i>	43
Figure 7: General representation of a- parameter values in percentage <i>Source: researcher</i>	44
Figure 8: Graphical presentation of the item difficult levels in percentages related to the item difficult level ranges	47
Figure 9: ICC's for item number 1-10	49
Figure 10: ICC's for item number 11-20	49
Figure 11: ICC's item number 21-30.....	50
Figure 12: Graphical representation of c-parameter categories	51
Figure 13: Graphical representation of items in relation to all the three key parameters of IRT.	55

LIST OF TABLES

Table 1: English pass rates for the year 2018-2022	3
Table 2: Discrimination power of test items and recommendations.....	13
Table 3: Categories of P-values of items and possible recommendations in CTT	14
Table 4: Interpreting item difficult level in IRT	16
Table 5: General summary of overall MSCE pass rate from 2019-2022	24
Table 6: KMO and Bartlett's Test Results.....	37
Table 7: Summary of loadings and Eigenvalues which were extracted	38
Table 8: Interpretation of discrimination index values	41
Table 9: Very easy items.....	44
Table 10: Easy items with b-parameter ranging less than -2	45
Table 11: Moderately difficult items (average items) with b-parameters ranging from -1 to +1	45
Table 12: Difficult items with b-parameter +1 to +2.....	46
Table 13: Very difficult items with b parameters greater than +2	46
Table 14: Summary of results in relation to acceptable c-parameter range of $0 < c < 0.35$.	50

LIST OF APPENDICES

Appendix 1: Letter to MANEB, requesting authorization for data collection instrument	66
Appendix 2: Authorization letter from MANEB	68
Appendix 3: Data collection instrument	69
Appendix 4: Permission to conduct research.....	73
Appendix 5: Letter of introduction	75
Appendix 6: Informed Consent.....	76
Appendix 7: Confidentiality	77
Appendix 8: Chikalata Chakuvomeleza za Kudziwisidwa Za Kafukufuku	78
Appendix 9: A Summary of Loadings and Eigenvalues.....	84
Appendix 10: Table of Results after Regression	85

LIST OF ACRONYMS AND ABBREVIATIONS

ICC	Item characteristics curve
IRT	Item response theory
MANEB	Malawi National Examination Board
KMO	Kaiser-Meyer-Olkin
MSCE	Malawi School Certificate Education
SPSS	Statistical Packages for Social Scientists
SEED	South Eastern Education Division
3 PLM	Parametric Logistic Model

CHAPTER ONE

INTRODUCTION TO STUDY

1.1 Chapter overview

This chapter provides the introduction which gives the background information to the research topic in question, a statement of the problem, purpose of the study, research objectives, research questions, significance of the study, study delimitations and definitions of key terms (words) used in this write-up.

1.2 Introduction

Malawi National Examinations Board here in referred to as MANEB, is a government agency mandated by law to “conduct academic and other examinations at the conclusion of any approved course in Malawi as it may be considered desirable in the public interest,” (Malawi National Examinations Board Act, section 4 (9)(a). The Board was founded in April 1987 to conduct valid and reliable examinations for certification, selection and placement purposes. Besides that, it was also founded to provide professional advice relating to assessment and examinations (Mabumbe, 2022).

One of the academic examinations conducted by MANEB at the conclusion of an approved course is the Malawi School Certificate of Education (MSCE). MSCE is written by students who are completing their final year of secondary school education in

Malawi. For one to qualify for the award of an MSCE, he or she has to satisfy two conditions. The first condition is passing any six subjects, including English, but with a credit in at least one of the subjects. The second condition is passing any five subjects, including English, but with a credit in at least three of the subjects. As seen in the stated conditions for the award of an MSCE, it can be concluded that the English subject is an enabler. To put it in simple words, no candidate is awarded an MSCE unless they pass this subject. Therefore, it is a great concern whenever students perform poorly in this subject as it has a direct effect on the overall pass rate of the entire examination exercise.

In 2020 MSCE examinations, candidates did not perform well in English and subsequently the overall passing rate was also very poor. The Minister of Education; Agness Nyalonje, admitted that the results were the worst in the past decade and attributed the poor performance to circumstances that surrounded the administration of the 2020 MSCE which she said were beyond the Ministry's control, (Gwede, 2021). Factors like Covid-19 and cancellation of examinations were claimed to have contributed to the poor passing rate by different stakeholders. However, literature reviewed by the researcher so far is silent on the examination quality which was administered hence the interest in conducting item analysis in relation to key parameters of IRT.

Item analysis is very important in a way that it helps to check if the examination items functioned as intended and also it helps to determine the quality of the paper, hence the main focus in this study.

1.3 Problem statement

Results of the 2020 MSCE examination indicate that students did not perform well in English.

Table 1: English pass rates for the year 2018-2022

2018 ENGLISH PASS RATE			
Registered ENG	ENG Sat	Passed ENG	ENG Pass percentage
207941	198463	181992	91.70072003
2019 ENGLISH PASS RATE			
Registered ENG	ENG Sat	Passed ENG	ENG Pass percentage
97770	93043	75661	81.31831519
2020 ENGLISH PASS RATE			
Registered ENG	ENG Sat	Passed ENG	ENG Pass percentage
153834	138444	97646	70.53104504
2021 ENGLISH PASS RATE			
Registered ENG	ENG Sat	Passed ENG	ENG Pass percentage
128746	123366	101160	81.99990273

2022 ENGLISH PASS RATE			
Registered ENG	ENG Sat	Passed ENG	ENG Pass percentage
152484	148374	127065	85.63831938

Source: MANEB

Table 1 summarizes how students performed in English at MSCE from 2018 to 2022. From the table, it can be vividly seen that in 2020, students did not perform well as compared to the other 4 years.

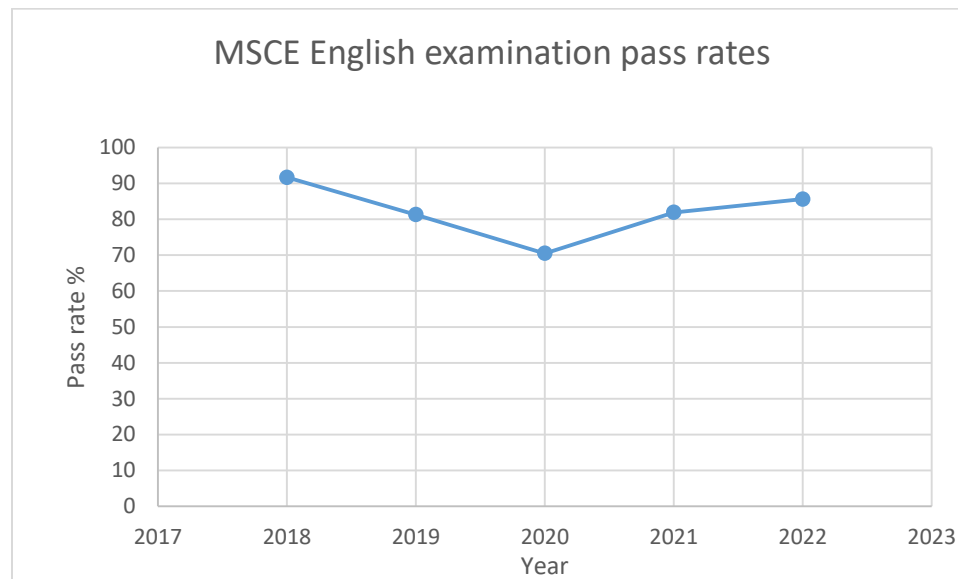


Figure 1: Graphical representation of English examination pass rate. Source: *researcher*

The poor pass rate in 2020 examinations has been attributed to several factors. For example, the Minister of Education; Agnes Nyalonje, attributed the students' poor performance to three main factors. In the first place, she attested to the fact that the Covid-19 pandemic had a negative bearing on the results as the school calendar was

disrupted and examinations were shifted from July to October, 2020. Worse still, schools had to close from March to September 2020. Consequently, some schools may not have completed the syllabus. The second factor was the cancellation of October 2020 MSCE due to massive examination leakage which left most students disoriented and confused. Lastly, Nyalonje attributed the poor performance to strict controls and security which left no room for minor examination leakage and cheating (Ministry of Education, 2020). Similarly, education rights activist and expert, Benedicto Kondowe attributed 2020's MSCE poor results to inadequate resources and infrastructure, poor leadership and management, inability to promptly adapt and respond to disaster, inadequate and demotivated teachers, poor supervision in schools, poor parenting, absenteeism, and drop out and repetition (Pembamoyo, 2021). None of the two authorities thought that English could have been one of the factors. But as it has been highlighted, English is a key subject of the process. And for that, it must be checked if it is one of these factors. A determination should be made whether it played a role in the student's unsatisfactory overall performance in the subject and the entire 2020 MSCE examinations.

Without trivializing the said external factors, Hambleton (2005) argues that factors inherent to the quality of the papers administered can also affect the quality of examination results. The quality of an examination paper can be assessed by analyzing psychometric parameters of examination papers in relation to item-difficulty, item discrimination and pseudo-guessing parameters by using the Three Parameter Logistic model (3-PL Model). However, existing literature as reviewed by the researcher is silent if this was done in 2020 MSCE examination papers, particularly English. Cognizant of

this gap in the literature, the researcher intended to fill it by assessing the psychometric parameters in the 2020 English MSCE Paper I.

1.4 Justification for the choice of English

The researcher decided to focus on English because English is a key subject in the Malawi's curriculum and according to MANEB, for one to qualify for the award of an MSCE, he or she has to satisfy the prior-stated conditions: firstly, one must pass any six subjects including English but must have a credit in at least one of the passed subjects, secondly, one must passing any five subjects including English but must have a credit in at least three of the subjects. This shows that English subject determines whether one will pass the MSCE examinations and be awarded an MSCE or not. Again, according to the pre-research enquiry done by the researcher in a number of schools in the South East Education Division (SEED), most students did not perform well in English in the 2020 MANEB examinations. This in a way contributed to the poor overall pass rate, making English subject an area of interest in this study.

1.5 Purpose of study

The purpose of this study was to assess the psychometric parameters of 2020 MSCE English Paper I in relation to item-difficulty, item discrimination and pseudo-guessing parameters using the Three Parameter Logistic Model (3-PL model).

1.6 Research Objectives

1.6.1 Main Objective

To assess the levels of IRT key parameters of the 2020 MSCE English examination Paper I.

1.6.2 Specific objectives

The study had the following specific study objectives:

- i. To analyze the level of item difficulty of the 2020 MSCE English examination Paper I.
- ii. To assess the discrimination index of the 2020 MSCE English examination Paper I.
- iii. To describe the pseudo-guessing level of the 2020 MSCE English examination Paper I.

1.7 Research questions

This research was guided by the following research questions;

1.7.1 Main Research Question

What are the levels of IRT key parameters of 2020 MSCE English Paper I?

1.7.2 Specific Research Questions

- i. What was the level of item difficulty of 2020 MSCE English examination Paper I?
- ii. What was the discrimination index of 2020 MSCE English examination Paper I?
- iii. What was the pseudo-guessing level of 2020 MSCE English examination Paper I?

1.8 Significance of the Study

This study is significant in the following ways: Firstly, it may give a clear insight to the public on the quality of the 2020 MSCE English Paper I examination by providing constructive feedback about item quality, increasing the effectiveness of the examination and supporting the examination's validity and reliability. Secondly, it may help to inform

teachers and other stakeholders like MANEB of the need for item analysis. Thirdly, the results of item analysis may be valuable in improving items which could be used again in later tests, by eliminating ambiguous or misleading items. Lastly, this study might create a benchmark for researchers who would want to conduct research in the related field. Mudavanhu (2017) stressed that researchers cannot perform effective studies without an adequate grasp of the topic of interest. A literature review provides a starting point for any future research as it highlights what has already been done in a particular field and helps one to identify the research gap that is still unexplored or under-researched.

1.9 Operational Definition of terms

Difficult level: the percentage of students that answered the item correctly (Boupathiraj, and Chellamani, 2013)

Discrimination power: the ability of an item to differentiate among students based on how well they know the material being tested, (Ferrando, 2012)

Dichotomously scored items: These are items which have only two possible outcomes; fully correct or fully incorrect, (Mellenbergh, 1977)

Examinees: candidates who sit for a particular selection test.

Experts: Secondary School English teachers who are well knowledgeable about the subject matter.

Item analysis: is a process in which both students' answers and test questions are examined to assess the quality and quantity of the items and the test as a whole. (Suckie, 2017).

Item Response Theory: is a paradigm in psychometrics that deals with the design, analysis and scoring of tests, questionnaires, and similar instruments measuring abilities, attitudes, or other variables (Baker, 2001).

Pseudo-guessing parameter: the probability of getting the item correct by guessing alone, (Hambleton, 1997).

Scores: marks obtained by examinees on a particular selection test

Test: an instrument used to judge achievement among learners (Nitko, 1996)

Test items: questions that make up an examination, (Nitko, 1996)

1.10 Study Limitations

This study had the following possible limitations; Firstly, the examination environment or mood was different from the mood set by MANEB during administration of MSCE and students may not have put much effort compared to the real national examination. . Secondly, taking into account that when developing end of term tests some teachers pick questions from past papers, it is possible that some of the students might have already been exposed to some of the questions in 2020 English Paper I before the research was conducted. Furthermore, the result would not be generalized to 2020 MSCE English examination papers administered that year since the study only focused on English Paper I. It has to be stressed that during the MSCE examinations, twenty-two subjects are

offered and administered. Most subjects comprise two examination papers except Chichewa and English which have two papers each. Therefore, it would not be feasible for the researcher to assess the psychometric parameters in all the subjects. However, the results of this study would still be informative for future development of examinations. Lastly, the researcher accepts the use of English Paper I only as a limitation. The use of Paper I only takes into consideration that the 3PL model used in Item Response Theory for analysis is for dichotomously scored items and not polytomously scored items. So, considering all the papers under this study would pose a very big challenge during analysis. To this end, the researcher only focused on English, especially the dichotomously scored items which was in line with the IRT model for data analysis in this study.

1.11 Chapter Summary

This chapter has given the background to the problem under scrutiny that is IRT key parameters (a, b and c- parameters). The relevance of this study has been explained by stating the research problem, purpose and significance of the study and the research questions (objectives). The dimensions that this study did not meet have been highlighted. For clarity, definitions of key words have been provided.

CHAPTER TWO

REVIEW OF RELATED LITERATURE

2.1 Chapter Overview

This chapter gives an insight into some issues tackled briefly in Chapter 1. It explains more about IRT and its key parameters and how it relates to examination validity. It also presents how Item Characteristics Curves (ICCs) are developed using the 3-Parametric Logistic Model (3PLM) and how they are used in the analysis of examination quality in relation to item difficulty level, item discrimination level and pseudo-guessing parameters

2.2 English as a Subject

English is a core subject in the general education curriculum in Malawi (Hanoi, 2018). It is one of the twenty-two subjects offered at the MSCE level. It is one of the required subjects a candidate has to pass to be awarded an MSCE.

