



**INVESTIGATING CONTENT VALIDITY AND RELIABILITY OF MODIFIED
TEACHER-MADE BIOLOGY JCE MOCK TEST ITEMS FOR LEARNERS
WITH VISUAL IMPAIRMENT.**

MASTER OF EDUCATION (TESTING, MEASUREMENT AND EVALUATION)

THESIS

By

ROBIN FRANCIS CHATAIKA

Bachelor of Special Needs Education -Catholic University of Malawi

Submitted In Partial Fulfilment of the Requirements for the Degree of Master of
Education (Testing, Measurement and Evaluation)

UNIVERSITY OF MALAWI

FEBRUARY 2025

DECLARATION

I, the undersigned, declare that this thesis is my own work and that all the sources that I have used or quoted have been indicated and acknowledged by means of a complete reference.

ROBIN FRANCIS CHATAIKA

Full Legal Name

Signature

Date

CERTIFICATE OF APPROVAL

The undersigned certify that this Thesis represents the student's own work and effort and has been submitted with my approval.

Signature: _____ Date: _____

GERSON MUTALA PHIRI, PhD (Research Fellow)

Main Supervisor

DEDICATION

This research study is dedicated to my wife, Emma, for her unwavering support and love.

To my late mother, Mrs. Effita Chataika, and my late grandmother, for their enduring dedication to my upbringing and for instilling in me the importance of education.

I also dedicate this work to my beloved children; Elizabeth, Patricia, Charity, Chifuniro, Francis, and Blessings. May this work inspire you to strive for greater achievements in your educational journeys.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to all those who have contributed to the successful completion of this Thesis.

First and foremost, I would like to thank my Supervisor, Dr Gerson Mutala Phiri for his invaluable guidance, support, and encouragement throughout this research. His expertise and insights have been instrumental in shaping this work.

On a personal note, I would like to thank my family i.e. my wife Emma and my children Elizabeth, Patricia, Charity, Chifuniro, Francis and Blessings for their unwavering support and encouragement throughout my academic journey. Their love and belief in me have been a constant source of motivation.

Lastly, I am grateful to all those who have, directly or indirectly, supported and inspired me throughout this endeavor. Your contributions, no matter how big or small, have been invaluable.

ABSTRACT

This Study investigated the validity and reliability of Modified Teacher-Made Biology Junior Certificate of Education (JCE) Mock Test Items which had diagrams, pictures and tables tailored for learners with visual impairment in selected secondary schools within the South West Education Division of Malawi. The primary purpose of the study was to determine if these modified test items, maintained the validity and reliability of the original items. Findings indicated significant discrepancies: 70% of the modified items were less reliable and valid than the original versions. The inconsistency was evident in the difficulty levels of the modified items, with some items proving more difficult and others less difficult than the original items. Additionally, item characteristic curve analysis revealed that some modified items were overly challenging even for learners without visual impairments. Reliability analysis highlighted poor performance in some items, with low discrimination indices (e.g., QM10 = 0.08, QM19 = 0.19). Content validity assessments demonstrated that some modified items did not measure the same concepts as their original items for example; they were derived from different topics and objectives. For instance, question QM20 was not aligned with the original item's topic and objective. The study concluded that most modified test items failed to maintain the original difficulty range, making them unreliable. The findings underscored the need for comprehensive guidelines and enhanced in-service training to develop valid and reliable test items for learners with visual impairments, ensuring equitable assessment practices.

TABLE OF CONTENTS

ABSTRACT.....	vi
TABLE OF CONTENTS	vii
LIST OF FIGURES	xi
LIST OF TABLES	xii
LIST OF APPENDICES	xiii
LIST OF ACRONYMS AND ABBREVIATIONS	xiv
CHAPTER ONE	1
INTRODUCTION.....	1
1.0 Chapter Overview	1
1.1 Background to the Study	1
1.2 Statement of the Problem.....	2
1.3 Preliminary Findings.....	3
1.4 Purpose of the Study	4
1.5 Research Objectives.....	5
1.5.1 Main Research Objective	5
1.5.2 Specific Research Objectives.....	5
1.6 Significance of the Study	5
1.7 Limitations and delimitations	6
1.8 Definitions of Operational Terms	6
1.9 Chapter Summary	8
CHAPTER TWO	9

LITERATURE REVIEW	9
2.0 Chapter Overview	9
2.1 Definition of a test and its functions	9
2.2 Test item construction.....	11
2.2.1 Test item analysis.....	11
2.2.2 Test item modification	12
2.2.3 Test validity	15
2.2.4 The reliability of the modified test items.....	17
2.2.5 Test item difficulty index.....	19
2.2.6 Test item Discrimination Index	22
2.3 Theoretical framework.....	24
2.3.1 Item Response Theory	25
2.3.2 The One-Parameter Logistic (1PL) Model	25
2.3.3 The Two-Parameter Logistic (2PL) Model.....	25
2.3.4 The Three-Parameter Logistic (3PL) Model.....	26
2.3.5 Item Characteristic Curve	27
2.4 Chapter Summary	29
CHAPTER THREE	30
METHODOLOGY	30
3.0 Chapter Overview	30
3.1 Research Paradigm.....	30
3.2 Design of the Study.....	31
3.3 Study population	32
3.4 Sampling technique and sample size	33
3.5 Research data management and dissemination.....	35

3.7 Data Generation and Instrumentation	37
3.9 Ethical Considerations.....	39
3.10 Chapter Summary.....	41
CHAPTER FOUR.....	42
RESULTS AND DISCUSSIONS OF THE FINDINGS	42
4.0 Chapter Overview	42
4.1 Reliability Analysis of Difficulty Indices for Modified Biology JCE Mock Test Items for Learners with Visual Impairment	42
4.2 Exploring the effectiveness of modified Biology JCE Mock test items in assessing learners with and without visual impairments using item characteristic curves.....	43
4.3 Assessing the Reliability of Modified Biology JCE Mock Test Items for Learners with Visual Impairment: Analyzing Item Difficulty Index.....	49
4.3.1 Analysing Less Difficult Modified Test Items with Diagrams, Pictures, and Tables	51
4.3.2 Analysing More Difficult Modified Test Items with Diagrams, Pictures, and Tables	52
4.3.3 Analysing Same Difficult Index Modified Test Items with Diagrams, Pictures, and Tables.....	54
4.4 Analysis of Item Discrimination in Modified Biology JCE Mock Test Questions.....	56
4.5 Assessing the Guessing Level in Modified Test Items.....	58
4.6 Evaluating the Content Validity of Modified Biology JCE Mock Test Items for Learners with Visual Impairments	59
4.6.1 Content Validity Based on Comparative Analysis (CA) of Modified Test Items and Original Test Items: An Examination with Diagrams, Pictures, and Tables	60

4.6.2	Evaluating Content Validity through Expert Review: Insights from Subject Matter Experts and Specialist Teachers for Learners with Visual Impairment.....	67
4.7	Chapter summary	73
CHAPTER FIVE		74
CONCLUSION AND RECOMMENDATIONS FOR FURTHER STUDIES.....		74
5.0	Chapter Overview	74
5.1	Findings of the Study	74
5.2	Conclusions.....	75
5.3	Recommendations for the Study’s contribution to knowledge.....	76
5.4	Proposed Recommendations for Further Research Studies	77
5.5	Chapter Summary	77
REFERENCES.....		78
APPENDICES		87