2.3 A Test

A test is a measuring tool or instrument in education, (Gronlund, 1985). A test is used in determining students' knowledge, ability, competence and experience in and outside the formal learning settings. Among others, results from assessment are used for making a range of important decisions like placement, grading and certification. Understandably, the most serious consequence of failing an exam is that one cannot move up to the next

level and education opportunities become restricted. It is, therefore, imperative to ensure that tests to be administered are of good quality. There are several parameters which are used when evaluating the quality of a test. Some of them are validity, reliability, item difficulty level, and item discrimination index (Ahmand, 2016). The quality of a test depends upon the individual items of a test. Item analysis is a post-administration examination of a test.

2.4 Key parameters of IRT

2.4.1 Item discrimination

Item discrimination is the ability of an item to separate examinees with high ability from those with low ability (McCowan, 1999). There are several ways of computing an item discrimination but one of the most common way is the point-biserial correlation. Items that discriminate between students with different degrees of mastery based on the course content are desirable and will improve reliability. An item is considered to be discriminating if the “better” students tend to answer the item correctly while the “poorer” students tend to respond incorrectly (Boopathiraj and Chellamani, 2013).

Item discrimination shows whether the test items differentiate between people of varying degrees of knowledge and ability. It may be defined as the percentage of the “high” group passing the item minus the percentage of the “low” group passing the same item. The discriminating power of a test item refers to the degree to which success or failure of an item indicates possession of the ability being measured. In other words, the ability of the test items measures the better and poorer examinees of items. The index of discriminating power (D) indicates the degree to which an item discriminates between high and low

achiever on a single administration of the test. Research studies by Huang, (2003) and Obinne, (2011) found that good items to be included in the selection test should have discrimination indices (a-parameters) that are reliable and in the acceptable range, from 0.5 to 2 in the IRT psychometric framework (Mudavanhu, 2017). High discrimination levels indicate that the items discriminate well between low and high-skilled individuals (examinees). If the indices of the item discrimination (a-parameters) are above 1, they are normally very desirable for good test items that produce a very quality test. Items with discrimination indices between 0.75 and 1 could be considered in the test construction but with lots of caution.

Table 2: Discrimination power of test items and recommendations

Discrimination Index	Quality of item
$a \geq 1.70$	Excellent item
$1.35 \leq a \leq 1.69$	Good item; little or no revision is required
$0.65 \leq a \leq 1.34$	Moderate; little or no revision required
$0.35 \leq a \leq 0.64$	Item is marginal and needs revision
$a \leq 0.34$	Poor item
Negative value	Worst

Source: Bichi (2018)

Note: this study adopted this interpretation of the a-parameters in its results and discussion.

2.4.2 Item difficulty level

Item difficulty is the percentage of people who answer an item correctly. It is the relative frequency with which examinees choose the correct response. Item difficulty has a

powerful effect on both the variability of test scores and the precision with which test scores discriminate among groups of examinees (Thorndike, Cunningham, Thorndike, & Hagen, 1991). Item difficulty is also calculated using a proven formula. It measures the proportion of examinees who respond to an item correctly. On the flip side of the coin, in item difficulty, if most students answered an item correctly then the item was an easy one. If many students answered an item incorrectly then it should have been a difficult one (Brown, 1983). According to classical test theory, the higher the values of the difficulty index the easier the item. The easiest is administered first to give a sense of accomplishment and a feeling of an optimistic start (Maheshwari, 2013). McAlpine, (2002) did research on item analysis and according to his findings, he stressed that it is undesirable to have facility values (P-value) of a test above 0.85 or below 0.15. This means that a good quality test should have items with item difficulty within the range of 0.15 to 0.85 that is in CTT. Items with P-values below or above the range should not be included since they are regarded as too difficult or too easy items respectively for the examinees to get the correct responses, hence, compromising the quality of the test and producing low reliability and erroneous predictive validity of the test. But importantly a test should have questions which vary in their difficulty, so that all points of the ability stratum may be fully tested.

Table 3: Categories of P-values of items and possible recommendations in CTT

Difficult value	Quality	Recommendation
Below 0.20	Very difficult/ misleading	Discard
0.20 to 0.50	Good	Retain
0.51 to 0.80	Best	Retain
0.81 to 100	Very easy/ poor item	Discard

Source: Suruchi and Rana (2014)

It should be further noted that an item that is either too difficult that all the learners fail or too easy that all the learners pass should be evaluated as a bad item. When all the test items are very difficult, the majority of test scores will be on the lower side. On the other hand, when all the test items are too easy, the majority of scores will be on the higher side. As a result, there will be little or no variation at all (Matlock-Hetzel, 1997) and such an item does in any way contribute to measuring individual differences.

In IRT, the item difficult parameter is both a shape parameter of the item response function but also an important way to evaluate the performance of an item in a test. Note that the key parameter that is utilized in all three IRT models is the item difficult parameter, b . According to Laila, (2022), the b parameter determines the construct level we would expect examinees to have a probability (assuming no guessing) of getting the keyed item response. The probability of a correct response of an examinee getting a correct response is as a result of interaction between the ability of an examinee's and the item difficult parameter. This implies that as b parameter increases, it also requires students with high ability to have 50% chance of getting the item correct. Difficult items are located to the right or to the higher end of the ability scale while easier items are located to the left or to the lower end of the ability scale. The typical values of item difficulty difficult range from -3 to $+3$, and items with b values near -3 are very easy items and should not be included in an examination, whilst those near $+3$ corresponds to the items that are very difficult for the examinee. But in extreme cases, the index may go beyond the aforementioned index.

Table 4: Interpreting item difficult level in IRT

Difficult value (b)	Interpretation
$-3.00 \leq -2.00$	Very easy
$-2.00 \leq -1.00$	Easy
$-1.00 \leq 1.00$	Moderately difficult
$1.00 \leq 2.00$	Difficult
$b > 2.00$	Very difficult

Source: Bich (2018).

This study interpreted the item difficult values based on Bich, (2018) classification as illustrated in table 4. Rini, (2009) did research on Item Analysis of Achievement Test in Final Test for The Seventh Grade Students of SMP where he found out that the mean of the difficulty level was 0.60. So, the English summative test items were classified as medium items in terms of their difficulty level. Then the mean of the discrimination power was 0.48, and concluded that the items were still able to discriminate the clever students and the poorer ones. Based on the result, the writer suggested to the teachers as the test makers to prepare test items far in advance before they give it to the students. They should also pay attention to the writing of multiple choice items and the characteristics of a good language. Finally, the writer draws a conclusion that the items in the English final test for the Seventh Grade Students of SMP N 1 Moga Pemalang could still be used as an instrument of evaluation with some revisions.

Kurniawan, (2021) conducted a research on 'Item Analysis of Teacher Made Test in Biology Subject' and found that the evaluation results showed that only 37.5% of the

items were good enough and good characteristics. 63.5% of the items were poor and of bad characteristics. The criteria for item difficulty level was $0.3 > P$ is difficult, $0.7 > P > 0.3$ is moderate, and $P > 0.7$ is easy. The criteria for discriminating power were 0.4-1 is very satisfactory, 0.3-0.39 is satisfactory, 0.2-0.29 is unsatisfactory, and negative-0.19 is poor. The distractor is concluded to be effective if a minimum of 5% of all examinees are selected. The researcher recommended trainings organized by schools or education offices on the competence to make quality tests.

Mulyani, et al (2020) did a research on 'Quality Analysis Of Teacher Made Tests In Financial Accounting Subjects' and the results showed that there are still many teacher-made tests that are not suitable for use because they do not meet the level of difficulty, discriminatory power but had good distractors.

Research by Adedoyin and Mokobi (2013), which examined three statistical parameters that constitute quality of test items; item difficulty, item discrimination and pseudo guessing parameters using 3 parameter model of IRT framework on Botswana National Examinations found that over 97% of the total items included in the Junior Certificate Mathematics National Examination of 2010 were poor items. The research established that item difficulty indices (b-parameters) within the range -0.5 to +0.5 were items with fair or medium difficult levels, Items with b-parameters below -1 were labeled easy items (poor) and those with b-parameters greater than +1 were regarded as very hard items (poor) to be answered correctly by examinees. This knowledge impacted the Examination Boards to verify the items before they were included in the test or national examinations in Botswana. The researchers concluded that items with poor p-values or b-parameters

should be improved or modified before they are included in the test since they compromise the quality and function of the test. It was, therefore, recommended that examination bodies should consider improving the quality of their test items by conducting IRT psychometric analysis for validation purposes on test items before use.

Konala, (2020) examined the quality of Form One selection test for Faith Mission Secondary School by computing and analyzing the b-parameter, a-parameter and reliability level of the test items. He concluded that the examinations were very hard test, 89% of the items were poor in terms of a- parameter and, of low reliability and validity.

2.4.3 Pseudo guessing

Guessing behavior is an issue discussed widely with regard to multiple choice tests. In IRT, the c-parameter is also known as the guessing parameter. It represents the probability of a test-taker correctly answering an item even if they do not know the answer. Its primary effect is on number of correct scores for examinees at lower levels of proficiency. The item guessing parameter c, is the lowest value that an ICC curve attains. For example, an examinee who randomly select responses to items that have four response (multiple) choices can answer these items correctly about 1 out of 4 times, meaning that the probability of guessing correctly is about 0.25 (Carlson & Davier, 2013). The c-parameter complements a and b parameters by accounting for the likelihood of guessing since it quantifies the chance that an individual with no knowledge of the correct answer would still provide a correct response through random guessing. The c-parameter typically ranges between 0 and 1, with higher values indicating a higher probability of guessing correctly. A value of 0 indicates no guessing while a value of 1

suggests that a respondent has an equal chance of guessing the correct response as providing a correct response due to actual knowledge. Practically, values above 0.35 are considered unacceptable, hence the range $0 < c < 0.35$ is the one that is acceptable (Baker: 2001). A value higher than the results one find when he divides 1 by number of item options, often indicates that a distractor is not performing.

2.5. The three-parameter logistic model (3 PLM)

3PLM is a statistical model commonly used in regression analysis. It is an extension of the standard logistic regression model and allows for greater flexibility in modeling the relationship between predictor variables and a binary outcome. In this model, instead of assuming a linear relationship, it assumes a non-linear relationship using three parameters (a, b and c parameters) which are also of interest in this study, hence the choice of the researcher. The aforementioned parameters determine the shape, scale and location of the logistic curve. By estimating these three parameters, the three parametric model can provide a more accurate representation of the relationship between responses and underlying latent trait. It is worth noting that the three parametric logistic model is just one of many possible models used in regression analysis, and its choice depends on specific question and data characteristics under study. This model builds upon the two-parameter model.

2.6. Item Characteristics Curve (ICC)

The concept of Item Characteristics Curve (ICC) is used in IRT to show the relationship between examinee ability and performance on an item. In IRT, ability and item parameters are both estimated based on examinees' response patterns on the test

(Crooker, 1986). This concept is commonly used in educational measurement and psychometrics to evaluate the performance and quality of test items. Graphically, it represents the regression of the item score on examinees' ability, which is known as the item response function. This function is plotted with the ability level of examinee along the x-axis, against the probability of answering an item correctly on the y-axis. Each examinee is considered to have an ability score which places him/her somewhere on the ability scale presented by the Greek letter theta (Θ). At each ability level, there is a certain probability that an examinee with that ability will answer the item correctly. This probability is indicated as $P(\Theta)$. Typically, this probability is smaller for individuals with low ability than for those with higher ability levels. Therefore, if the probability function $P(\Theta)$ is plotted against ability level, the result takes the typically S-shaped curve which forms the ICC Curve. At the lower ability levels, the curve shows a low probability of answering the item correctly. As the ability level increases, the probability of answering the item correctly also increases. However, at higher ability levels, the curve reaches a plateau when even individuals with high ability have a constant probability of answering the item correctly. The steepness of the curve which is the slope, represents the discrimination power of the item. Items with steeper curves discriminate well between examinees with different ability level, while items with flatter curves discriminate poorly. The location of the curve on the ability continuum indicates the difficulty level of the item. Difficult items are located towards the right side of the continuum, while easier items are located towards the left side.

Figures 2, 3 and 4 are graphical presentation ICC's just to appreciate how the ICC's looks like in relation to the b- parameter which locates the curve.

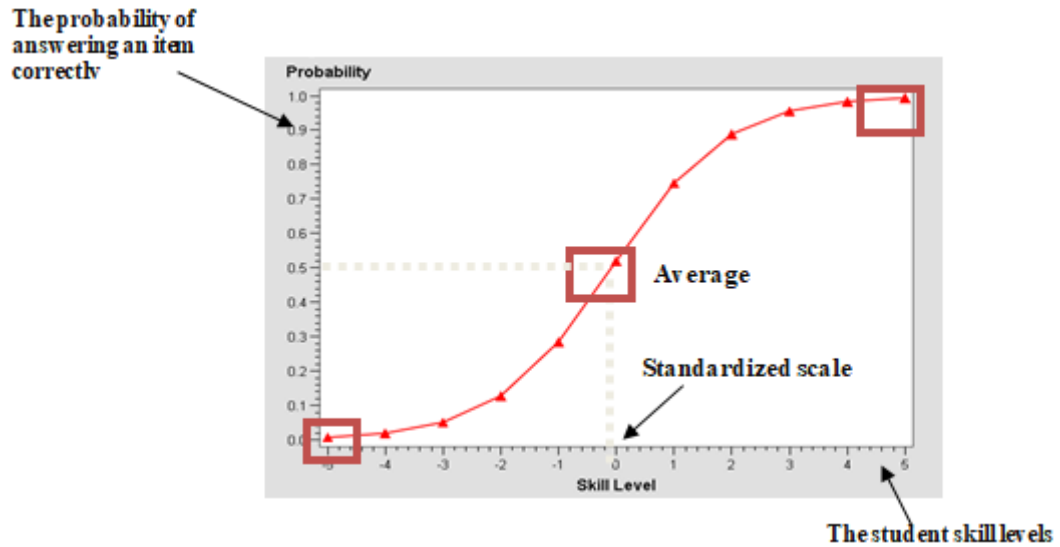


Figure 2: ICC for an average item

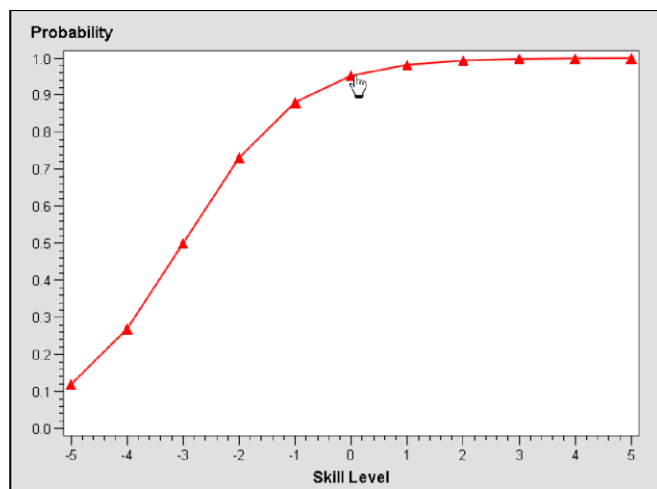


Figure 3: ICC for an easy item

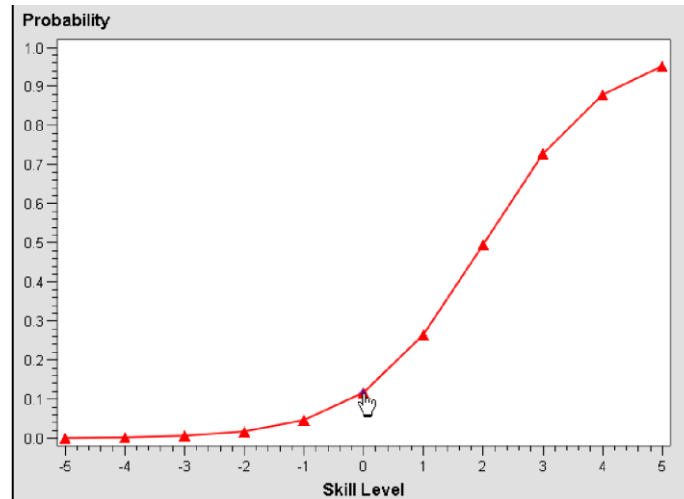


Figure 4: ICC for a difficult item

By analyzing the item characteristics curves of multiple test items, test developers can identify items that perform well and contribute effectively to the measurement of the construct being measured. They can also identify problematic items that may need revision or removal from the test. Overall, the item characteristics curve provides valuable information about the psychometric properties of test items and helps in ensuring the validity and reliability of assessments. These graphs will also guide the researcher the interpretation of the ICC's which will be developed later in the results and discussion. It should be noted that each item in this study, will have its own ICC separately.

2.7 Item analysis

Different scholars have come up with varied definitions of the term Item Analysis. For example, McCowan, (1999) defined item analysis as a set of statistical techniques to examine the performance of individual items. Item analysis is a post-test administration activity and it begins after the test has been administered and scored. Item analysis as a

process of “testing the item” ascertains specifically whether the item is functioning properly in measuring what the entire test is measuring hence item analysis tells about the quality of an item. It provides information concerning how well each item in the test functions (Maheshwari, 2013). It is a statistical technique used in educational and psychological testing to evaluate the quality of individual test items. Firstly, the data which are the scores or individual responses are collected and then calculations of required parameters follow. Thus, item analysis provides valuable insights into the properties of individual test items, allowing test developers to improve the overall quality and validity of assessments. It is commonly used in educational settings to refine and enhance the effectiveness of tests and examinations.

Item analysis indicates which item may be too easy or too difficult and which may fail for other reasons. This makes it transparent to discriminate clearly between the better and the poorer examinees. It is imperative to bring it to light that the purposes of item analysis are twofold: Firstly, it enables examiners to test improvement and evaluation procedures. Secondly, it helps one to plan, revise and improve instructions since item analysis indicates items or material which students have mastered and not.

Item analysis provides a way of measuring the quality of items and seeing how appropriate they were for the examinees and how well they measured their ability. IRT offers some advantages over CTT and has been widely used to analyze dichotomous types of data in educational testing. In general, item analysis procedures provide a basis

for increasing skills in test quality and test construction. Table 5 below shows the trends on MSCE passing percentages of students from 2019 to 2022.

Table 5: General summary of overall MSCE pass rate from 2019-2022

Year	2018	2019	2020	2021	2022
Passing %	63.23	50.36	41.42	51.99	61.66

Source, MANEB

From the table representing MSCE overall passing percentages, the average passing percentage so far is 51.35%. This means that almost half of the students who sat for the MSCE in the past four academic years failed the examination and they do not have an MSCE certificate. The worst among the pass rates is therefore the 2020 results. So far, literature reviewed by the researcher shows that 2020 MSCE English examination paper quality assessment has not been done yet. It is therefore the aim of this research to assess the quality of 2020 MSCE English Paper I by analyzing the key parameters of IRT, if at all the quality of the paper contributed to the poor performance in the 2020 examination.

2.8 Theoretical Framework

This study was guided by IRT, which is also known as Modern Mental Test Theory or Latent Trait Theory. It is a body of theory that describes the application of mathematical models that expresses the relationship between an individual's response to an item and the underlying latent trait, also called construct or ability (Ahmad, 2022). IRT theory has assumptions, one being that of unidimensionality of the measured trait, local independence, item invariance and monotocity (Hambleton: 2019). He also emphasized that an assumption common to the IRT models most widely used is that only one ability is measured by the items that make up the test. This is called the assumption of

unidimensionality. A concept related to unidimensionality is that of local independence. Under IRT, the term ‘ability’ denotes a latent trait that underlies the responses of examinees on the items of an instrument. This latent trait is expressed as beta (β) or theta (θ) according to the model used. With an appropriate IRT model, the ability level of an examinee can be accurately estimated with any set of items that measure ability. This is because, for a given ability, estimates of item characteristics hold regardless of the group being tested. It means a group of respondents of low ability will produce the same ICCs as a group of high ability (Baker, 2001). Conversely, the characteristics of test items for example difficulty, discrimination and guessing are accurately evaluated with any sample of examinees. The item parameters are not dependent upon the ability level of the examinees responding to the item.

The Item Response Theory (IRT) field provides a framework for modeling and analyzing item response data. It is a modern method of estimating both examinees' ability and item parameters since it gives invariant properties that can be used across different groups of examinees. IRT is a powerful tool used in the measurement of examinee ability, selection of test items and for equating tests. This is item-level modeling that gives IRT its advantages over Classical Test Theory (CTT). IRT is based on strong mathematical and statistical assumptions. It is only when these assumptions are met, at least to a reasonable degree, can item response theory methods be implemented effectively for analyzing educational and psychological test data and for drawing inferences about properties of the tests and the performance of individuals. Checking model assumptions and assessing the model data fit are routine in statistical endeavors (Hambleton, 1985). In a study, IRT is

used to select a model (3PLM) to use, dimensionality of items, item dependence, analysis of data and interpretation of findings and choice of which software (BILOG) is to be used for analysis.

2.9 Chapter Summary

Chapter 2 has reviewed the literature in relation to the concepts of item discrimination, item difficulty, pseudo-guessing, ICCs, and item analysis and also established a literature gap that needs to be filled with this study. From the literature reviewed by the researcher, it is clear that several scholars have been conducting research on item quality analysis and it is evident that quality of examination is a necessary component of assessment to obtain accurate results, upon which high-stakes decisions are made. The literature reviewed so far, is silent on item analysis in relation to the key parameters of IRT of 2020 MSCE English paper 1, therefore there was still a need to fill the gap by doing a quality assessment of the 2020 MSCE English paper 1 examination in relation to a, b and c parameter.

CHAPTER 3

METHODOLOGY

3.1 Chapter overview

This chapter presents study approach, research design, population, sampling procedures, data collection techniques, instrumentation, data analysis and ethical considerations that guided the research study.

3.2 Research Paradigm

This study used a positivist paradigm since the researcher believes that there is one objective reality out there yet to be discovered. A quantitative research approach was employed as the data to be collected, which are scores, provides objective information. According to Creswell, (2023), quantitative research tradition involves the collection of numerical data and the use of mathematical models as the methodology of data analysis. This approach helps in collecting objective data for statistical analysis as the researcher's questions. Data obtained is used objectively to measure reality after which meaning is created through the objectivity discovered in the interpretation of the collected data and also responding accordingly to the research questions posed.

3.3 Research design and methodology

The research adopted a cross-sectional descriptive research design in which data was collected from different groups of students but within the same period of time to examine

the quality of MSCE English Paper I examinations in terms of item difficulty, item discrimination, and pseudo-guessing parameters using IRT. Descriptive research describes and interprets the current status and is concerned with conditions that exist, practices that prevail, or trends that are developing (Siedlecki, 2020).

This was achieved through item analysis which uses statistics and expert judgment to evaluate tests based on the quality of individual items, item sets, and entire sets of items, as well as the relationship of each item to other items. Thompson and Levitov, (1985, p. 163) contend that item analysis investigates the performance of items considered individually either in relation to some external criterion or in relation to the remaining items on the test. Item analysis seeks to give some ideas of how well a test has performed relative to its purposes. Information from this research provides insight to the stakeholders on the quality of the 2020 MSCE English examination paper.

3.3 Sample size

The research had a sample size of 891 students who were drawn from secondary schools located in SEED. Such is the case since IRT requires a sample size to be large (around 1000) in order to obtain accurate item-parameter estimates that result in accurate estimates of ability, upon which some high-stakes decisions are made (Hambleton, 1989). Weiss, (2016) concluded that the sample size requirement for the estimation of item parameters should have a minimum of 500 calibration examinees in order to achieve adequate item parameter. Therefore, a sample size of 891 used in this study falls within the recommended range. The sample was drawn from one national secondary school, two district secondary schools, two conventional secondary schools, two community day

secondary schools and two private secondary schools. The study drew a very large sample of examinees to increase viability and consistence of test parameters (Lord, 1980). The schools are Phina Girl's Secondary School, Panga Secondary School, Mira Secondary School, Nona Secondary School, Widzi Day Secondary School, Lowe Urban Day Secondary School, Chipe Community Day Secondary School, Singa Community Day Secondary School, Mamia Profile Private Secondary School and Lili Private Secondary School. Note that the names of the schools used above are just pseudo-names as per ethics requirement. The study drew a sample from these different schools, to have a good representation of schools in SEED. It is worth pointing out that students are selected to secondary schools according to how well they have performed. That is, the best performing students go to national secondary schools, then to district secondary schools, then to day secondary schools, then to community day secondary schools and in the end and those who fail to secure places in the government schools go to private secondary schools, hence a need to have students samples from the aforementioned schools. That aside, the 2020 MSCE examinations were administered to all schools with learners of different learning abilities. Therefore, the choice of the schools is to have a good representation of the students who sat for the examinations in 2020. The sample size was around 90 students from each school, where both boys and girls sat for the examination. It should be noted that some schools had Form 4 students who were less than 90, more especially CDSS's, and that gap was filled up by those schools which had more than 90 Form 4 students. This was done in-order to have a large sample which is required for accurate IRT analysis and because IRT is group independent.

3.4 Sampling strategy

A random sampling strategy was employed in the selection of the students. Bonet, (2000) states that random sampling enables researchers to make a probability statement based on data collected at random from the targeted demographic. The students from each participating schools were selected using systematic random sampling more especially those schools with more Form 4 students. Mostafa, (2017) defined systematic random sampling as a common technique in which the researcher samples every k th element. In this study, the researcher developed a sampling frame for each school containing all the assigned numbers of the students (N) and then divided the size of the frame by the desired sample size (90) and then chose every k th element in the frame to create the sample of students who sat for the exams. The schools were purposively selected based on the required criteria of having schools from different quality levels of secondary schools (e.g. national, district, day, CDSS and private secondary schools).

3.5 Data Collection Techniques

2020 MSCE English Paper I dichotomously scored items were administered to the 2024 Form 4 students. It should be well noted that one of the assumptions of IRT is that it is group independent, that is, the results/scores of an examination are not affected by the population on which the test is being administered. IRT group independence means that relationships between item responses and latent traits are consistence across different groups of individuals (Nguyen, 2015). In most cases, IRT assumes group invariance which means that the item parameters (such as a , b and c parameters) are the same for all groups. The examination was marked by subject matter experts, and coincidentally, were

also those who took part in marking 2020 MSCE English Examination and the scores collected were analyzed accordingly.

3.6 Data Analysis

Data was entered into Microsoft Excel to develop a dichotomous score sheet which was later exported to STATA where preliminary descriptive statistical analysis was done and later IRT psychometric analysis. IRT assumptions of local independence and unidimensionality were tested and verified. The researcher used KMO and Bartlett's test to test for local independence that is to check if the items were not related to each other. Later, principal factor analysis was used to test for unidimensionality. Note that, there are some classical approaches for checking for unidimensionality with factor analysis as the widely used technique (Ziegler & Hagemann, 2015). A test is unidimensional if the systematic differences observed in the answers to the items are considered manifestation of the differences on the latent construct (Cronbach, 1947) and that is if the differences within the item variance are only due to one variance source which is the latent variable. Eigenvalues were computed and then first factor value was compared with the other factors. A scree plot was later developed so as to have a graphical representation of the values. However, other IRT assumptions of monotonicity and item invariance were not checked because it works better with grouped data which was not the case in this study. On IRT psychometric analysis, item responses were regressed on the underlying trait in order to come up with the key parameters of IRT for each item and this was done using 3PLM since literature has shown that it is the general model and has all the three parameters concerned in the current study. Based on the scores which were collected, psychometric parameters like item difficult (b-parameter), item discrimination (a-

parameter), and pseudo-guessing (c-parameter) were computed, ICC were developed and later interpreted

3.7 Ethical Consideration

In educational research, ethics is concerned with ensuring that the interests and the well-being of research participants are not harmed. Gajjar, (2013) defined ethics as a methods, procedural or perspective for deciding how to act and for analyzing complex problems and issues. At the beginning of the study, the researcher obtained ethical clearance from the University of Malawi Ethics and Research Committee (UNIMAREC). Thereafter, the researcher obtained permission to conduct the study in schools from the Education Division Manager (EDM) offices. Upon visiting the schools, the researcher explained clearly her intentions and sought permission from the school authorities. The researcher took permissions from all stakeholders. Respondent's anonymity and confidentiality was taken care of, names of the schools and participants in the study were assigned false names. All data generated was considered confidential and solely used for the purposes of this research study. Issues of voluntary participation and participation withdrawal were taken into account.

3.8 Reliability and validity of data collection instrument

Reliability and reliability form psychometric properties of measurement scales that are very important in estimating adequacy and accuracy procedures of a scientific research as argued by Bajpai, (2014). Kubai, (2019) defined reliability as the degree to which the measure of a construct is consistence or dependable, that is, being able to produce similar results when the instrument is administered even to different people, on different occasions and under different conditions. On the other hand, validity is the extent to

which a measure adequately represents the underlying construct that it is supposed to measure (Drost, 2011). To ensure reliability of the instrument, the researcher conducted a pilot study which comprised a sample size of 100 students. Later on after conducting the actual study, the results were compared and they were consistent. Students were able to respond to all the questions, omissions and errors were checked in the data base and corrections were made. All these were done so as to ensure paper reliability. The sample size was large enough and helped in increasing the consistency of the results. Note that, reliability is not only a feature of the measuring instrument, but it is also a feature of the results of the measuring instrument.

The instrument used was a MANEB examination paper which undergoes a series of steps starting from item development, proofreading, and piloting just to mention a few before the exam is administered and all this is done by experts to make sure that the items measure the construct it intends to measure. Also, to ensure the validity of the results, students were told well in advance to prepare for the exams. Therefore, it is evident that the research instrument used in this study was reliable and also valid.

3.9 Data Management Plan

Research data management is simply the effective handling of information that is created in the course of research. It is usually an integral part of the research process and the researcher in this study will make sure that the data is well managed to avoid the risk of data loss, increase research efficiency but also to improve the research integrity as well as act as validation for research results. This was achieved by making sure that the collected

data in this research which are scores, were well kept in named files both hardcopies and softcopies for easy access by the researcher and the soft copy valuable data was backed up to Google drive in case of any emergency. The information that was irrelevant was deleted and passwords were used as a way of controlling who has access to the collected data. This was done throughout the entire cycle of data from the point of data creation and it will continue through dissemination, publication and archiving and this will be done with regard to research ethics as well, where issues of confidentiality will be addressed.