LIST OF FIGURES

Figure 2.1 Item Characteristic Curves	27
Figure 4.1 Reliability on difficulty and discrimination on ICCs for test items QM8, QM9, QM14, QD8, QD9 and QD14.	44
Figure 4.2 Reliability on difficulty and discrimination of ICCs for test items QM6, QM18, QD6, and QD18 at 0.5 p-value.	46
Figure 4.3 Reliability on difficulty and discrimination on ICCs for test items QM13, QM20, QD13, and QD20 at 0.5 p-value.	47
Figure 4.4 Reliability on difficulty and discrimination on ICCs for test items QM10, QM19, QD10, and QD19 at 0.5 p-value.	48
Figure 4.5 Reliability on difficulty and discrimination on ICCs for test items QM5 and QD5 at 0.5 p-value.	49
Figure 4.6 Test item difficult level	55
Figure 4.7 Summary of the content validity based on comparative analysis (CA)	66
Figure 4.8 Knowledge in test item modification for test items with DPTs for learners with VI.	69
Figure 4.9 I learnt test item modification for learners with visual impairment in college.	70
Figure 4.10 Have knowledge in test item modification.	71
Figure 4.11 it is important to have knowledge in test item modification.	72

LIST OF TABLES

Table 1.1 Biology 2022 mock examination summary results for learners with and without VI for secondary schools A and B.	4
Table 2.1: Performance of learners who use braille in science subjects.....	14
Table 2.2: Item difficulty indices interpretation	21
Table 2.3: The discrimination values and their corresponding interpretation and recommendations (Zhao, 2006).	23
Table 3.1 Learners with visual impairment	32
Table 4.1 Reliability of Difficulty and Discrimination on Latent Trait for Items QM8, QM9, QM14, QD8, QD9 and QD14 at a 0.5 p-value.....	43
Table 4.2 Reliability of Difficulty and Discrimination on Latent Trait for Items QM6, QM13, QM18 and QM20 at a 0.5 p-value.....	45
Table 4.3 Reliability of Difficulty and Discrimination on Latent Trait for Items QM5, QM10, and QM19 at a 0.5 p-value	47
Table 4.4 A comparison analysis of p-values for Item Difficulty between Modified and Original Items	50
Tables 4.5 Frequency of QM19 and QD19 Scores	53
Table 4.6 Percentage Distribution of Items Across Same, More, and Less Difficult Ranges.....	54
Table 4.7 Item Discrimination Table Sorted in Ascending Order	56
Table 4.8 Guessing Parameter of Logistics Model	58
Table 4.9 Data Test of Normality	60
Table 15: Table 4.10 Goodness-of-Fit Analysis	62
Table 4.11 Parameter Estimates for the Statistical Model.....	63
Table 4.12 Mean Statistics for Comparative Analysis (CA) of Modified and Original Test Items.....	64
Table 4.13 Content Validity Analysis Based on the SMEs and the STLVI	68
Table 4.14 Knowledge on Test Item Development and Modification (TDM).....	69

LIST OF APPENDICES

APPENDIX 1: Modified Teacher-made Biology JCE Mock Test Items.	87
APPENDIX 2: Original Biology JCE Mock Test Items with diagrams, pictures and tables.	89
APPENDIX 3: Likert Scale Instrument.....	91
APPENDIX 4: University of Malawi Research Ethics and Regulatory Committee Approval and Permit for Protocol No. P.07/23/276	93
APPENDIX 5: Department of Education Foundation Letter of Introduction: Robin F. Chataika	95
APPENDIX 6:A letter requesting permission for data collection to the South West Education Division Manager (EDM)	96
APPENDIX 7: Authority to Conduct Research in South West Education Division (SWED).....	97

LIST OF ACRONYMS AND ABBREVIATIONS

BSS	: Blantyre Secondary School
CA	: Continuous Assessment
DIF	: Differential Item Functioning
DPT	: Diagram, Picture and Table
EDMs	: Education Division Managers
GoF	: Goodness of Fit
ICC	: Item Characteristic Curve
IE	: Inclusive Education
IQ	: Intelligence Quotient
IRT	: Item Response Theory
JCE	: Junior Certificate of Education
MANEB	: Malawi National Examinations Board
MTMBTIs	: Modified Teacher-Made Biology Test Items
NSS	: Ngabu Secondary School
PLM	: Parameter Logistic Models
QD	: Question with Diagram
QM	: Modified Question

SMEs	: Subject Matter Experts
SNE	: Special Needs Education
SNS	: Special Needs Specialist
STLWVI	: Specialist Teachers for Learners with Visual impairment
SWED	: South West Education Division
TDM	: Test Development and Modification
UNIMAREC	: University of Malawi Research Ethics Committee
VI	: Visual Impairment

CHAPTER ONE

INTRODUCTION

1.0 Chapter Overview

This Chapter discusses the background of Modified Teacher-Made Biology Test Items (MTMBTI) for Biology Junior Certificate of Education (JCE) Mock that were in Diagram, Picture and Table (DPT) format and were modified to textural statements which were converted into Braille in an inclusive education setting. In addition, it discusses the statement of the problem, purpose of the study, main research objective, specific research objectives, and significance of the study, limitations and delimitations; and definitions of operational terms.

1.1 Background to the Study

The study investigated content validity and reliability of Modified Teacher-made Biology JCE Mock test items for learners with visual impairment. This was a case of selected secondary schools in the South West Education Division (SWED) in Malawi. In Inclusive Education (IE), there are learners with diverse needs and one of which are learners with visual impairment. Inclusive Education (IE) is a process of addressing and responding to the diverse of needs of all learners through increasing participation in learning and reducing exclusion within and from education (UNESCO, 2005). These

learners are categorized into three groups i.e. learners with low vision and use large print, learners with low vision and they use Braille and learners with complete loss of light perception and they use Braille. The researcher concentrated on the latter category. The test items which have diagrams, pictures and tables in the original print copy are modified into text format then converted into Braille. So the process of modifying such test items requires subject matter expertise to ensure test validity and reliability, hence the modification of such items was investigated. For ensuring reliability, the study looked at the test items written by learners with visual impairment if they have the same weight (i.e. same difficulty and discrimination level) to their friends without visual impairment. In the process of investigating the validity of MTMBTIs the researcher looked at the test items if they were developed from the same topic and objective as the original items. This was to ensure that both learners with and without visual impairments had items which were taken from the same topic and objectives.

1.2 Statement of the Problem

Biology is one of the science subjects offered in secondary schools. In a regular classroom set up, some Biology test items do have some diagrams, pictures and tables. This also applies in an Inclusive Education (IE) class. Learners with visual impairment always have problems in understanding concepts that are presented in diagrams, pictures and tables. In view of this such test items with diagrams, pictures and tables are modified into text format for these learners with visual impairment. Biology relies a lot on sight and touch in both theory, practical lessons and assessments. The visually impaired learners must overcome many obstacles involving sight if they are to succeed from Biology test instructions. As a result of that the subject teachers do modify such test items

that have diagrams, pictures and tables. Therefore, there is need to investigate the validity and reliability of the modified test items in terms of difficulty level and discrimination if they maintain the same to the original items. In addition, if the modified test items also are developed from the same topic and objectives of the original items.

Test modifications can change the way in which test items are presented to the learner's method of responding to test items or the process a learner uses to derive responses to test items (Elliott et al., 2010). Therefore, modifications to test items can significantly impact both the level of difficulty and the interpretation of questions, which ultimately affects the validity and reliability of assessment outcomes. Allman, C. (2009), states for instance, when a complex diagram is converted into a textual description, it may inadvertently provide hints regarding the answer or necessitate the use of additional cognitive skills that are not pertinent to the specific construct being evaluated. This could lead to confusion or misinterpretation of the intended assessment goals.

When administering test items with diagrams, pictures and tables to visually impaired learners, it is vital to modify them appropriately so that they maintain the validity and reliability in relation to the original items. Failure to do so, may lead to learners being assessed on different constructs and level of difficulty and discrimination from their sighted peers which could affect the accuracy of the scores and impact the content validity of the test.