3.10 Research dissemination strategy

Much as this research is for academic purposes and that the results will be presented accordingly to the academia body for grading. But still, given an opportunity, the researcher wishes to inform the stakeholders in this case the students, parents, schools and even MANEB about the quality of 2020 MSCE English Paper 1 in relation to the key parameters (a, b and c parameters) of Item Response Theory.

3.11 Research risks

A risk can be defined as the potential for an event or situation to occur that may have a negative impact on the objectives, goals or outcomes. This research might face a risk of social and legal consequences, where it might uncover controversial or sensitive information that can have social or legal implications. It may also face the risk of misinterpretations of research outcomes contrary to the initial intent. This could lead to misinformation, unethical practices or policy decisions based on flawed interpretation.

But above all, the researcher hopes to be able to manage the risks so as to minimize their potential consequences and be able to make informed decisions.

3.12 Foreseeable and unforeseeable circumstances

The researcher handled well all the circumstances which surrounded this study such as ethical consideration, availability of resources, time constraints, collaboration and communication, technological advancements, among others.

CHAPTER 4

RESULTS AND DISCUSSION OF THE STUDY

4.1 Chapter overview

This chapter presents the results for the study according to the research questions, which had to do with the determination of the item difficulty level (a-parameter), item discrimination (b-parameter) and pseudo-guessing parameter (c-parameter). Firstly, results on checking IRT assumptions of unidimensionality and local independence will be presented.

4.2 Results

The study involved a total of 891 participants who were students from different schools within SEED. Out of 891 participants, 453 were males and 438 were females, representing 50.84% and 49.16% respectively. IRT methods worked well with large samples as argued by Embretson and Riese, (2000) that the sample should not be less than 500 examinees for stable IRT item parameter estimates. The participants responded to an instrument that had 30 multiple choice items of 2020 MSCE English Paper I.

4.2.1 Checking for IRT assumptions

Dichotomously scored data sheet which was generated in excel sheet was imported to SPSS where the checking of local independence and unidimensionality was done.

4.2.1.1 Checking for Local Independence.

This is done to check if the responses given to the separate items in a test are mutually independent given a certain level of ability. Note that, there are a number of ways for checking local independence of test items, but this study adopted a KMO and Bartlett's Test. KMO is conducted to measure sample adequacy and also examine the strength of the partial correlation between variables. According to Reddy, (2019), Bartlett's test is used to test the null hypothesis in which a significant statistical test of not greater than 0.05 ($\text{sig} < 0.05$) shows that the correlation matrix is indeed not an identity matrix and that is suitable for factor analysis.

Table 6: KMO and Bartlett's Test Results

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.701
Bartlett's Test for Sphericity	Approximate Chi-square	1350.989
	df	435
	sig	.000

From the results above, a KMO value of 0.701 was obtained. This entails that there is a presence of an average correlation and it is plausible to conduct factor analysis. Babae, (2010) recommended that the value of KMO should be greater than 0.5 ($\text{KMO} > 0.5$) to support the validity of the used variables. The results shows Bartlett's test with a significance value of 0.000 which is less than 0.05 the correlation matrix is not identical and it holds for factor analysis. The results show that the items were not dependent of each other and hence the assumption of local independence holds looking at the values of both KMO and Bartlett's Test for Sphericity.

4.2.1.2 Checking for Unidimensionality

Unidimensionality in Item Response Theory (IRT) means that the test measures a single dimension (Livingston, 2020). However, the degree of unidimensionality can vary, and this can impact IRT calibration and scoring. Unidimensionality can be checked through eigenvalue plots of the inter-item correlation matrix to determine whether a dominant first factor exists, comparing the eigenvalue plots obtained using procedures like real data and simulated data, DIMTEST, or Multidimensional Scaling (MDS). This study adopted the use of eigenvalue plots of the inter-item correlation matrix with a Varimax rotation to extract the components (check Appendix 10 for all components). To have a dominant component, the first factor should account for more than 20% of variability in the data or the first eigenvalue should be four times larger than the second eigenvalue.

Table 7: Summary of loadings and Eigenvalues which were extracted

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	2.715	9.049	9.049
2	1.533	5.109	14.158
3	1.352	4.508	18.666
4	1.278	4.261	22.927
5	1.225	4.085	27.011
6	1.185	3.949	30.960
7	1.109	3.697	34.656
8	1.099	3.664	38.321
9	1.072	3.573	41.893
10	1.047	3.489	45.382
11	1.028	3.428	48.811
12	1.023	3.411	52.222

Extraction Method: Principal Component Analysis.

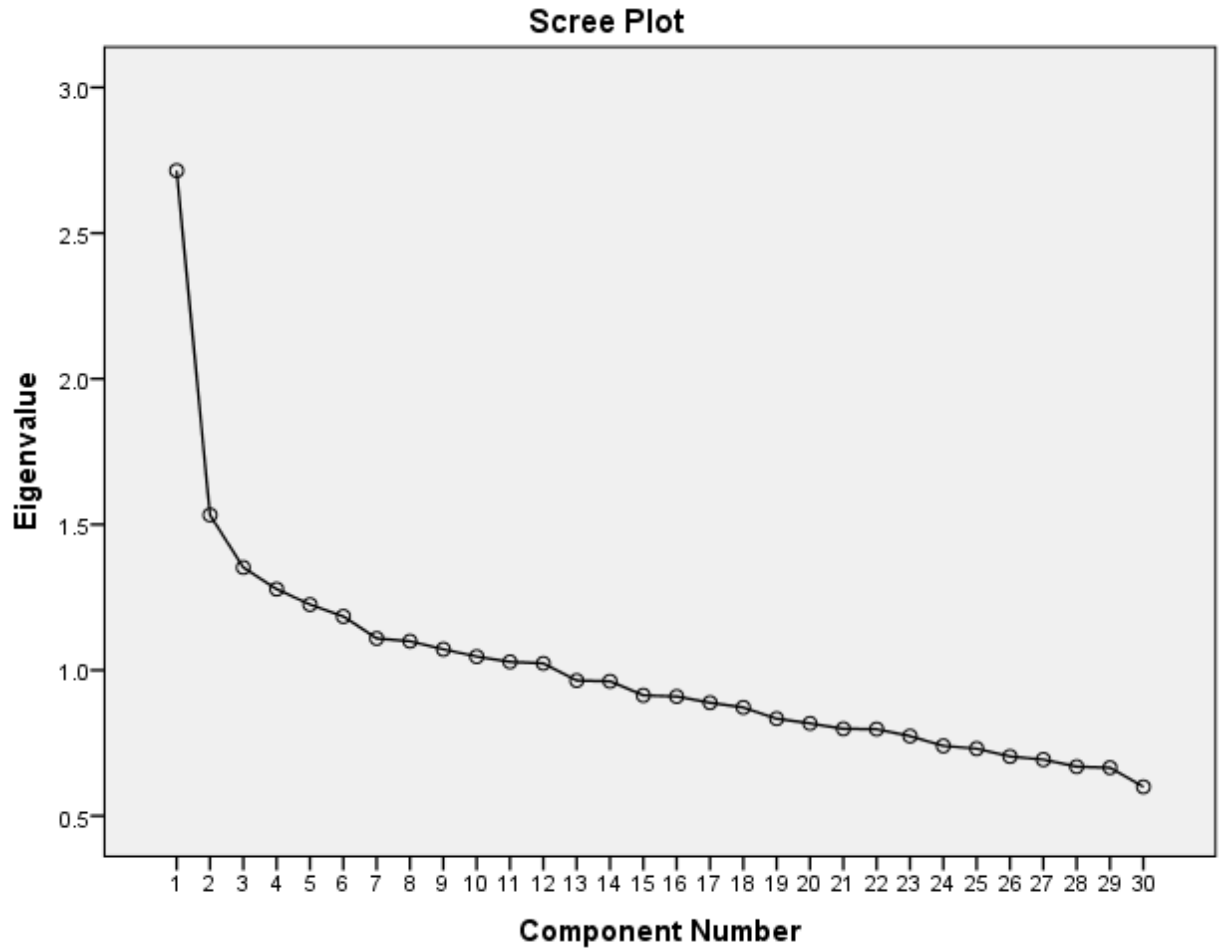


Figure 5: Scree Plot

Principal component analysis extracted 12 components of Eigen value less than one. Unidimensionality is an important assumption in Rasch and IRT models which explains that all items of a scale should measure a single unidimensional latent trait, that is, only one latent trait should explain the variability in the observed responses. From the 12 items which were extracted, they provide the differences of examinees and cumulatively provide 52.22% explanation of the test. The first component explains 9.049 % variance. This verifies that there is a total variability and it is indeed a dominant first factor and therefore the assumption of unidimensionality holds. Zeiger, (2015) emphasized that when unidimensionality holds, then the test reflects the target construct.

Now that the assumptions of local independence and unidimensionality were obtained, then it ascertains that the scale measures different levels of the latent traits and that there is no overlapping of the test items hence, the differences in observing correct responses between respondents will be due to variation in their latent trait. This is the first step before going into the regression specification of interest under this study. In this case, the researcher proceeded to apply IRT framework in data analysis.

4.2.2 Assessing IRT key parameters.

From excel, a dichotomously score sheet was developed and later imported to Stata for analysis. Then it was fitted to the 3 Parameter Logistic Model, which was most suitable PLM for the determination of the key parameters of IRT which was of interest in this study. This calibration provided values of a, b and c parameters at 95% confidence level. These parameters are the values which are found in the first column abbreviated coef. standing for the coefficient which is produced after regressing the data on 3PLM. Note that these values will be marred to the Item Characteristic Curves which will be developed from the same items for a further discussion. The table 8 first present all the difficult values which were obtained and later after the values will be consolidated in another table that is according to the difficult ranges presented in the literature review and their interpretation or recommendation will be made.

4.2.2.1 Item consolidation according to discrimination parameter ranges and interpretation

The item discrimination power (values) (a-parameters) that can be included in the selection test should range from +0.5 to +2 (Adedoyin & Mokobi, 2013). High discrimination level indicates that the items discriminate well between low and high skilled individuals (examinees) and they are best items for selection test. For very good quality items, the values of the item discrimination (**a**-parameters) should be between +1 and +2. They are normally more desirable values for better quality selection test and items with a-parameter values above +0.75 but less than +1 are less desirable quality items. Those items with a-parameters between +0.5 and +0.75 are questionable though they are sometimes acceptable in a selection test when they don't have alternative items.

Table 8: Interpretation of discrimination index values

Item # and it's a-parameter	Interpretation
Item 5 (1.693)	Good item (a parameter is +1.35 to +2)
Item 1(0.685), item 7(0.746), item 8 (0.759), Item 9 (1.131), item 10(0.951), item 15(0.785),item 16 (0.803), item 19 (1.103), item 21 (0.728), item 22 (0.842), item 23 (0.858), item 24 (0.764) item 25 (0.976), item 26 (0.686), item 29 (0.683), item 30 (0.665)	Moderate items (a parameter is +0.65 to +1.34)
item 3 (0.465), item 4 (0.572), item 6 (1.410), Item 11(0.466), item 12 (0.623), item 13	Marginal items (a parameter +0.35 to +0.64)

(0.528), item 14(0.644), item 17 (0.610), item 20 (0.361),	
Item 2 (0.273), item 18 (0.313)	Poor items ($a \leq 0.34$)
Item 27 (-0.232), item 28 (-0.302)	Worst items (negative a parameters)

In table 8, using theoretical framework scale recommended by Bichi, (2018), only 1 item out of 30 items was a good item of a parameter greater than 1.35 representing 3.3 %. 16 items out of 30 items under study were moderate items which may require a little improvement to make them good/excellent items, representing 53.33% of the total items under study. 9 out 30 items were marginal items which need to be improved thoroughly if they are to serve the intended purpose of discriminating students in relation to their knowledge; otherwise they are not supposed to be part of the test items. These marginal items represent 30% of the total items. And then 2 out of 30 items were poor items, which need to be eliminated. Thus, they were not supposed to be part of the test items. Poor items represent 6.7% of the total items. To the extreme, there are 2 items with negative a-parameters. These were the worst of all the items and they literally didn't serve any purpose of discriminating students. These items represent 6.7 % of the total items which were analyzed.

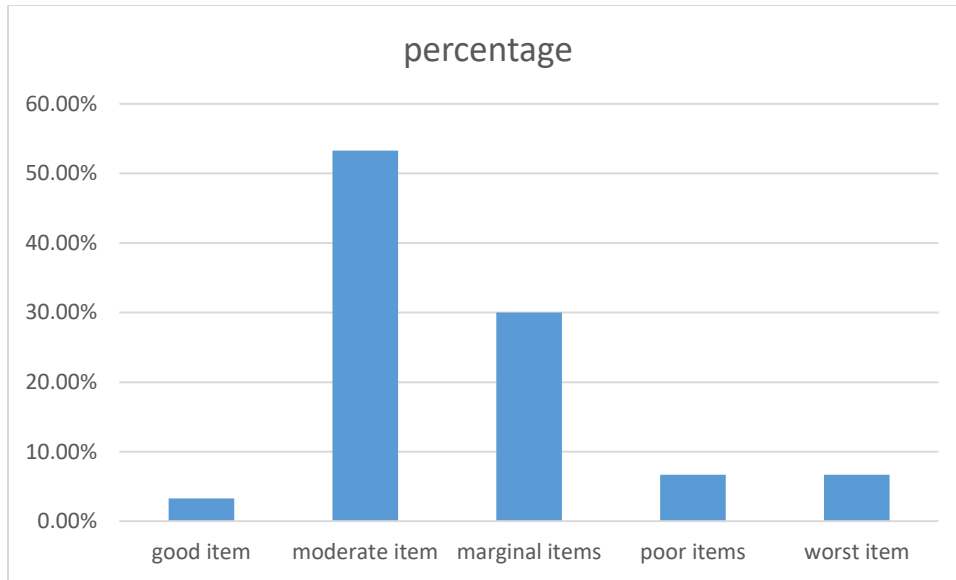


Figure 6: Graphical representation of a- parameters values in percentages.
Source: researcher

Looking at the bar graphs in figure 6, there were more moderate items than the rest. Generally, 56.6% of the items were reasonable to be part of the examination items in 2020 MSCE English Paper 1, more especially dichotomously scored items and about 43.4% of the total items were items which need a thoroughly improvement or which need to be eliminated for good.

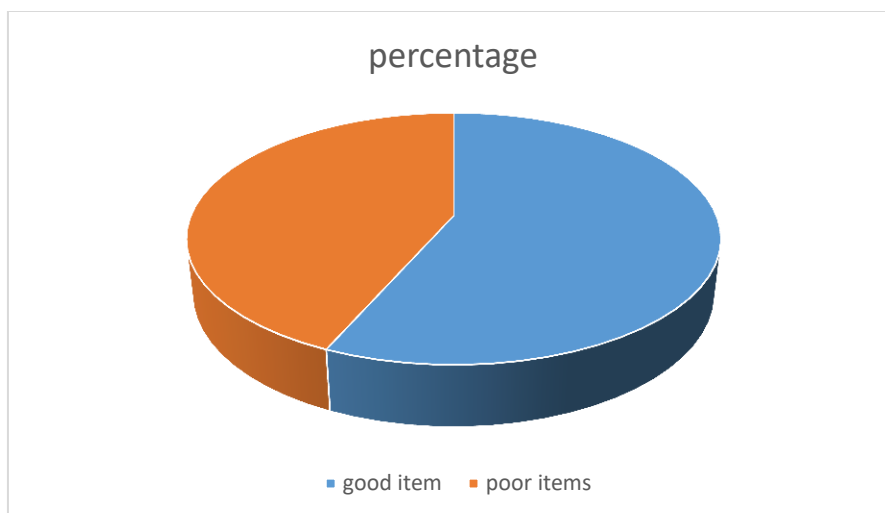


Figure 7: General representation of a- parameter values in percentage
Source: *researcher*

In summary, 43.4 % of the items administered in 2020 MSCE English Paper 1 more especially dichotomously scored items, failed to discriminate the students according to their knowledge and this directly compromised the quality of the paper and in the end contributed to the poor performance of students in 2020 MSCE examinations.

4.2.3. Interpretation of the indices in relation to b-parameter

Table 9: Very easy items

Item Number	b-parameter
1	-3.558
2	-2.641
8	-3.104
22	-2.023
26	-2,311
28	-3.553

Table 9 presents items which were regarded as very easy items. 6 out of 30 items were regarded as very easy items representing 20% of the total items which were analyzed. These items were not supposed to be included in the MSCE examinations. Thus, they are even supposed to be removed from the item bank.

Table 10: Easy items with b-parameter ranging less than -2

Item #	b-parameter	Item #	b-parameter
7	-1.477	23	-1.175
17	-1.095	24	-1.980
21	-1.651		

Table 10 presents items which were regarded as easy items. 5 out of 30 items which were analyzed were easy items representing 16.7% of the total items. These items need to be revised if they are to be included in the MANEB test item banks.

Table 11: Moderately difficult items (average items) with b-parameters ranging from -1 to +1

Item #	b-parameter	Item #	b-parameter
3	-0.420	18	0.317
6	-0.881	19	0.750
9	0.404	20	0.598
10	0.162	22	-0.902
11	-0.392	25	0.618
14	-0.834	29	0.384
16	0.306	30	0.009

Table 11 represents moderately difficult items. 14 out 30 items were regarded as moderately difficult items, also referred to as average items. This represents 46.7% of the total items analyzed. These were good items fit for assessing the Form 4 students who sat for 2020 MSCE English Paper I examinations.

Table 12: Difficult items with b-parameter +1 to +2

Item #	b-parameter	Item #	b-parameter
4	2.957	13	1.278
5	2.083	15	1.041

Table 12 represents difficult items. 4 out 30 items were regarded as difficult items, representing 13.3% of the total items which were analyzed. These items required students with high abilities amongst the Form 4 students in order to get correct responses. They were not supposed to be included in the test for good quality test parameters. They need to be revised if they are to be included in the test bank.

Table 13: Very difficult items with b parameters greater than +2

Item #	b-parameter
12	10.095

Table 13 presents very difficult items. 1 out of the 30 items was an extreme case, a very difficult item. Representing 3.3% of the total items analyzed. Demanding student of extreme levels of intelligence to respond to this item correctly. This items should literally be removed from the test items for it did not serve the intended purpose.

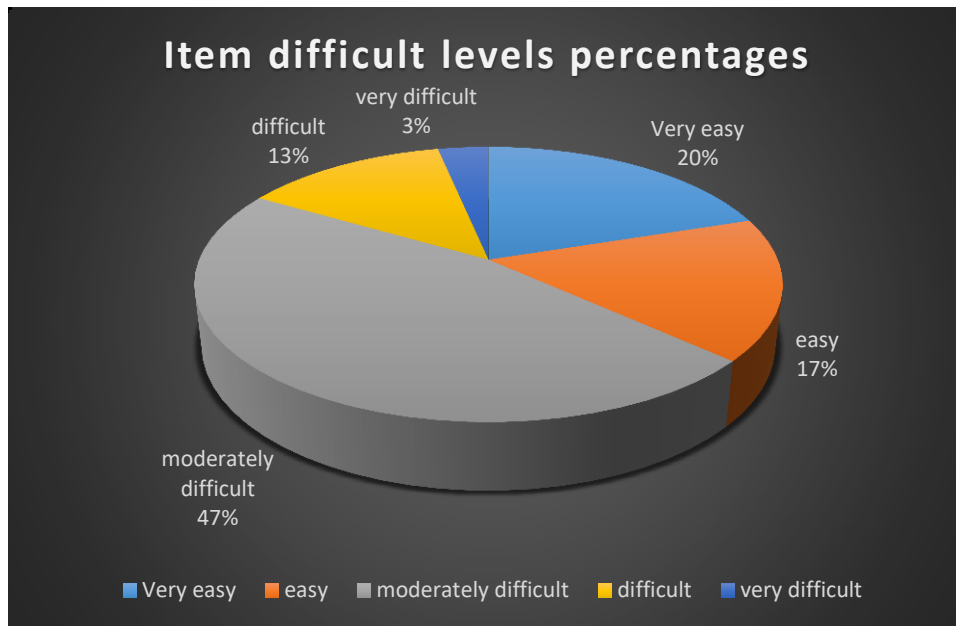


Figure 8: Graphical presentation of the item difficult levels in percentages related to the item difficult level ranges

Source: researcher

From figure 8, 47% of the items were fit enough in terms of difficult level to be part of the test items which were administered to the Form 4 students who sat for MANEB 2020 MSCE scrutiny of English Paper I, more especially dichotomously scored items. About 17% of the items were easy, 20% were very easy, 13% of the items were difficult items and 3% were very difficult making a total of 53% of the total items under study lacked quality in relation to IRT key parameters and this in a way compromised the quality of 2020 MSCE English Paper I examinations, that is according Bich, (2018) theoretical framework scale. It should be noted that, when test is hard, both high ability and low ability examinees fail to locate correct responses hence both get low scores. This means that it is difficult to identify individual ability levels of each examinee and likely to affect the decision making resulting into making erroneous judgment if the results are used by authority. Similarly, when a selection test is easy both high ability and low ability examinees get high scores hence the test fail again to identify examinees individual abilities resulting into all examinees seem to have same abilities that brings errors in decision making (Razak et al., 2012).

4.3 Interpretation of c-parameter in relation to Item Characteristic Curves

Further, the fitted data was used to develop item characteristics curve for each item and below were the results. It should be noted that selection of test items is a vital step in test development procedure and test items could be selected either through the item card or Item Characteristic Curve (ICC). In recent times, ICC has become a vital tool in the selection of tests items as it shows graphically the psychometric properties of the test items. At a glance the curve gives such information as discriminating, difficulty and guessing value (depending on the model). This study works on modeling of 3-Parameter Model of Item Response Theory (IRT) to generate the item characteristics curve using Stata packages in addition to the windows Microsoft excel application. Analyzing psychometric properties of items before administering it on examinees is a step towards achievement of test reliability and validity. It should be noted that, the application of Item Characteristic Curve (ICC) in test items selection gives a graphical presentation of the test items from which the c-parameter can be deduced. This provides information about the relationship between the ability of the test takers and the probability of getting the item right. Therefore, the ICC's which were developed were the ones helping in determining the level of getting the item correct through guessing which provided the study with the c-parameter of IRT. Based on the scores collected in this study, below are ICC's for all the items which were developed. Each graph carries ICC's for 10 items.

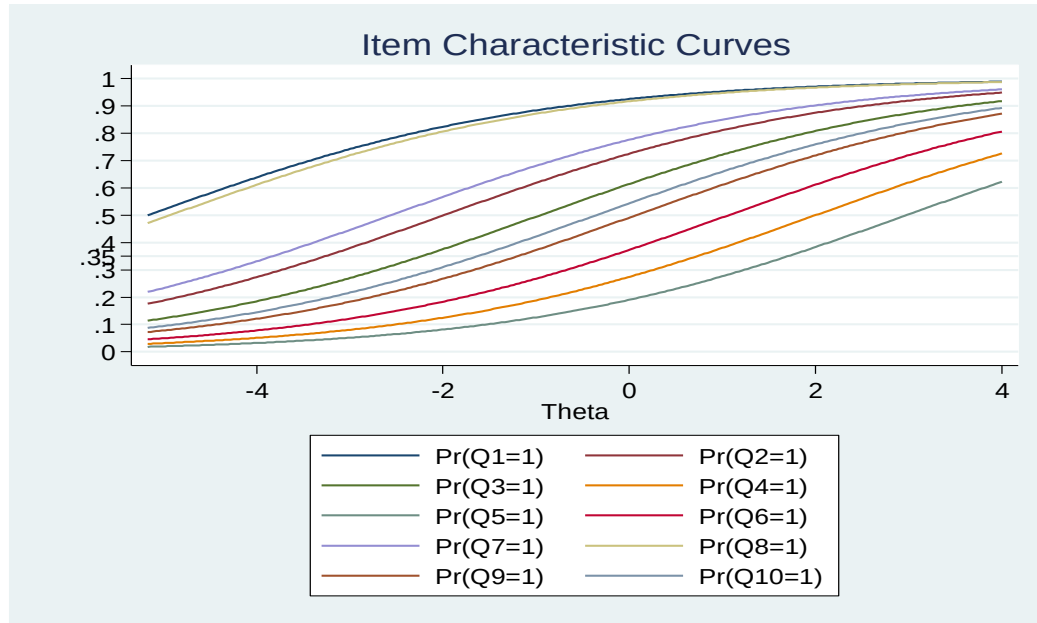


Figure 9: ICC's for item number 1-10

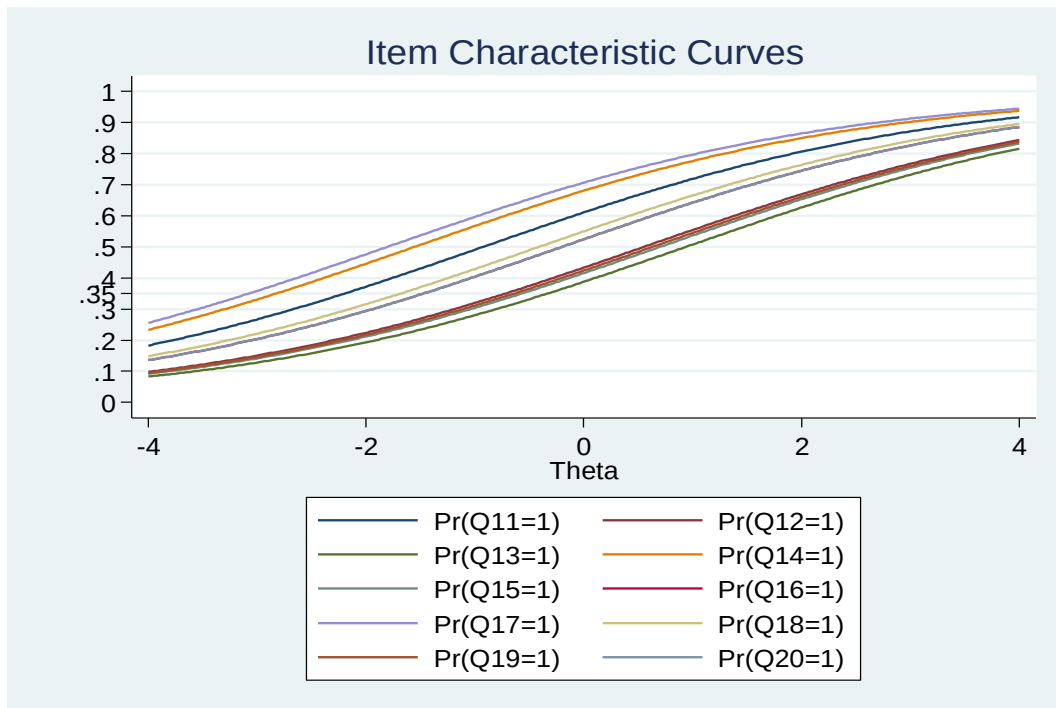


Figure 10: ICC's for item number 11-20

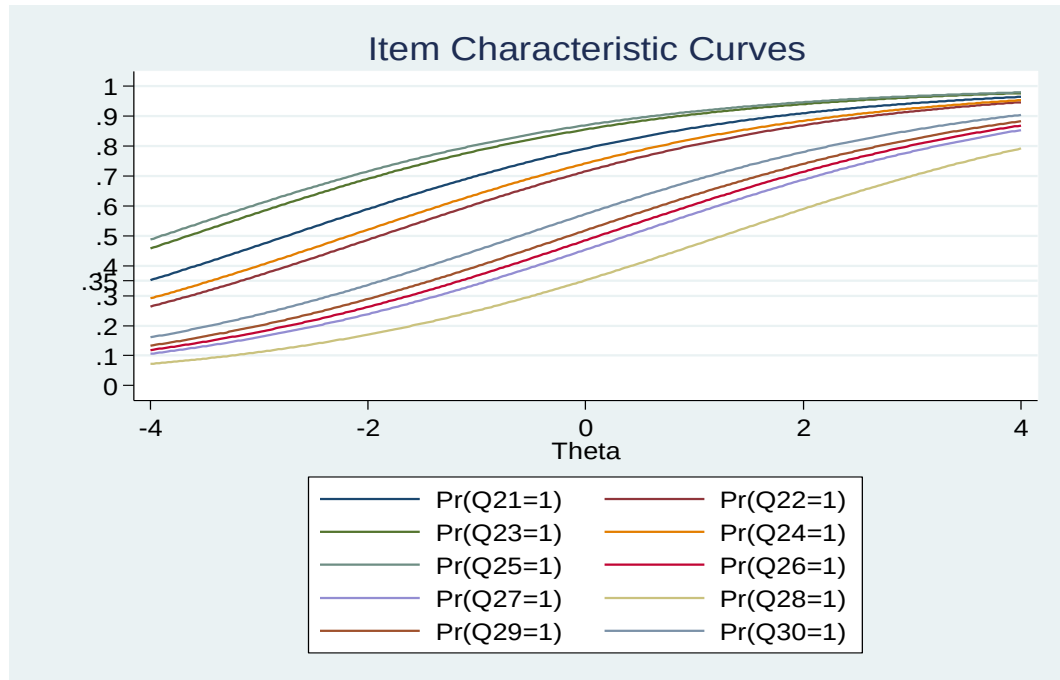


Figure 11: ICC's item number 21-30

Table 14: Summary of results in relation to acceptable c-parameter range of $0 < c < 0.35$

Items with c parameter within the acceptable range of c-parameter	Item not within acceptable range of c-parameter
Item 2, item 3, item 4, item 5, item 6, item 7, item8, item 9, item 10, item 11, item 12, item 13, item 14, item 15, item 16, item 17, item 18, item 19, item 20, item 22, item 24, item 25, item 26, item 27, item 28, item 29	Item 1, item 8, item 21, item 23 and item 30

4.3.1 Interpretation of the Pseudo Guessing Parameter in Relation to the Table

14

The ICC's also provided information on c-parameter by looking at the asymptote of the curves and considering the recommended range of c-parameter of greater than 0 but less than 0.35 that is according to (Baker, 2001). 25 out of 30 items under study falls within the acceptable of c-parameter and 5 out 30 items were beyond the acceptable range, representing 83.3% and 16.7% respectively. 16.7% of the items had a high probability of students getting them correct through guessing and were not supposed to be part of the test items administered in 2020 MSCE English Paper I.