1.3 Preliminary Findings

The preliminary findings on the scores of 2022 Biology JCE Mock test results for learners with and without visual impairment at Secondary School (A) and Secondary

School (B) were noted that at secondary school (A) the score range for learners without visual impairment (WVI) was 48-93 while those with VI was 26 - 56. Similarly, at secondary school (B), it was noted that the score range was 41 - 81 for learners WVI while those with VI it was 17-58.

Table 1.1 Biology 2022 mock examination summary results for learners with and without VI for secondary schools A and B.

School	Sighted learners	VI learners	Total	Without VI score range	With VI score range
A	107	5	112	48-93	26-56
B	84	3	87	41-81	17-58

Therefore, this study investigated the validity and reliability of Modified Teacher-Made Biology Test Items (MTMBTI) that had pictures, tables and diagrams (DPT) for learners with visual impairment if they had the same difficulty level and discrimination.

1.4 Purpose of the Study

The purpose of the study was to investigate the content validity and reliability of Modified-teacher made Biology Test Items (MTMBTIs) for learners with visual impairment. Biology test items contain some diagrams, pictures and tables which are too difficult for learners with visual impairment to understand. Learners with visual impairment might not be able to make the drawings or label them. Such test items with diagrams, pictures and tables are modified into text format where learners with visual impairment could easily read and understand them when writing examinations or tests.

Therefore, the test items should conform to content validity and reliability for same difficulty level and discrimination as it was in the original print copy.

1.5 Research Objectives

1.5.1 Main Research Objective

To investigate content validity and reliability of teacher-made modified Biology JCE Mock test items for learners with visual impairment.

1.5.2 Specific Research Objectives

The main research objective was addressed by the following specific objectives:

- i. To analyse the difficulty level of modified Biology JCE Mock test items
- ii. To explore the effectiveness of modified Biology JCE Mock test items in assessing learners with and without visual impairments using item characteristic curves.
- iii. To investigate the reliability of modified Biology JCE Mock test items for learners with Visual Impairment.
- iv. To investigate content validity of modified Biology JCE Mock test items for learners with Visual Impairment.

1.6 Significance of the Study

The Study would help to inform and guide the subject matter experts (SMEs) and specialist teachers for learners with visual impairment (STLWVI) to be consistent on modification of test items that have diagrams, pictures and tables into Braille for learners with visual impairment. It would also help to guide policy makers to come up with the blueprint for test item modification for learners with visual impairment. In addition, the

study would guide teacher training institutions to develop guidelines for modification of Biology test items that have diagrams, pictures and tables into text format for learners with visual impairment. It would help teachers graduating from the teacher training institutions to have relevant skills in modifying such test items that have diagrams, pictures and tables into Braille.

Furthermore the study would inform examining institutions to ensure that modification of test items that have diagrams, pictures and tables into text format for Braille are valid and reliable as were in the original print copy. The Study would help to improve the credibility of examinations taken by learners with visual impairment, hence improving the quality of tests and examinations in inclusive education.

1.7 Limitations and delimitations

The Study was conducted in the sampled secondary schools in the South West Education Division (SWED), and not in all the six education divisions in Malawi. Therefore, the results of the Study might not be generalized to all secondary schools in the country. Furthermore, only the Form Two learners participated in the administration of Modified Teacher-Made Test Items for Biology JCE Mock. In addition, the Study used Biology which is one of the subjects taught at JCE. .

1.8 Definitions of Operational Terms

a-Parameters: A statistic measure that gives an ability level of an item to discriminate learners with high ability levels from those with lower ability levels (Siri & Freddano, 2011).

b-Parameters: A statistic measure that gives a level of difficult of an item (Siri & Freddano, 2011).

Difficulty level: The percentage of examinees that answered the item correctly.

Discrimination Index (D): This is an instrument to measure the difference in item difficulty between groups of learners with high and low marks (Macdonald & Paunonen, 2002).

Item analysis: A process which evaluates responses of learners to individual test items in order to assess their quality and the quality of the test as a whole (Siri & Freddano, 2011).

Item Characteristic Curve (ICC): This is a curve that is used to present psychometric properties of test items (Philip & Ojo, 2017).

Item Response Theory (IRT): This is a measurement framework used in the design and analysis of educational and psychological assessments (i.e. achievement tests, rating scales, inventories, or other instruments) that measure mental traits (Ogunsakini & Shogbesani, 2018).

Reliability of a test: Is the degree to which a test is consistent, stable, dependable or trustworthy in measuring what it is measuring (Osuji and Okonkwo, 2006).

Subject matter experts: This is an expert with special skills, knowledge, and experience in a particular field of study. (Hopkins & Unger, 2017).

Test: This is a detailed or small scale task carried out to identify the candidate's level of performance (Manichander, 2016).

Test item modifications: These are changes in testing procedures or formats that provide learners opportunity to participate in testing procedures (Elliott et al. 2010).

Validity: This is the extent to which a test measures what it claims to be measuring, the extent to which it is possible to make appropriate inferences from the test score (Coaley, 2010).

1.9 Chapter Summary

The Chapter highlighted the background of the study; statement of the problem; purpose of the study; main research objective; specific research objectives; significance of the study; limitations and delimitations; and definitions of the operational terms.

CHAPTER TWO

LITERATURE REVIEW

2.0 Chapter Overview

This Chapter provides definitions of a test, test construction, test item analysis, test item modification, test validity and reliability of modified test items. In addition, it discusses the IRT logistic models; test item difficulty index, test item discrimination index and pseudo-guessing. It also discusses the classical test theory and theoretical framework.

2.1 Definition of a test and its functions

A test is thus one of the assessment instruments used in different institutions such as schools. It is used in getting quantitative data. Assessment is important for evaluating a learner's performance and understanding. It provides useful understanding of a learner's knowledge of the material taught and their ability to apply acquired knowledge in real situations (Manichander, 2016). Tests are conducted to measure the person's ability in performing certain tasks after learning has taken place.

A test is an instrument for assessment in schools which serves different purposes and functions. One of the purposes of assessment is to help decide how well learners have learnt a given content or how far the objective earlier set out has been achieved quantitatively (Abdulmalik, 2013).

Data obtained in assessment or tests serve various educational functions in the school such as: the classroom function which determines the level of achievement, effectiveness of the teacher, teaching methods and instructional materials. It also motivates a learner when is successful and it is used to predict learners performance in new situations (Abdulmalik, 2013).

The assessment also provides guidance function in a situation where it can provide the teacher with diagnostic data about individual learners in his class in terms of learners' strengths, weaknesses and interests. On the other hand, assessment serves the administrative function where it serves as communication of information when data collected are used in reports to parents. In addition, it helps in making appropriate decisions and recommendations on curricula packages and curricula activities. On another note, a test could form the basis upon which streaming, grading, selection and placement are based (Abdulmalik, 2013). The later function of a test is what happens in secondary schools. For example, when a learner passes the form two examinations then he or she is placed into the senior section i.e. form three where he proceeds to form four. After doing extremely well in form four then the learner is graded and selected for university placement. This is why the researcher was interested to investigate the validity and reliability of modified test items' difficulty level and discrimination for Biology JCE Mock for learners with visual impairment. It was from that understanding of test functions that would ensure that test modifications should be done while upholding the principles of test item construction.

2.2 Test item construction

Test item construction refers to the process of well-crafted test items that their scores provide valid inferences about examinee's mental attributes such as achievement, ability, and aptitude whereby the items must reflect a specific psychological construct or domain of content (Osterlind, 2002). As such, the modified Biology test items that had diagrams, pictures and tables must maintain same attributes as they were in the original print copy.

There are four major issues in test item construction and one of which is the presentation of methods for determining the quality of test items. This may be categorized into two i.e. procedures for gauging the proper content for test items, which revolve around concerns of validity. Secondly, the procedures for examining test items for either random errors or systematic bias, which reflect considerations of reliability (Osterlind, 2002). Both of these issues are addressed either by judgmental procedures or statistical models.

2.2.1 *Test item analysis*

Test item analysis refers to the process by which test items are examined and critically analyzed for their validity, reliability and level of difficult (Osterlind, 2002). Its purpose is to identify and reduce the sources of error in measurement. Test item analysis is conducted to gauge the quality of test items and discard those which are unacceptable, repair those which can be improved, and retain those which meet criteria of merit (Osterlind, 2002). The researcher focused at the teacher-made Biology JCE Mock test items that had diagrams, pictures and tables which were modified into text format if they were valid, reliable and of the same difficult level.

There are two ways of conducting test item analysis and these are numerical and judgmental analysis. The numerical test item analysis is whereby the statistical properties of particular test items are examined in relation to a response distribution. In this, the test items are done in a field trial for examination development (Osterlind, 2002). The primary purpose for field trials of test items is to collect appropriate data for reviewing them.

Judgmental item test analysis approach analyses test items by asking subject matter experts to comment on particular test items if they are valid and reliable (Osterlind, 2002). The researcher conducted the judgmental approach on the modified Biology test items with diagrams, pictures and tables to find out whether items were testing the same content as they were in the original print copy. The researcher involved three Biology Subject Matter Experts from the school A, in the judgmental approach and used the Likert Scale instrument. In addition, the researcher administered the teacher-made modified test items to the Form Two learners in the sampled schools in order find the reliability of the modified items.

2.2.2 *Test item modification*

Test item modifications are changes in testing procedures or formats that provide learners with visual impairment an equal opportunity to participate in testing procedures and to demonstrate their knowledge and abilities (Elliott et al. 2010). For example, test items in Biology comprise of diagrams, pictures and tables which are difficult for learners with visual impairment who use Braille writing code. It is difficult for the learners to tactically

follow the diagrams, pictures and tables and may not be able to draw them in Braille using their Perkins Braille Machines or Hand Frame.

Test modifications can change the way in which test items are presented to the learner, the learner's method of responding to test items or the process a learner uses to derive responses to test items (Committee on Special Education, 2016). Similarly, Biology print test items that have diagrams, pictures or tables are modified into Braille text format for learners with visual impairment to access them with a better understanding of the content and constructs.

Modification of teacher-made test items for Biology JCE Mock that have diagrams, pictures and tables or any other subject area is done by the Subject Matter Experts (SMEs) who are the subject teachers such as Biology teachers. A subject-matter expert has special skills, knowledge, and experience in a particular field of study like in this case Biology. The SMEs provide the knowledge and expertise in a specific subject and technical areas for any assignment (Hopkins & Unger, 2017). However, test item modification into Braille for learners with visual impairment is done with advice from the specialist teachers for learners with visual impairment. These specialist teachers have a specific qualification that is over and above their initial teaching qualification in order to develop and deliver specialized educational programs for learners with visual impairments (McLinden, et.al. 2017).

Therefore, the SME provides the expertise of the knowledge of the subject matter and content. On the other hand, specialist teacher for learners with visual impairment provides the technical know-how of the adaptability and modification of the structural

test items which can easily be approached by learners with visual impairment (McLinden, et.al. 2017).

Habulezi, J. et al. (2017) provide a thorough analysis of the unsatisfactory academic performance displayed by learners with visual impairments in science subjects, a concern that has become increasingly widespread in schools across Botswana. The data on the performance of learners with visual impairments who used Braille from 2010 to 2016 reveals troubling trends.

As shown in Table 2.1, learners using Braille in Botswana encounter similar challenges to those faced by their counterparts in Malawi, as indicated in the preliminary research findings. Alarming, the pass rates for learners with visual impairments in Botswana during this period consistently fall below acceptable levels.

Table 2.1: Performance of learners who use braille in science subjects

Year	# of learners sat for exam	number passed	number failed	pass %	fail %	credit pass # (A-C)	quantity pass # (A-E)
2016	9	0	9	0	100	0	0
2015	11	1	10	9	91	0	1
2014	8	2	6	25	75	0	2
2013	6	2	4	33	67	0	2
2012	5	1	4	17	83	0	1
2011	4	1	3	25	75	0	1
2010	3	1	2	33	67	0	1

Source: Special Education Department, 2017

Additionally, Habulezi et al. (2017) identify one of the primary reasons for this poor performance in science subjects as the insufficient adaptation and modification of educational materials. Many test items with diagrams, pictures, and tables that have not

been adequately modified to formats accessible to visually impaired learners. This lack of appropriate modification becomes particularly evident when these learners attempt to answer test items that include visual elements; significantly hindering their understanding and performance hence it affects the validity and reliability of such modified items.

Stone et.al (2010) investigated the impact of test accommodation on an English Language assessment for learners with visual impairment. Their study examined differential item functioning (DIF) using the Mantel-Haenszel method. Only one item at each grade was flagged as displaying large DIF, in each case favouring learners without disabilities. The results identified areas for improvement mainly in formatting and consistency. On the other hand, the current study Therefore, the current study went further to explore the validity and reliability of modified teacher-made Biology test items that have diagrams, pictures and tables. This determined the difficulty level and discrimination of modified teacher-made test items. The well modified items are expected to assess the same skill and have equal value and validity. In support to the same, Allman, (2009) states modified items must maintain the correct answer in the same position as that of the original test item.

2.2.3 Test validity

Test validity is the extent to which a test measures what it claims to be measuring, the extent to which it is possible to make appropriate inferences from the test score (Coaley, 2010). It focuses at the accuracy of a measure. It is further stated that it is about the significance of validity and its models, including content, construct and criterion-related validity, as well as about subordinate forms such as convergent discriminant validity.

There are four main types of test validity which are construct validity, content validity, face validity and criterion validity (Middleton, 2022). Therefore, the study focused on content validity.

Content validity is a crucial aspect of test development that plays a pivotal role in defining the performance domain of interest. This typically involves the selection of a panel of qualified experts in the relevant content domain who can provide invaluable feedback on the test items (Crocker & Algina, 2008). Crocker & Algina, further explain the process of content validation as provides a structured framework for the thorough process of matching test items to the performance domain, as well as collecting and summarizing the data generated from the matching process.

There are two standards for ensuring content validity which are the sampling of the items and the method of constructing the items. These two standards use two types of judgements i.e. the measure of the extent of each item for defining the traits; and the set of items that represent all aspects of the traits (Yaghmale, 2003). As such, the researcher investigated content validity to address the objectives of the modified teacher-made Biology JCE Mock test items if they were the same to the original items that had diagrams, pictures and tables. In addition, the researcher assessed the items difficulty level and item discrimination.

The purpose of investigating content validity is to determine whether the items adequately represent a performance domain or construct of specific interest. In content validity, a typical procedure is to have a panel of independent experts to judge whether the items adequately sample the domain of interest (Crocker & Algina, 2008). For example, the objectives from which the test items are derived. In this case, the researcher

involved the Biology Subject Matter Experts who made a judgement using the Likert Scale instrument.

2.2.4 The reliability of the modified test items

Assessment performance on modified test items for learners with visual impairments must support the same inferences regarding learner proficiency as those derived from test items given to learners without visual impairments (Winter, et al., 2019). This underscores the importance of ensuring that modified test items adhere to comparable levels of reliability on difficulty and discrimination as the original items. Moreover, it is essential that these modified items are constructed from the same topics and objectives, thereby measuring the same skills as their original items with diagrams, pictures and tables.