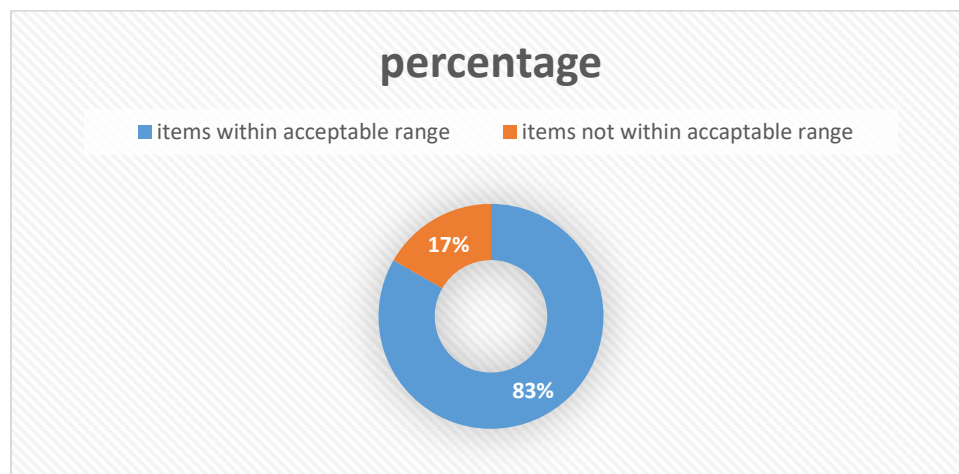


Figure 12: Graphical representation of c-parameter categories

Source; *researcher*

Almost 83% of the items under study were okay in relation to the c-parameter and provided acceptable probability range of getting the items correct through guess work. While 17% of the items had their probability values beyond the acceptable range of $0 < c < 0.35$, implying that they promoted guess work.

4.4 General discussion of the results

Based on the results above, 47% of the items were fair enough, suiting both students of low ability and high ability getting the items correct, that is according to Bichi's, (2018) theoretical framework scale. These items fairly discriminated the students according to their level of knowledge and provided a minimal chance of getting the item correct due to guessing. Note that, tests with average difficulty level ably discriminate higher achievers and lower achievers.

16% of the items were very difficult, requiring those students with high ability getting the items correctly. Showing that it failed to discriminate the students according to their knowledge level. These items provided a very slight chance of guessing. Should be noted that, some unrelated aspects of the items were irrelatively difficult for some individuals or groups. Example of such construct irrelevant difficult are unfamiliar context in a particular test. This causes one to obtain a lower score. Such items needs to be improved or eliminated from the item bank. Those items reduced quality of test in general according to Adedoyin and Makobi (2013) who found that a test with item difficulty parameters outside the range of -1 and +1 is a poor test.

Out of the items, 37% were very easy even students of low ability managed to get the items correct but then with very high probability of them guessing the correct response. Some of the items under this category, one wonders how students of high ability failed to get the items correctly hence even failing the whole exam in the end which is evidence that those with low ability got the item right by guessing. Concept of guessing occurs

when unrelated clues in items or tasks permit some individuals to respond correctly or appropriately in ways that are irrelevant to the construct being measured. Example of construct irrelevance easiness are guessing or test wiseness. These things compromise the reliability and validity of the test. Holland and Winner (1993) advocate that construct-irrelevant variance for individuals and groups is a major source of bias when interpreting test score. Differences in the construct irrelevant easiness or difficulty for different groups are the central focus even in testing. These items failed to discriminate the students according to their ability level. These items needs to be improved, if they are to be used in the next examinations.

As it is, almost 53% of the items administered had areas that needed improvement or totally being removed from the test items which were administered in 2020 MSCE English Paper I examinations. Improvement of test items is done so that they become average to accommodate average and high ability student's levels. This could be the one of the reasons why in 2020 MSCE examination and English in this case students did not perform well. The quality of test items has a large impact on reliability of the test and a good test must have items which are of appropriate difficulty level (Hills, 1991). Poor items often reduce reliability while good items tend to increase reliability.

Thirdly, 56.6% of the items managed to discriminate the students according to their knowledge levels well while about 43.4% of the items under study failed to. It is worth pointing out that, item analysis helps examiners to detect items with flaws as well as those that decrease the validity of the test. Items with negative discrimination are the ones that decrease the validity of the test, like in this case item number 27 has the

discrimination index of -0.2, similar to item 28 whose b parameter is -0.3. These items with negative discrimination are the worst items and chance given these items should be removed from the item bank for they did not serve the intended purpose. This is in line with what McCowan (1999) who said that negative coefficients are the worst and when more students in the lower group than in the upper group select the right answer to an item, the item actually has negative validity. Note that, item analysis enhances the internal consistency of the test and in turn, improves its reliability as well as its validity, hence items should be pretested and thoroughly analyzed before administered. Test items should be able to discriminate higher scorers and lower scorers. In order to determine whether an item is excellent, good or poor one has to evaluate the item's discrimination power. Items that discriminate between students with different degrees of mastery of subject matter are most desirable and they improve test reliability. An item is considered to be discriminating if the "better" students tend to answer the item correctly while the "poorer" students tend to respond incorrectly (Boopathiraj and Chellamani, 2013).

4.4.1 Items with all the key parameters of IRT

Only 7 out of the 30 items under study which had all three key parameters of IRT that is according to the analysis which has been done. They are within the range of -1 to +1, that's moderate item range according to item difficulty level scale. In terms of discrimination they are within the recommended average range of 0.65 – 1.34 of item discrimination index range. These 7 items have lower asymptote, indicating that there was no or very minimal room for getting the answers correct through guessing by those students with low ability. These 7 items are; item 9, 10, 16, 19, 22, 29 and 30

representing 23.3% of the total test items under this study. This means that about 76.7% of the total items under study had shortfalls, they did not meet the recommended ranges of IRT key parameters index.

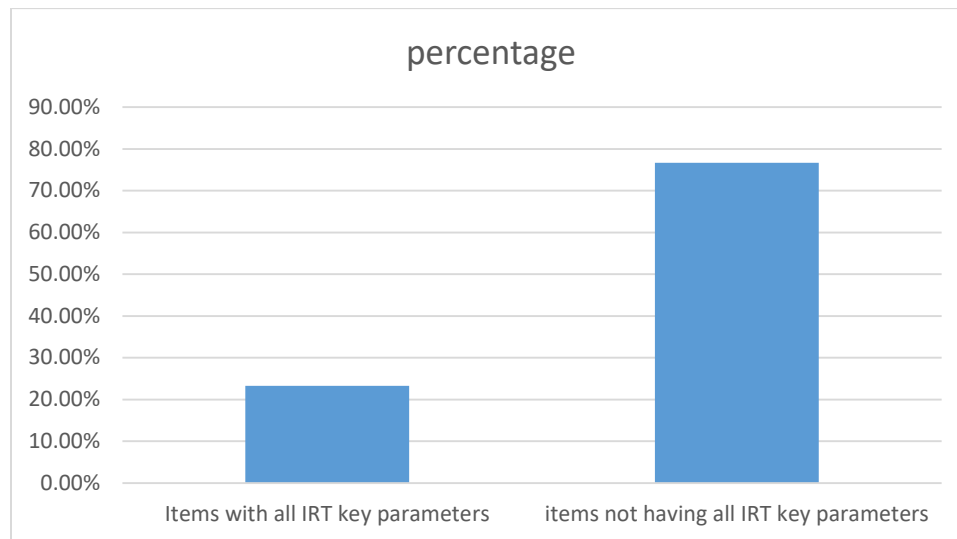


Figure 13: Graphical representation of items in relation to all the three key parameters of IRT.

*Source: **Researcher***

4.5 Chapter summary

Summing it up, the data was better fitted in 3PLM, hence all the parameters of interest in this study were determined using this model. ICC's were developed and analyzed. The analysis showed that, some items were very easy, some fair and some very difficult. Some items fairly discriminated the students according to their ability while others failed to do so. Lastly, for some items students managed to get them correct due to guessing. In one way or the other, the flawed items compromised the quality of the paper and in return this among other factors contributed negatively performance of the students in 2020 MSCE English paper 1 examination and even the overall pass rate.

CHAPTER 5

SUMMARY, CONCLUSIONS AND IMPLICATIONS.

5.1. Chapter overview

This chapter presents a summary of the findings through a highlight of the research questions, relationship of these results to similar researches, implications of the findings to practice and lastly, delimitations of the study and directions for the future research.

5.2 Summary of results

It should always be remembered that IRT studies are done with an intention to improve the item quality in terms of difficulty level, item discrimination level and pseudo-guessing parameter of the items which is also connected to improving the validity of the examination instrument. This study attempted to analyze MSCE 2020 English Paper I, administered by MANEB, more especially dichotomously scored items. The purpose was to see if the instrument was ok in relation to three key parameters of IRT.

5.2.1 To what extent in terms of a parameters does the MSCE 2020 English paper 1 had?

About 56.6% of the items on 2020 MSCE English Paper I were average and manage to fairly discriminate the students according to their knowledge. About 43.4% failed to discriminate well the students according to their knowledge because some were either very easy or very difficult and their p-values fall out beyond the recommended

discrimination index range for a fair examination item that is in relation to Bichi's (2018) theoretical framework scale. Remember that the numerical value of discrimination power (Dp) may range from -1.00 to +1.00. Thus, items with Dp values above 0.39 are rated Excellent, those from 0.30 to 0.39 are said to be Good, those from 0.20 to 0.29 are rated as Mediocre, those from 0.0 to 0.20 are rated as Poor and items less than -0.01 are rated as Worst.

5.2.2 To what extent in terms of b parameters does the MSCE 2020 English paper 1 had?

2020 MSCE English Paper I, had about 3% of the dichotomously scored items which were very difficult, 13% were difficult requiring those with high ability to be able to get the items correctly. 20% were easy, 17 % of the items were very easy and 47 % of the items were of moderate difficult level. Griswold (1990) also agrees that test items need to be at a reasonable difficult level. Here, we see that analyzing item difficulty is important because it has a profound effect on both the variability of test scores and the precision with which test scores discriminate among different groups of examinees (Thorndike et al, 1991).

5.2.3 To what extent in terms of c parameters does the MSCE 2020 English paper 1 had?

Basing on the probability level at the asymptote of the ICC curves, about 80% of the total items under study, had their c-parameters within the acceptable range of $0 < c < 0.35$, they did have a low probability of students getting them correct through guessing. The remaining 20% of the items had a high probability of students getting them correct through guessing. Implying that in total, 20% of the items were not supposed to be part of

the test items administered in 2020 MSCE English Paper I. They need to be improved or eliminated from the MANEB item bank, for they have shown to promote guess work according to the asymptote of their ICC's.

5.3. Implication of findings

It is possible that similar results would be obtained if the instruments for 2020 MSCE examinations were analyzed. The results obtained in this study gives confidence to say that the quality of English Paper I administered in 2020 was compromised and in one way or the other contributing to the poor pass rate of students in the subject area but also in the overall pass rate in general. Bearing in mind other social-economic factors that may have also influence the failure of the students in 2020, but the effect of the testing instrument cannot go unnoticed.

As seen from the IRT key parameters analysis, most of the items were of poor quality in terms of difficult level. About 53% of the items were either too easy or too difficult. Item flaws introduce the systematic error of construct irrelevant variance to assessment, thereby reducing the validity evidence for examinations and penalizing some examinees, that's according to Downing (2004). The remaining 47% were of moderate difficult level and they served their purpose well.

Additionally, most of the items about 43.4% failed to discriminate the students according to their knowledge and also gave room for guessing. 56.6% of the test items under study were of average discrimination index. Note that items that highly discriminate are the one's which are desirable. All things being equal high discriminating items also produce

high scores of reliability, Downing (2004) emphasized. Generally, very few items were very ok in terms of all the three key parameters of IRT. This is worrisome since unfairness of the test items affects even the validity of the instrument. It's possible that some students did not qualify for an MSCE certificate in 2020 due to the poor quality of some items. It should be noted that MSCE examination is a high stake examination and the results obtained from it are used to make high stake decisions, hence a need for quality assurance.

Subjecting the MSCE examination instruments to item analysis of the three key parameters of IRT and most importantly to other psychometric analysis in general before they are administered may help to assure the quality of the papers administered and in the end the validity of the examinations. In this study, some items are of good quality, some needs to be improved if they are to be part of the MANEB item bank and some needs to be eliminated for good. Conducting these studies helps bring to light test quality in terms of item difficulty, item discrimination, and item distractors and also helps in perfecting examiners skills in test construction

5.4 Conclusion and Future Directions for Research

This study analyzed the key parameters of IRT of 2020 MSCE English paper 1. The parameters are item difficulty level (a-parameter), item discrimination level (b-parameter) and also pseudo-guessing level (c-parameter), where these parameters are considered as the core-elements of examination validity. The definition of these parameters were given and more interests were shown on the same by regressing the scores, develop and interpreting the ICC's using the 3PL model.

Most items analyzed shown that they were not fair (about 53%) and some were okay (about 47%) in terms of difficult level. Some of the items (43.4%) items failed to discriminate the students in relation to their ability. Nevertheless, some items (63.33%) has a high probability of student getting them correct through guessing, in which students with low ability managed to get the items correct. Implying that 36.67% of the items were ok to be part of examination items, as they have low probability of students getting the item correct through guessing. In as much as some external factors may have contributed to the failure of the students in 2020 MSCE English examinations, the tool for testing (the instrument in this case) plays a great role, therefore, its contribution to the final decision made cannot be underestimated.

It is very important that items for examinations should be analyzed thoroughly for psychometric properties including difficult level, discrimination level and also pseudo-guessing level and if any item shows to lack these qualities among others, should be revised or eliminated after a thorough psychometric analysis. In case where items cannot be pretested for security reasons, then those responsible for selecting the items for the final paper, should carefully choose items which are of good quality. Recommended levels of $a > 0.3$, $0.3 < b < 0.35$ and $0 < c < 0.35$ for discrimination, for difficult level and for guessing parameters respectively should always be considered in examination composition by examiners and in this case MANEB

This study leads to the following future inquiries to be considered; firstly, future IRT key parameters should be done on the same items using 4PL model to also test for The slip parameter which is a fourth key parameter of IRT. Currently, the 4PL model is rarely

used, since the interpretation of the higher asymptote (on ICC) regarded as slip, carelessness or disinterests is questionable. So further research should be conducted to clear this questionable mist. Note that, the 4PLM gives room for explaining why respondents with high ability may miss the correct choice to an easy item.

Secondly, future research should be conducted on other MSCE subject's examinations which were administered in 2020 by MANEB, to check their qualities in relation to the same a, b and c key parameters of IRT.

Thirdly, future research on other psychometric properties like reliability and validity should be conducted on all the examinations subjects administered by MANEB in 2020 MSCE examination, to see if at all the quality of the papers was compromised or not.

5.5 Chapter summary

Chapter 5 has given the summary and conclusion of this research study, which has shown that 2020 MSCE English paper 1 administered by MANEB had items which needs to be improved or eliminated as they lack the required quality in relation to IRT key parameters (a, b and c parameters). The results have implications as discussed, which are crucial in maintaining the validity of the examinations. For further inquiry regarding studies of similar nature, directions have been suggested.

REFERENCES

- Ahmad, H. (2016). Is 3PL Item Response Theory an Appropriate Model for Dichotomous Item Analysis of the Anatomy & Physiology Final Examination. *Jurnal Pendidikan Sains Dan Matematik Malaysia*, 6(1), 13-23
- Adedoyin, O.O. & Makobi, T. (2020). Using IRT psychometric analysis in examining the quality of junior certificate Mathematics multiple choice examination test items. *International Journal of Asian Social Science*, 3 (4), 992-1011.
- Ajira. W. (2022). *Maneb MSCE Results 2022 Pass list Download*. MANEB
- Babae. N. (2010). *Investigating Effective Factors and Presenting Practical Guideline to Adoption of Mobile ticketing* (Master's thesis). Lulea University of Technology.
- Baker, F.B (2001). *The Basics of Item Response Theory*. University of Campinus.
- Bichi, A. A (2018). Item Response Theory: An Introduction to Latent Trait Models to Test and Item Development. *International Journal of Evaluation and Research in Education*, 7(2), 142-51
- Bonet. D. G. & Wright. T. A (2000). Sample size requirement for estimating Pearson, Kendal and Spearman's correlation. *Psychometrics*, 65(1), 23-28
- Boopathiraj, C. & Chellamani, K. (2013). Analysis of Test Items on Difficulty Level and Discrimination Index in the Test for Research in Education. *International Journal of Social Science and Interdisciplinary Research*, 2 (2), 69-75
- Bajpai, S.R. & Bajpai, R.C. (2014). Goodness of Measurement: Reliability and Validity. *International Journal of Medical Science and Public Health*, 3(2), 112-115
- Cronbach, L. J. (1947). Test "reliability": its meaning and determination. *Psychometrika*, 12(1), 1-16. <https://doi.org/10.1007/BF02289289>
- Downing, S.M. (2004). *The Effects of Violating Standard Item Writing Principles on Test and Students*. <http://doi.org/10.100/s10459-004-4019-5>.

- Drost, E.A. (2011). Validity and Reliability in Social Science Research. *Education Research and Perspectives*, 38(1), 105-124
- Embretson, S.E., & Reise, S.P. (2000). *Item Response Theory for Psychologists*. Lawrence Erlbaum Associates.
- Ferrando, J. (2012). *Assessing the Discrimination Power*. Rovira I Virgili University.
- Gronlund, N. E. (1985). *Measurement and evaluation in teaching* (5th ed.). Macmillan.
- Gajjar, N. B. (2013). Ethical Consideration in Research. *Education*, 2(7), 8-15.
- Gwede, W. (2021). *MSCE exams 60% fail rate worst in the past decade*.
<https://www.nyasatimes.com/msce-exams-60-fail-rate-worst-in-the-past-decadesays-education-minister/>
- Hambleton, R. K. (1994). Item Response Theory: A Broad Psychometric Framework for Measurement Advances, 6(3), 535-556
- Hambleton, R.K. & Slater, S.C. (1997). Item Response Theory models and test practices: Current International status and future directions. *European Journal of Psychology Assessment*, 13(1), 21-28
- Hambleton, R. K. & Swaminathan, H. (1985). *Item Response Theory: Principles and Application*. Kluwer
- Hambleton, R.K, Merenda P.F & Spelberger, C.D. (2005). *Assessing the Quality Tests. Revision Of the EFPA review model*
- Hanoi, B.I.S. (2018). *General Education: English Language Curriculum*. British Council
- Hills, J. R. (1991). Apathy Concerning Grading and Testing. *Phi Delta Kappan*, 72(7), 540-545.
- Konala, L.J (2020). *Examined the Quality of Form One Selection Test for Faith Mission Secondary School* (Master's thesis). University of Malawi.
- Kubai, E. (2019). Reliabilty and Validity of Research Instrument Correspondance to kubaeiedwin@yahoo.com.

- https://www.researchgate.net/publication/335827941_Reliability_and_Validity_of_Research_Instruments_Correspondence_to_kubaiedwinyahoocom
- Laila, I. (2022, December 18). Psychometrics: *Parameters and Models of Item Response Theory*. <https://www.youtube.com/watch?v=ubPKsQ9xwjE>
- Livingston, S.A. (2020). Basic concepts of item response theory: *A Nonmathematical Introduction*. Education Testing Services.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Lawrence Erlbaum.
- McCowan, R.J. & McCowan, S.C. (1999). *Item Analysis for Criterion-Referenced Tests*. State University of New York.
- Mabumbe, M.A.G. (2022). *MANEB Examination Results Malawi*. MANEB
- Mellenbergh, G.J. (1977). *Dichotomous Decisions: Review of Education Research*. Academic press
- Mostafa, S. (2017). *Recent Developments in Systematic Sampling*. <http://doi.org/10.1080/15598608.2017.1353456>
- Mudavanhu, Y. (2019). *Quality of Literature Review and Discussions of Findings*. <https://files.eric.ed.gov/fulltext/EJ1132190.pdf>
- Mulyani, H. (2020). *Quality Analysis of Teacher-made Test in Financial Accounting Subject at Vocational High Schools*. <https://doi.org/10.21831/jpv.v10i1.29382>.
- Nguyen, T. H. (2015). *An Introduction to Item Response Theory for Patient Reported Outcome Measurement*. <http://doi.org/10.1007/540271-013-0041-0>
- Nitko, A.J. (1996). *Education assessment of students* (2nd ed). Prentice-Hall.
- Pembamoyo, S (2021). *Breaking news: MSCE results out*. <https://www.facebook.com/NationOnlineMw/posts/breaking-newsmsce-results-outthe-minister-of-education-science-and-technology-ag/4112993462054846/>

- Razak, N.A., Khairanib, A.Z, & Thien, L.M. (2012). Examining quality of mathematics test items using rasch model: Preliminarily analysis. *Procedia Journal - Social and Behavioral Sciences*, 69 (2012), 2205 – 2214
- Reddy, L. S. (2019). *Performing the KMO and Bartlett's Test for Factors Estimating the Warehouse Efficiency, Inventory and Customer Contentment for E-retail Supply Chain*. Jain University, India
- Rezigalla, A. (2022). *Item Analysis: Concept and Application*. <http://doi.org/10.5772/intechopen.100138>.
- Rini, N.I. (2009). *Item Analysis of Achievement Test in Final Test for The Seventh Grade Students of SMP N 1 Moga Pematang*. Semarang State University.
- Siedlecki, S. L (2020). *Understanding Descriptive Research Designs and Methods*. <http://doi.org/10.1097/NUR.000000000000493>
- Suckie, L. (2017). *Making Multiple Choice Test More effective*. Schreyer Institute for Teaching excellence. The Pennsylvania State University.
- Thompson, B., & Levitov, J. E. (1985). Using Microcomputers to Score and Evaluate Test Items. *Collegiate Microcomputer*, 3(1), 163-168.
- Thorndike, R. M., Cunningham, G. K., Thorndike, R. L., & Hagen, E. P. (1991). *Measurement and Evaluation in Psychology and Education* (5th ed.). MacMillan.
- Weiss, J. D, Shengyu, J. & Wang, C. (2010). *Sample Size Requirements for Estimation of Item Parameters in the Multidimensional Graded Response Model*.

APPENDICES

Appendix 1: Letter to MANEB, requesting authorization for data collection instrument

Christina Mukoko
(MED/MEV/07/21)
C/O University of Malawi
School of Education
P.O Box 280
Zomba.
6th August 2023.

The Director
Malawi National Examination Board
P.O Box 191
Zomba.

Dear Madam,

REQUEST TO USE 2020 MALAWI SCHOOL CERTIFICATE OF EDUCATION ENGLISH PAPER I

I am a student in the school of Education, Department of Education Foundations at University of Malawi, pursuing a Masters in Testing, Measurement and Evaluation. Currently, I am in my second year of study and I am supposed to work on a thesis as per requirement.

The topic of my study is “Assessing Item Response Theory Key Parameters of 2020 Malawi School Certificate of Education English Paper I”. My area of interest it’s much

on item analysis in relation to difficulty level, discrimination level and pseudo-guessing parameters.

My writing is to request that I use the 2020 Malawi School Certificate of Education English paper I. The use of the instrument and the results are strictly for research purposes.

Your consideration will be highly appreciated.

Yours Sincerely,

Christina Mukoko

Cell: 0881968838

Email: xtmukoko@gmail

Appendix 2: Authorization letter from MANEB



THE MALAWI NATIONAL EXAMINATIONS BOARD

P.O. Box 191, ZOMBA, MALAWI. TEL. (265) 01 525 277, FAX: (265) 01 525 351

e-mail: executivedirector@maneb.edu.mw

EXECUTIVE DIRECTOR

Dorothy Cynthia Nampota, *Bed Malawi, MA London, PhD Bath*

All communications should be addressed to:
The Executive Director

In reply please quote:

Your Ref:

Our Ref: C/5/1/9A

CONFIDENTIAL

28th August, 2023

Christina Mukoko (MED/MEV/07/21)
C/o University of Malawi
School of Education
P O Box 280
ZOMBA

Dear Madam,

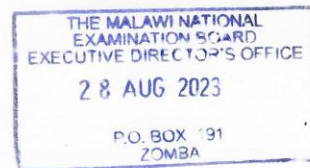
REQUEST TO USE 2020 MALAWI SCHOOL CERTIFICATE OF EDUCATION ENGLISH PAPER I

Following your request to use the Malawi National Examinations Board past question paper referred above as an instrument for data collection, I hereby allow you to use it for the requested purpose. You can source it from any secondary school.

Wishing you all the best in your academic research.

Yours faithfully,

Prof Dorothy Nampota
EXECUTIVE DIRECTOR



Appendix 3: Data collection instrument

2020

EXAMINATION NO.: _____
Page 2 of 8

M052/1

Section A (30 marks)

Answer all questions in this section by encircling the letter representing the right answer to each question.

Questions 1 – 18

In each of the following questions, choose the word(s) (A, B, C or D) that best complete(s) each of the sentences.

1. Lindiwe would come if you asked her, _____?
A. would she
B. hadn't she
C. had she
D. wouldn't she
2. They used to play together, _____?
A. were they
B. weren't they
C. didn't they
D. did they
3. We are concerned _____ drug abuse in our country.
A. for
B. about
C. on
D. by
4. I asked my father to blow _____ the paraffin lamp.
A. up
B. off
C. out
D. over
5. The headteacher was amazed _____ the performance of the school.
A. at
B. with
C. about
D. on
6. He promised to return _____ a week's time.
A. during
B. by
C. in
D. before
7. He jumped across the ditch which separated one compound _____ the other.
A. from
B. with
C. by
D. to
8. Tionge seemed pleased when she met her aunt _____ the first time.
A. to
B. for
C. in
D. on

Continued/...

9. You cannot lean ____ your parents forever.
A. against
B. at
C. to
D. on
10. Chipeta's request was _____ by the headteacher.
A. passed
B. agreed
C. proofed
D. approved
11. The duty of responsible _____ houses is to disseminate information.
A. media
B. press
C. air
D. announcing
12. He _____ good character and won the hearts of many people.
A. expressed
B. portrayed
C. showed
D. displayed
13. The hunters _____ nothing when they returned home.
A. would kill
B. would have killed
C. killed
D. had killed
14. No sooner _____ I left the place than it started raining.
A. did
B. had
C. have
D. were
15. She would go home at noon if her father _____ back in good time.
A. came
B. will come
C. has come
D. would come
16. Choso will soon _____ the challenges he is facing.
A. get out
B. get by
C. get off
D. get over
17. The boys _____ the angry mob that came to the school.
A. stood up
B. stood against
C. stood for
D. stood over
18. The company was bankrupt and workers were _____ for two months.
A. laid out
B. laid off
C. laid on
D. laid down

Continued/...

Questions 19 to 22

In each of the following questions, choose the part of speech (A, B, C or D) that best describes each of the underlined words in the sentences.

19. The lion left its kill to the scavenging hyena.
A. verb
B. noun
C. pronoun
D. adjective
20. Time is drawing near for the examinations.
A. preposition
B. adverb
C. adjective
D. verb
21. Despite the evidence against him, the man insisted on his innocence.
A. conjunction
B. verb
C. noun
D. adjective
22. The village you are looking for is beyond that hill.
A. verb
B. noun
C. preposition
D. adverb

dasascom

Questions 23 to 26

In each of the following questions, choose the order of adjectives (A, B, C or D) that best completes the sentences.

23. Students admired a _____ hut when they visited the museum.
A. large traditional Lhomwe
B. Lhomwe large traditional
C. traditional Lhomwe large
D. Lhomwe traditional large
24. Lumbani lost her _____ ring.
A. diamond very beautiful engagement
B. very beautiful engagement diamond
C. diamond engagement very beautiful
D. very beautiful diamond engagement
25. The teacher is writing on a _____ chalkboard.
A. big black classroom
B. black big classroom
C. classroom black big
D. classroom big black
26. Our teacher has bought _____ textbook.
A. a new beautiful interesting
B. an interesting beautiful new
C. a new interesting beautiful
D. an interesting new beautiful

Continued/...

Questions 27 to 30

In each of the following questions, choose the option (A, B, C or D) that best describes and gives the function of the underlined phrase or clause.

27. He labours that he may get basic needs for his family.
A. noun clause, object of 'labours'
B. noun phrase, object of 'labours'
C. adverb clause, modifying 'labours'
D. adverb phrase, modifying 'labours'
28. The prisoner revealed how he escaped from jail.
A. noun phrase, object of 'revealed'
B. adverb phrase, modifying 'revealed'
C. adverb clause, modifying 'revealed'
D. noun clause, object of 'revealed'
29. Chisomo waved at us from the inside of a telephone booth.
A. adjective clause qualifying "Chisomo"
B. adverb phrase, modifying 'waved'
C. adjective phrase, qualifying "Chisomo"
D. noun phrase, object of 'waved'
30. Although she was tired she finished the race.
A. noun clause, subject of 'finished'
B. noun clause, complement of 'finished'
C. adverb clause, modifying 'finished'
D. adjective clause, qualifying 'race'

Continued/...