However, (Winter et al., 2019) findings exposed that although given the typically small size of special groups of English learners, such as blind and low vision learners, traditional calibration and item linking techniques were often incapable of ensuring sufficient levels of comparability. Therefore, this provides basis for ensuring that modified items are reliable and valid.

Reliability of a test may be defined as the degree to which a test is consistent, stable, dependable or trustworthy in measuring what it is measuring (Osuji and Okonkwo, 2006). Therefore, reliability is the consistency of a measure of test items.

There are different methods that explain reliability. Some of them are Reliability as Equivalence, Reliability as Stability, Reliability as Internal Consistency, Split-half, Kuder-Richardson-20 and 21 (KR-20 and 21) and Cronbach's alpha (Bichi, 2016)).

Coefficient alpha also called Cronbach's alpha is widely used to measure reliability coefficient. It estimates the reliability of a test-score from a single test administration using information from the relationship among test items. It provides a reliability estimate based on the covariation among internal items and is also called an internal-consistency coefficient (Webb, et.al, 2006).

Internal-consistency reliability involves a full length test, given on a single sitting which is divided into parallel parts (Cronbach, 2004).

Cronbach's Alpha formula is claimed as one of the best analysis method that can be used to gauge the reliability of educational and psychological measurements and it assesses the consistency of scores from one condition to another. Procedures like alpha are known as internal consistency analyses (Bichi, 2016)). The reliability is computed with coefficient alpha, defined as:

$$\alpha = \left(\frac{K}{K-1} \right) \left(1 - \frac{\sum S_i^2}{S_x^2} \right) \text{ Source: (Bichi, 2016)).}$$

where: k : represents number of items on the test;

$\sum S_i^2$ sum of the variances of the different parts of the test (item i) and

S_x^2 variance of the test scores (Bichi, 2016)).

Reliability is the extent to which test scores are not affected by chance factors like the specific questions or problems that were on the edition of the test as compared with those on other test editions (Livingston, 2018). For example, a test may be affected by modification such as the one in the modified teacher-made test items for Biology JCE

Mock for learners with visual impairment. This is why the researcher was interested in investigating the reliability of such modified Biology test items for learners with visual impairment as compared to those that were in the original print copy edition and establish their reliability.

Ayanwale & Ndlovu, (2021) investigated the scalability of a modified cognitive multiple-choice test. It was observed that the test could not be scaled unidimensionally due to the low scalability of some items and the results were recommended for further modification in order to provide monotonic characteristics.

Therefore, this study aimed at focusing on content validity, reliability, difficulty level and discrimination of the modified teacher-made test items that had diagrams, pictures and tables for learners with visual impairment hence the study was carried out.

2.2.5 *Test item difficulty index*

Research indicates that examinations with an excessive number of items that are either too easy or too difficult can negatively impact their validity and reliability (Watering, 2006). In view of this, Watering (2006) further suggests, this problem can be resolved by analyzing item difficulty, distractor effectiveness, and discrimination among the items.

The validity and reliability of examination can be adversely impacted by the mismatch between the level of cognition in the assessment and the educational task. This mismatch can appear in the form of too many easy or difficult items (Rezigalla, et.al, 2020).

To ensure high-quality assessment and evaluation, it is crucial to consider the level of difficulty of examination items, the effectiveness of distractors, and the discrimination among the items. Therefore, the current study investigated the validity and reliability of

Modified Teacher-Made Biology Mock Test Items for learners with visual impairment through item difficulty index and discrimination.

In computing test item difficulty level, items are dichotomously scored such as **1** representing the right answer or **0** representing the wrong answer (Philip & Ojo, 2017).

The difficulty index is the proportion of examinees that answer the item correctly:

$$\text{Item difficulty Index} = \frac{\text{number of candidates that got the item right}}{\text{total number of candidates}}$$

The *p* (proportion) value ranges from 0 to 1. When multiplied by 100, *p-value* converts to a percentage, which is the percentage of learners who got the item correct. The higher the *p-value*, the easier the item (Hotiu, 2006). This means the higher the difficulty index, the easier the item is understood to be. Those with a *p-value* between 20 and 90% are considered as good and acceptable. This is interpreted as:

Range	Inference to item
0.85 – 1.00	Very Easy
0.70 – 0.84	Easy
0.30 – 0.69	Optimum
0.15 – 0.29	Hard
0.00 – 0.14	Very Hard

(Philip & Ojo, 2017).

Selection of test items is a vital step in test development procedure. Test items could be selected using Item Characteristic Curve (ICC). In recent times ICC, has become a vital tool in the selection of test items as it shows graphically the psychometric properties of the test items. At a glance the curve gives such information as difficulty, and discriminating guessing value depending on the model (Philip & Ojo, 2017). This works on modeling of 2-parameter model of Item Response Theory (IRT) to generate the item

characteristics curve using a statistical software package for data analysis (STATA). In addition, windows Microsoft excel application was used. The ICC curve is a vital tool in determining the suitability of item for selection in a test.

Bichi, (2016) explains item difficulty index is the proportion of examinees taking the test, who get an item or answer it correctly. It further explains the larger the percentage of getting an item correctly, the easier the item is and the less the percentage the more difficult the item is. To compute item difficulty index, divide the number of examinees answering the item correctly by the total number of examinees answering the item. An item answered correctly by 75% of the examinees would have a difficulty index or *p-value*, of .75, whereas an item answered correctly by 40% of the examinees would have a lower item difficulty or *p-value*, of .40 (Bichi, 2016).

For example, Bichi, (2016) provides a general guideline for the interpretation of an item difficulty index as is provided in the table below:

Table 2.2: Item difficulty indices interpretation

Difficulty Index (p)	Interpretation
$P \leq 0.30$	Difficult
$0.31 \leq 0.70$	Moderately difficult
$P > 0.70$	Easy

(Bichi, 2016)

The item difficulty is denoted as *p* and is symbolically given as:

$$P = \frac{R}{N}$$

where P = is the difficulty of a certain item

R = is the number of examinees who get that item correct and

N = is the total number of examinees.

Therefore, the IRT model analysis investigated item test difficulty level and how modified teacher-made test items were able to discriminate between learners in the upper group and those in the lower group of intelligence quotient.

2.2.6 *Test item Discrimination Index*

Macdonald & Paunonen, (2002) explain Discrimination Index (D) is an instrument to measure the difference in item difficulty between groups of learners with high and low scores. It indicates the extent of an item to differentiate learners with different ability levels. Discrimination indices of each test item provide information regarding what the learners have learned and enable teachers to determine and correct the faulty test items (Azzopardi & Azzopardi, 2019). It provides a valuable tool in designing the test.

The discrimination index varies between -1 and 1, where the item should have a positive discrimination index of at least 0.2. If the item equals to 0, it means that there is no discrimination (Macdonald & Paunonen, 2002). Test items with negative indices need to be revised and here is one of the examples how discrimination index is calculated:

$$\text{Discrimination index} = \left[\frac{\sum H - \sum L}{N(\text{Score}_{\max} - \text{Score}_{\min})} \right]$$

Source: Johari et al.,

2011

H = total score for 25% of learners in the high achievement group.

L = total score for 25% of learners in the low achievement group.

N = 25% of total numbers of learner tested.

$Score_{max}$ = maximum (full) marks for the item.

$Score_{min}$ = minimum marks for the item (Macdonald & Paunonen, 2002)

It implies that if the test item measures the same ability or competence, it is expected to have those in a high overall test score to have a high probability of being able to answer the test item. Therefore, a good test item should discriminate between those who score high on the test and those who score low (Azzopardi & Azzopardi, 2019).