Appendix 4: Permission to conduct research



VICE-CHANCELLOR
Prof. Samson Sajidu, BSc MIw, MPhil Cantab, PhD MIw

Our Ref: P.08/23/289

Your Ref.:

6th November 2023

Ms Christina Mukoko
Education Foundation Department
University of Malawi
P.O. Box 280
Zomba

Email: xtmukoko@gmail.com

Dear Ms Mukoko

**RESEARCH ETHICS AND REGULATORY APPROVAL AND PERMIT FOR
PROTOCOL NO. P.08/23/289. ASSESSING THE ITEM RESPONSE THEORY
KEY PARAMETERS OF 2020 MALAWI SCHOOL CERTIFICATE OF
EDUCATION (MSCE) ENGLISH PAPER 1.**

Having satisfied all the relevant ethical and regulatory requirements, I am pleased to inform you that the above-referred research protocol has officially been approved. You are now permitted to proceed with its implementation. Should there be any amendments to the approved protocol in the course of implementing it, you shall be required to seek approval of such amendments before implementation of the same.

This approval is valid for **one year** from the date of issuance of this approval. If the study goes beyond one year, an annual approval for continuation shall be required to be sought from the University of Malawi Research Ethics Committee (UNIMAREC) in a format that is available at the Secretariat.

Once the study is finalized, you are required to furnish the Committee and the Vice Chancellor with a final report of the study. The committee reserves the right to carry out a compliance inspection of this approved protocol at any time as may be deemed by it. As such, you are expected to properly maintain all study documents including consent forms.

UNIMAREC wishes you a successful implementation of your study.

Yours Sincerely,



Dr Victoria Ndolo
CHAIRPERSON OF UNIMAREC

CC: Vice Chancellor
Registrar
Director of Finance and Investments
Acting Head of Research
UNIMAREC Administrator
UNIMAREC Compliance Officer



Appendix 5: Letter of introduction

Zomba Urban Day Secondary
School

P.O Box 350

Zomba.

19th June 2023

University of Malawi research and ethics committee

P.O Box 280

Zomba

Dear Sir,

LETTER OF INTRODUCTION

I am a postgraduate student at the University of Malawi in the school of education, education foundations department. I am conducting a study on assessing the item response theory key parameters of 2020 Malawi .School Certificate of Education (MSCE) English paper I.

I am a holder of Bachelor of Science in Agriculture obtained from LUANAR in 2016. Currently I am teaching at Zomba Urban Day Secondary School. I am hereby kindly submitting my research proposal for review by the university of Malawi research and ethics committee.

Looking forward to your favorable consideration.

Sincerely,

Christina Mukoko

Appendix 6: Informed Consent

Informed consent form

TITLE OF STUDY

ASSESSING THE ITEM RESPONSE THEORY KEY PARAMETERS OF 2020 MALAWI SCHOOL CERTIFICATE OF EDUCATION (MSCE) ENGLISH PAPER I

PRINCIPAL INVESTIGATOR

Christina Mukoko

Education Foundations Department

University of Malawi, P.O Box 280, Zomba.

+265 (0) 881968838/ +265 (0) 999802194

med-mev-07-21@unima.ac.mw

PURPOSE OF STUDY

Kindly, you are being asked to take part in a research study, but then before you get engaged in this study, it is important that you take note the purpose of the research and what it is all about. Feel free to ask the researcher if there is anything that want to have a better understanding on. In this study, your participation will be about 45 minutes, in which 10 minutes will be used as a pre-test preparation time, 30 minutes for the test and 5 minutes for the post- test conciliation.

The purpose of this study is to assessing the item response theory key parameters of 2020 Malawi .School Certificate of Education (MSCE) English paper I. A number of factors been attributed to the poor performance of students in 2020 MSCE Examination, but so far according to the literature, none is said about the psychometric properties of 2020 MSCE English paper, hence the interest of the study.

VOLUNTARY PARTICIPATION

Note that, your participation in this study is voluntary. It is up to you to decide whether or not to take part in this study. If you decide to take part in this study, you will be asked to sign a consent form. After you sign the consent form, you are still free to withdraw at any time and without giving a reason. If you withdraw from the study before data collection is completed, your data will be returned to you or destroyed.



Appendix 7: CONFIDENTIALITY

CONFIDENTIALITY

Your responses to this study will be anonymous. Please do not write any identifying information. Will be the researcher responsibility to preserve your confidentiality.

CONTACT INFORMATION

If you have questions at any time about this study, or you experience adverse effects as the result of participating in this study, you may contact the researcher on whose contact information is provided on the first page **OR** my supervisor **Dr. S . Winiko** {+265(0) 884230956 or swiniko@unima.ac.mw}

CONSENT

I have read and I understand the provided information also have had the opportunity to ask questions. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving a reason and without cost. I voluntarily agree to take part in this study.

Participant's signature _____ Date _____

Age _____

Sex: _____

Investigator's signature _____ Date _____



Appendix 8: CHIKALATA CHAKUVOMELEZA ZA KUDZIWISIDWA ZA KAFUKUFUKU

Appendix 8: Chikalata chakuvomereza mwakudziwitsidwa za kafukufuku

MUTU WA KAFUKUFUKU

Kuyang'ana zofunikira pa mayeso a chingerezi olembedwa pakumaliza sukulu ya secondary, a chaka cha 2021

MWINI KAFUKUFUKU

Christina Mukoko
Education Foundations Department
University of Malawi, P.O Bokosi 280,
Zomba.
+265 (0) 999802194 ndi 0881968838
med-mev-07-21@unima.ac.mw

CHOLINGA CHA KAFUKUFUKU

Mukufunsidwa kutenga nawo mbali mu kafukufuku. Musanapange chiganizo chotenga nawo gawo mu kafukufukuyu, ndikoyeneka kuti mumvetsetse bwino lomwe chifukwa chomwe kafukufukuyu akuchitikira komanso zomwe adzakhudze. Chonde funsani wofufuza ngati pali china chake chomwe sichikumveka bwino kapena kufuna kumva zina zambiri. Nthawi yotenga nawo mbali ndi pafupifupi mphindi makumi asanu (50). Cholinga cha kafukufukuyu n'kuyang'ana ngati mayeso a chingerezi olembedwa pakumaliza sukulu ya secondary, a chaka cha 2021 anali ndizofuniira zina zomwe mayeso amayeneka kukhala nazo monga; mayeso amayeneka kukhala ndizoyenelera zosiyanasiyana zina mwaizo ngati kuti kuvuta kwake molingana ndi level ya mwana yemwe akuyesedwa, otha kusiyantsa mwana yemwe akudziwa ndiomwe asakudziwa zomwe zinaphunzitsidwa komanso okuti mwana asamangolota mayankho mulimonse.



KUTENGA NAWO MBALI MODZIPELEKA

Kutenga nawo mbali mu kafukufukuyu ndi mwakufuna kwanu. Muli ndi ufulu osankha kutenga

nawo mbali mu kafukufukuyu kapena ayi. Ngati mwaganiza zotenga nawo gawo mu kafukufukuyu, mudzafunsidwa kusaina fomu yovomereza. Dziwani kuti ngakhale mutasaina fomu yovomereza kutenga nawo gawo mukafukufukuyu, muli omasuka kusiya nthawi iliyonse popanda kupereka chifukwa olo kulipira kalikonse. Mukatuluka mu kafukufukuyu tisanamalize

kusonkhanitsa mayankho onse,nde kuti mayankho anu tidzakubwenzerani kapena adzawonongedwa.

KUSUNGA

CHISINSI

Mayankho anu ku kafukufukuyu sakhala odziwika kwa anthu. Chonde musalembe chidziwitso chilichonse papepalapa kuti zithandizire kusunga chinsinsi. Wofufuza adzapanga kuthekera konse kuti asunge chinsinsi chanu.

ZAMBIRI

ZAMALUMIKIZIDWE

Ngati muli ndi mafunso nthawi ina iliyonse okhuza kafukufukuyu, kapena mukukumana ndi mavuto chifukwa chochita nawo kafukufukuyu, mungalankhule ndi wofufuza yemwe mauthenga ake akupezeka patsamba loyamba KAPENA wondiyang'anira **Dr. S.Winiko** kudzera pa +265 (0)993897894 kapena swiniko@unima.ac.mw KAPENA wapampando wa UNIMAREC

Dr. Victoria Ndolo, wapampando wa University of Malawi Research Ethics Committee (UNIMAREC), P.O. Box 280, Zomba. +265 995 0427 60.

KUVOMEREZA

Ndawerenga ndipo ndikumvetsetsa zomwe zaperekedwa ndipo ndapansidwanso mwayi wofunsa

mafunso. Ndikumvetsetsa kuti kutenga nawo mbali ndikudzifunira kwandekha komanso kuti ndili ndi ufulu wosiya nthawi iliyonse, popanda kupereka chifukwa komanso popanda kulipira



kanthu. Ndikuvomera mwakufunakwanga kutenga nawo mbali mu kafukufukuyu.

Nenani ngati mukuvomera kupitirizadi kapena ayi ndi kafukufukuyi? (chongani) EYA/ AYI

Dzina la otenga nawo gawo: _____ Zaka:

Chilengedwe:wamamuna/wam`mkazi

Posayina: _____

Tsiku:

Dzina la ofufuza

:Siginecha:

Tsiku:

ZIKOMO KWAMBIRI....



Informed assent form –interviews with participants under the age 18

STUDY TITLE: “ASSESSING ITEM RESPONSE THEORY (IRT) KEY PARAMETERS OF 2020 MALAWI SCHOOL CERTIFICATE OF EDUCATION (MSCE) ENGLISH PAPER I”

INTRODUCTION OF THE RESEARCHER

Good morning/afternoon my name is CHRISTINA MUKOKO, I’ am a master’s degree student from UNIMA and I am currently conducting a study titled “ASSESSING ITEM RESPONSE THEORY (IRT) KEY PARAMETERS OF 2020 MALAWI SCHOOL CERTIFICATE OF EDUCATION (MSCE) ENGLISH PAPER I”. Therefore, I am seeking permission to interview you.

PURPOSE OF THE STUDY

The purpose of this study is to assess the psychometric parameters of 2020 MSCE English I in relation to item difficulty, item discrimination and pseudo-guessing parameters using the three parameter logistic model (3-PL model). The researcher expects you to participate for about 40-50 minutes where you will be expected to sit for a 30 minutes examination paper, following all rules of an examination and under the guidance of the invigilators.

DISCOMFORT AND RISK

I have been in contact with the Education Division Manager (EDM), the school Head Teacher (HT), Parents Teacher Association (PTA) and the teachers and they have given me a go ahead to interact with you students. Currently, I would like to find out from you, if you are willing to participate in the study which will take about 40 minutes of your time. The question are multiple choice questions and you will be required to choose only one correct option.

This research has no major risks involved, if at all anything your time and effort to participate might be the only inconvenience. Just in case you happen to lose your writing material in the course of your participation, then the researcher will replace that accordingly. Otherwise, feel comfortable to participate

BENEFITS

Much as, you would be given the opportunity to share your experiences, but then there would be no direct benefit to you. However, your contributions may help in improving assessment



procedures in Malawi assessment boards. Should be well noted that, there is neither any payment for your participation, nor any cost.

VOLUNTARY PARTICIPATION

Your participation in this study is solely voluntary. It is your decision to participate or not and if you happen to agree to take part and in the course of the process, If you happen to decide to stop, you are very free to do so and without any costs. If you happen to withdraw in the process of data collection, your data will be returned to you or destroyed.

THE PLACE OF EXAM ADMINISTRATION

The examination will be administered within your school environment and this will be done after you agree to participate. Thereafter, the exam paper will be safely kept and marked subject matter experts (SME).

CONFIDENTIALITY

Your student identity will still remain confidential and your name would not be shared with anyone outside the research team. Your name would only be recorded in the Assent Form, which would be kept separate from the examination papers, but you are kindly asked to show your responses clearly. The only exceptions for disclosing your identity and both of them are rare would be:

1. Personal information may be disclosed if required by law
2. The Human Research Ethics Committee of the University may exceptionally require personal data to respond to a formal complaint, or for a compliance audit

DURATION OF PARTICIPATION

The examination will last for about an hour.

CONTACT INFORMATION

In case you want to ask any question or there is a pertaining issue e.g. research related injuries, please feel free to contact the researcher on '0999802194/ 0881968838' you can also contact **UNIMAREC Chairperson contact details:** Dr Victoria Ndolo, Chairperson of University of Malawi Research Ethics Committee (UNIMAREC) P.O. Box 280, Zomba. +265 995 0427 60



Certificate of Assent

I have read the information above, or it has been read to me. I have had the opportunity to ask questions about it and any questions I have asked have been answered to my satisfaction. I assent voluntarily to be a participant in this study.

☐ Yes ☐ No

Do you agree to take part in the study?

☐ Yes ☐ No

Name of the respondent _____

Gender; Male /Female

Signature of respondent _____

Thumb of the participant _____

Date _____

Consent of the witness: have witnessed the accurate reading of the assent form to the potential participant, and the individual has had the opportunity to ask questions. I confirm that the individual has given assent freely.

Name of the witness; _____

Signature of the witness: _____



Appendix 9: A SUMMARY OF LOADINGS AND EIGENVALUES

Summary of loadings and eigenvalues which were extracted

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	2.715	9.049	9.049
2	1.533	5.109	14.158
3	1.352	4.508	18.666
4	1.278	4.261	22.927
5	1.225	4.085	27.011
6	1.185	3.949	30.960
7	1.109	3.697	34.656
8	1.099	3.664	38.321
9	1.072	3.573	41.893
10	1.047	3.489	45.382
11	1.028	3.428	48.811
12	1.023	3.411	52.222
13	.964	3.215	55.437
14	.962	3.207	58.644
15	.914	3.045	61.689
16	.910	3.032	64.721
17	.888	2.962	67.683
18	.872	2.907	70.590
19	.834	2.779	73.368
20	.817	2.725	76.093
21	.799	2.663	78.756
22	.798	2.658	81.414
23	.774	2.579	83.993
24	.740	2.467	86.461
25	.731	2.435	88.896
26	.704	2.347	91.243
27	.693	2.310	93.553
28	.669	2.230	95.783
29	.666	2.218	98.001
30	.600	1.999	100.000

Appendix 10: TABLE OF RESULTS AFTER REGRESSION

Three-parameter logistic model **Number of obs** = **891**

QUESTION NUMBER	a- Parameter	b-parameter
1	-3.55777	0.6852103
2	-2.640654	0.2728989
3	-0.4204256	0.4647445
4	2.956605	0.5724109
5	2.082957	1.693362
6	0.8811758	1.4143812
7	-1.47737	0.7455611
8	-3.103833	0.7590662
9	0.4038845	1.130645
10	0.1619537	0.9507745
11	-0.3918536	0.4599773
12	10.09501	0.623477
13	1.728098	0.5280338
14	-0.8344394	0.644126
15	1.040804	0.7858276
16	0.3058697	0.8029223
17	0.6104631	-1.095373
18	0.3174166	0.6104631
19	0.7504262	1.103096
20	0.598215	0.3614616
21	-1.650701	0.7284925
22	-0.9018139	0.8424148
23	-2.022716	0.8583289
24	-1.17544	0.7641266
25	-1.979889	0.9748261
26	-0.6179708	0.6864739
27	-2.311115	-0.2321558
28	-3.552896	-0.3029954
29	0.3840715	0.6825081
30	-0.0086887	0.6653274