Table 2.3: The discrimination values and their corresponding interpretation and recommendations (Zhao, 2006).

Discrimination Index	Description	Recommendations
$D = \text{negative}$	Defective Item	Rejected or improved
$D < 0 - 0.19$	Poor discrimination	Poor items to be rejected
$D \text{ between } 0.2 - 0.29$	Acceptable discrimination	Marginal items usually need and subject to improvement
$D \text{ between } 0.3 - 0.39$	Good discrimination	Reasonably good but subject to improvement
$D = 0.4$	Very good discrimination	Very good items; accept
$D > 0.4$	Excellent discrimination	Very good items; accept

(Zhao, 2006)

Discrimination index is also calculated using the following formula as indicated below:

$$\text{Discriminating index} = \frac{Up - Lp}{\text{total number of test takers in the upper group}}$$

Up = number of test takers in the upper group that got the item right
 Lp = number of test takers in the lower group that got the item right. (Philip & Ojo, 2017)

Candidates' scores need to be sorted either in descending or ascending order of the total score when computing discrimination index using Microsoft Excel (Philip & Ojo, 2017).

Therefore, test items between the original print copy and the modified ones must be of the same level of difficulty. They should maintain the originality as it is in the original print paper.

2.3 Theoretical framework

Item Response Theory (IRT) is founded to be the theoretical framework for the study. IRT comprises a set of statistical models for measuring examinee abilities through their answers to a set of test items. IRT shows the relationship between examinee ability and performance on an item. The IRT framework under "invariant item parameter" principle generate item parameter estimates (a, b, c) which are item discrimination, item difficulty level and pseudo-guessing respectively. One of the most important advantages of IRT allows comparison between examinees who answered different test items. This property, known as invariance, is obtained by introducing separate parameters for the examinee abilities and item difficulties (Pena, et.al, 2018). In IRT, ability and item parameters are both estimated based on examinees' response patterns on the test items (Adedoyin, & Mokobi, 2013).

2.3.1 *Item Response Theory*

Item Response Theory (IRT) is a measurement framework used in the design and analysis of educational and psychological assessments (i.e. achievement tests, rating scales or inventories) that measure mental traits (Ogunsakini & Shogbesani, 2018). There are three IRT parameter logistic models (PL) namely; 1PL, 2PL and 3PL.

2.3.2 *The One-Parameter Logistic (1PL) Model*

The one-parameter (1PL) model also called Rasch Model in the IRT is the simplest and most widely used from the three models. The 1PL model estimates test item difficulty parameter ***bi*** (Ogunsakini & Shogbesani, 2018). The 1PL model is computed using:

$$P_i(\theta) = \frac{1}{1 + \exp(\theta - b_i)} \quad (\text{Ogunsakini \& Shogbesani, 2018})$$

i index refers to the item

P indicates the probability

b indicates the difficulty level of the test item and

θ is for learner's ability

2.3.3 *The Two-Parameter Logistic (2PL) Model*

The two-parameter logistic model presents the item discrimination parameter and the item difficulty level varies across items (David, 2013). Here is the formula how the 2PL Model is computed:

$$P_i(\theta) = \frac{1}{1 + \exp[-Da_i(\theta - b_i)]} \quad (\text{David, 2013})$$

The index **a** indicates the discrimination of the test item depending on the learner's ability while **b** indicates the item difficulty level of the test item.

2.3.4 *The Three-Parameter Logistic (3PL) Model*

The three parameters being used for this model are **a**, **b** and **c** which are the discrimination, difficulty and guessing parameters respectively. The **c** guessing parameter is useful for multiple-choice and true-false testing (Ogunsakini & Shogbesani, 2018). In view of this, the 3PL Model was a good model for the study because it used the multiple choice test items in order to investigate the validity and reliability of modified teacher-made Biology test items. The 3PL Model is computed as:

$$P_i(\theta) = c_i + \frac{1 - c_i}{1 + \exp[-Da_i(\theta - b_i)]} \quad (\text{Ogunsakini \& Shogbesani, 2018})$$

The **a_i** discriminates power of item **i**,

b_i is the difficulty index that represents the value of the individual parameter (e.g., cognitive ability of a learner).

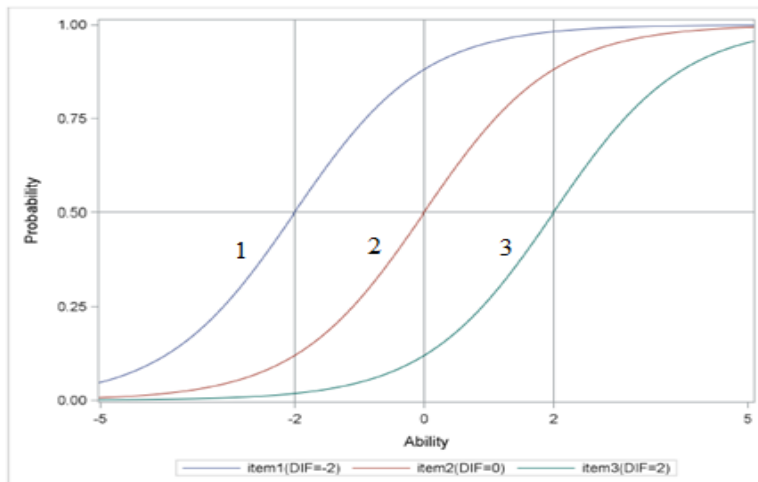
The function **1 / (1 + exp(-t))** is a logistic function, with **exp (-t)** denoting **e**, the natural exponent.

The **D** is a multiplicative constant, typically set to 1.7 or 1.702, because this value helps the 2PL model approximate the normal ogive model (Ogunsakini & Shogbesani, 2018).

Therefore, the researcher investigated the reliability of modified teacher-made Biology JCE Mock test items to determine the difficulty level of the items.

2.3.5 *Item Characteristic Curve*

Item Characteristic Curve (ICC) is used to select test items in test development procedure as it graphically presents the psychometric properties of the test items. For example, at a glance the curve gives such information as discriminating, difficulty and guessing value (Philip & Ojo, 2017).



Source: (Xinming and

Yung, 2014).

Figure 2.1 Item Characteristic Curves

Figure 2.1 shows the item characteristic curves (ICCs) for three items, with different difficulty parameters of -2 , 0 , and 2 . The test item 1 which has -2 probability of a correct response at 0.5 demonstrates that the test item was very easy that even those test item takers with low intelligence quotient (IQ) would get them correct. Test item 2 with a zero level of difficulty at 0.5 probability of a correct response demonstrates that those test item takers with an average IQ would get them correct. The test item 3 which has the difficult

level of 2 at 0.5 probability of a correct response shows that test takers only with a higher IQ may get it correct (Xinming and Yung, 2014).

Therefore, it is from this understanding that the researcher investigated the level of difficulty on the modified teacher-made test items for Biology JCE Mock for learners with visual impairment.

The ICCs for modified and original test items scores were compared in order to establish which items were more difficult or more discriminating than the other.

The IRT test characteristics curve models provide a mathematical equation for the relation of the probability of correct response to ability of the test taker. Each model employs one or more parameters whose numerical values define a particular item characteristic curve (Baker, 2001). Such mathematical models are required to develop a measurement theory that can rigorously be defined and is agreeable to further growth. In addition, these models and their parameters provide a means for communicating information about an item's technical properties (Baker, 2001). For each of the three models, the mathematical equation was used to compute the probability of correct response at several ability levels.

Therefore, the researcher investigated the modified test items that had diagrams, pictures and tables and determine content validity, difficulty level and discrimination.

Muzingo (2018) carried out a study to find out whether there was fairness in scoring of modified examinations by measuring the inter-rater agreement of scores obtained from two judges. The study findings indicated a moderate reliability index which showed a weak degree of agreement between Special Needs Specialist (SNS) rater and ordinary

rater in scoring of modified items. Therefore, the study made a recommendation to Malawi National Examinations Board (MANEB) to go for a reliable marking procedure in which the SNS raters should be considered due to their skills in Special Needs Education (SNE).

The current study was aimed to investigate the validity and reliability to determine level of difficulty and test item discrimination of modified teacher-made Biology test items with diagrams, pictures and tables which were modified into text form and then converted into Braille. Hence, the researcher investigated the validity and reliability on difficulty level and discrimination of teacher-made test items for Biology JCE Mock that had diagrams, pictures and tables that were modified into Braille.

The study used IRT for estimating the validity and reliability on item difficulty and discrimination of the modified teacher-made JCE Mock test items with DPT.

This IRT theoretical framework believes that the difficulty level, item discrimination, guessing and other item properties remain the same despite the modification of such test items with diagrams, pictures and tables (Haberman, 2016). So it helped to identify issues of test item biases for the modified teacher-made test items in Biology with DPT.

2.4 Chapter Summary

The Chapter discussed related literature on definitions of a test, test item construction, test item analysis, test item modification, validity and reliability. In addition, it discussed the item response theory, IRT logistic models which were test item difficulty level, item discrimination and pseudo-guessing. Finally, it discussed the classical test theory and the theoretical framework of the Study.

CHAPTER THREE

METHODOLOGY

3.0 Chapter Overview

The Chapter discusses the research paradigm, design of study, population and sampling procedures, sample population, sampling technique and sample size, research data management and dissemination, data generation and instrumentation; validity and reliability of the data collection instruments. It also discusses ethical considerations which helped to guide the research study.

3.1 Research Paradigm

The Study was grounded on positivism research paradigm. Paradigm is a set of beliefs or philosophical assumptions that guide researchers when conducting a study (Creswell, 2014). Morgan (2007) explains paradigm as a system of beliefs and practices that influence how researchers select both the questions they study and methods that they use to study them.

Positivism is also called a quantitative research tradition. Quantitative research is an approach for testing objective theories by examining the relationship among variables (Creswell, 2014). Quantitative approach was ideal for this study as data was collected in

form of numbers (i.e. scores). The study objectively measured and analyzed the scores using statistical procedures.

3.2 Design of the Study

The current Study applied a quantitative methodological approach. It examined the extent to which two or more variables co-vary, reflecting changes in one variable through changes in another. It is a correlational statistical test, used to establish a consistent tendency or pattern in two sets of data or variables (Creswell, 2014).

Additionally, a comprehensive overview of the variables of interest was conducted through a descriptive analysis, which played a vital role in offering a detailed understanding of the research subject. The study specifically focused on the difficulty index and discrimination of test items, modifying them from their original format of diagrams, pictures, and tables into textural statements. This descriptive analysis facilitated a thorough presentation of the test items, enabling effectively assess their difficulty levels.

On the same the researcher analysed the content validity through data that was collected from the Subject Matter Experts (SMEs) and Specialist Teachers (STs) for learners with visual impairment using the Likert Scale tool. The SMEs and STs made analysis on whether the MTMBTIs and the original items that had DPTs were from the same topic and objective or not. Content validity assesses whether a test is a representative of all aspects of the construct (Middleton, 2022). In this case, the researcher wanted to find out if the modified test items for Biology JCE Mock paper really were the true reflection of the content that the participants were supposed to be measured (Middleton, 2022). In

other words, the researcher examined if the modified test items with diagrams, pictures and tables were developed from the same topic and objective that had the original items. The researcher involved the Subject Matter Experts (SMEs) and the Specialist Teachers for learners with visual impairment to critically analyze the test items in the modified Biology JCE Mock paper if they were the true reflection of the topics and objectives as they appeared in the original print copy.

3.3 Study population

The target population of the study were learners in secondary schools in South West Education Division (SWED) which according to the preliminary findings made in 2023 from the South West Education Division Office showed there were a total of 38,483 learners where 20,380 were males and 18,103 were females. The conclusive findings remained constant at a total of 38,483 learners where 20,380 were males and 18,103 were females. The findings also showed that learners with visual impairment were very few in numbers in secondary schools that were sampled. There were only six Form Two visually impaired learners both at Secondary School A and Secondary School B as indicated in the table below:

Table 3.1 Learners with visual impairment

School	Male	Female	Total
Secondary School A	4	1	5
Secondary School B	1	0	1
TOTAL	5	1	6

In research design, inclusion and exclusion criteria are essential components which are used to define the characteristics of the study population and the conditions under which data would be collected. Inclusion criteria are attributes of subjects that are essential for their selection to participate in the study while exclusion criteria are responses of subjects that require their removal as subjects which do not meet the required criteria for the study population (Creswell & Creswell, 2017).

Therefore, the inclusion criterion was used to ensure that learners with visual impairment were included in the study while exclusion criterion was used to make sure that the form one, three and four learners were not included in the study because its main focus was on form two learners as these write the JCE national examinations. Both, inclusion and exclusion criteria ensured that the study results were valid, applicable, and obtained in an ethical manner.

These standards made it possible to verify that the study's findings were reliable, pertinent, and applicable to the intended audience. Since the study was carried out in two districts with a 1000 sample size then it was basically made possible for generalisation to a larger population. Consequently, the Form Two was the demographic feature of the study's target group. Out of the total 10,633 learners, 5,433 were boys and 5,190 were girls. Among them were six visually impaired learners, five of them were male and one female.

3.4 Sampling technique and sample size

The researcher used the probability sampling. It is one of the sampling techniques in which each sample has the same probability of being chosen (Curtin et al., 2005; Fowler,

2009). There are different methods in probability sampling and one of them that the researcher used was cluster sampling method. Cluster sampling is a sampling method where the entire population is divided into groups, or clusters, and a random sample of these clusters are selected (Singh & Masuku, 2014). It is a sampling technique used when “natural” but relatively homogeneous groupings are evident in a statistical population.

So the cluster sampling method was conducted in two stages. In the first stage the researcher grouped or clustered the secondary schools in the SWED into districts which were Blantyre, Mwanza, Neno, Chikwawa and Nsanje.

To choose the sample for the study, the researcher used the Cluster Sampling approach with great care in a two-step process. The secondary schools in the SWED were first divided into districts, of Blantyre, Mwanza, Neno, Chikwawa, and Nsanje constituting the clusters. In the second step, Blantyre and Chikwawa Clusters were randomly sampled from the five clusters. In order to have a total sample of nine secondary schools, the researcher randomly sampled five schools from Blantyre and four from Chikwawa Cluster. The Form Two class from the chosen schools served as the sample frame, which the researcher used to identify the study participants. Consequently, all Form Two learners with and without visual impairments from the selected schools were included in the study.

The following formula was applied in Cluster Sampling:

K = represents the total number of clusters in the population.

M_i = represents the number of units in cluster i .

N = represents the number of clusters selected in a simple random sample.

The overall population size is: $N = \sum_{i=1}^K M_i$. (Wu, & Thompson, 2020).

The researcher determined a sample size of 1000 participants using IRT theoretical framework which recommends the minimum sample of 500 samples if to generate accurate, valid, reliable and invariant parameters during data analysis (Lord, 1980).

Therefore, the researcher applied the strategy “imitating sample size of similar studies” as it was one of the strategies for determining sample size (Singh & Masuku, 2014). This approach is particularly useful when there is no access to extensive resources for conducting pilot studies or when working on a related research question. For example, Konala (2018) used 1003 population sample size in his study titled: “Examining the Quality of Form One Selection Test for Faith Mission Secondary Schools”. In addition, De Ayala, (2013) recommends 1000 population sample size if the 3PL was applied in order to generate accurate, valid, reliable and invariant parameters during data analysis. Therefore, the researcher adopted the 1000 sample size.

3.5 Research data management and dissemination

Research data management (RDM) is a term that comprises activities associated with the storage, organization, documentation, and dissemination of data. It is essential to efforts aimed at exploiting the value of scientific investment and addressing concerns related to the reliability of the research practice (Borghi J, et.al. 2018)).

On the other hand, research dissemination refers to a planned process that involves consideration of target audiences and the settings in which research findings are to be received. In addition, an appropriate communication and interaction with wider policy

and education service audiences is through ways that would facilitate research approval in decision-making processes and practice (Wilson, P.M., et al. 2010).

Therefore, the researcher followed stringent protocols to securely organize and store all data collected, accurately safeguarding against any unauthorized disclosure. On the same, the dissemination of the data would be restricted solely to the relevant groups.

Subsequently, the researcher would disseminate the research outcomes pertaining to the validity and reliability of the Modified Teacher-Made Biology Test Items. This would be achieved through the stakeholder meetings and participation in academic conferences. The aim is to ensure that the modified items with diagrams, pictures, and tables, exhibit equivalent levels of difficulty and discriminative capacity as the original items.

3.6 Validity and Reliability of data collection instruments

The researcher chose the Blantyre Secondary School 2020 modified teacher-made Biology JCE Mock test items because they were developed and moderated by Biology subject teachers from eleven different secondary schools in a collaborative effort at the cluster level. The collaborative approach ensured that the test items were clear and easily understood by learners, and any items that were not deemed comprehensible were restructured or removed altogether.

In the study, the researcher employed a Likert Scale instrument to collect data from subject matter experts (SMEs) and specialist teachers for learners with visual impairment (STLWVI). The instrument had two distinct sections, namely Section A and Section B. Section A aimed to evaluate the participants' knowledge and experience in test item development and modification (TDM). Section B was designed to conduct a comparative

analysis (CA) of modified and original test items that contained diagrams, pictures, and tables.

The main objective of Section B was to guarantee the content validity of the modified test items in relation to the original ones, particularly in terms of their alignment with the same topic and objectives.

3.7 Data Generation and Instrumentation

Data was collected from a total of nine sampled secondary schools, including National Secondary School (A), District Boarding Secondary School (B), Day Secondary Schools (C, D and E), as well as Community Secondary Schools (F, G, H and I). Modified Teacher-made Biology Test Items data collection tool was administered to 1000 Form Two participants, including those with and without visual impairments.

The Modified Teacher-Made Multiple-Choice Test Items for the Biology JCE Mock test tool (See appendix 1) had thirty multiple-choice items, both modified and unmodified featuring diagrams, pictures and tables. It was administered to the Form two learners with and without visual impairments in the selected secondary schools in SWED. It was administered with matched examination conditions as starting time, duration and classroom environment such as seating arrangement where desks were arranged in rows and columns with at least one metre apart. In addition, the examination was invigilated.

In addition, three subject matter experts were hired and gathered at the school A to score the multiple choice answered scripts. The researcher worked together with them by supervising the accuracy of scoring the answered scripts using the marking key provided. The researcher had to check the marked scripts to ensure the test items were correctly

scored. Overall, this attention to detail was critical in ensuring that the results reflected the true performance of the participants' abilities thereby to safeguard the validity and reliability of the whole process.

Then the scores were entered in a computer excel sheet and exported to a statistical software package (STATA). So, the analysis of validity and reliability was conducted using IRT parametric logistic models through test item analysis to generate item characteristic curves (ICCs) for item parameter estimates: item discrimination (a-value), item difficulty (b-value), item pseudo-guessing (c-value), item graphics, and test reliability (r) from scores.

In addition, the data from the Likert Scale instrument which was administered to Subject Matter Experts for Biology and Specialist teachers for learners with visual impairments was analyzed using the SPSS software to assess the reliability and validity of the Modified Teacher-made Biology Test Items respectively.

Therefore, the study evaluated the validity and reliability of the Modified Teacher-Made Biology Test Items for learners with visual impairment which was ascertained through item difficulty index and discrimination. It determined whether the modified test items were of the same difficulty level and discrimination to the original items that had diagrams, picture, and tables.

The Biology JCE Mock Paper data collection instrument had 20 multiple choice test items and 10 of them were modified which were referred to as Modified Questions (QM) and the other 10 items were the original items that had diagrams, pictures and tables which were not modified and were referred to as Questions with Diagrams (QD).

Therefore, QM items were QM5, QM6, QM8, QM9, QM10, QM13, QM15, QM18 QM19 and QM20. In addition, 10 original test items with diagrams, pictures, and tables (QD) were also administered, making a total of 30 multiple-choice test items. These were QD5, QD6, QD8, QD9, QD10, QD13, QD14, QD18, QD19 and QD20. Both items were used to see if they were assessing the same topic and objectives.

So, the reliability assessment focused on analyzing the difficulty level and discrimination of the modified Biology JCE Mock test items using the IRT 3PL Model item characteristics curves.

3.9 Ethical Considerations

The researcher put into account all ethical considerations as it is required in any research activity when it is being carried out. Research involves the process of collecting data from people and about people (Punch, 2005). As one way of observing ethical considerations the researcher sought consent from the South West Education (SWED) and the head-teachers of the schools. In addition, the researcher explained to the learners that their participation was voluntary and they could choose to withdraw even after they had already started taking part. The researcher also made sure that the participants were protected by seeing to it that the environment was free from any danger. Researchers need to protect their research participants; develop a trust with them; promote the integrity of research; guard against misconduct and impropriety that might reflect on their institutions; and cope with new, challenging problems (Israel & Hay, 2006).

On the other hand, the researcher respected the privacy and anonymity of participants for example, by not sharing their information with any other third party. Issues of plagiarism

were avoided where no information from other sources was used without permission. The participants were not forced in signing consent forms; and the researcher was sensitive to the needs of vulnerable groups such as children with visual impairment were provided the test items in Braille format.

Therefore, the researcher would not disclose any information that was collected from the schools such as names of the participants and their scores. Participants were kept anonymous throughout the research instead “pseudo-names” were used where the need arised. Participants had an opportunity to withdraw or not to take part in the research if they wished to do so during the period of data collection or thereafter.

In addition, for some participants in the study were young children, the researcher sought consent from the headteachers of the schools under study. The researcher got permission from the Education Division Managers (EDMs) for South West Education Division (SWED) in which Blantyre and Chikwawa Districts are found. In addition, the researcher also sought permission from the head-teachers of the selected secondary schools to conduct the study.

In the study, the core task of the University of Malawi Research Ethics Committee (UNIMAREC) was reviewing and approving research protocols for both ethical and scientific merit. Additionally, the committee diligently conducted on-site inspections at the sampled secondary schools involved in the study to ensure that the approved research protocols were being implemented in strict accordance with ethical standards guidelines. On the other hand, the Department of Educational Foundations provided a Letter of Introduction to serve as a confirmation that the researcher was a registered postgraduate

student. The letter served to request the concerned institutions and authorities to assist the researcher to collect the required data.

3.10 Chapter Summary

The Chapter has presented a research paradigm, design of study, the study population, sampling technique and sample size. The Chapter has also discussed research data management and dissemination, data generation and instrumentation; data analysis, validity and reliability of the research study and ethical considerations.

CHAPTER FOUR

RESULTS AND DISCUSSIONS OF THE FINDINGS

4.0 Chapter Overview

This Chapter presents the analysis of the results and discusses the research study findings based on the study topic. It discusses the reliability analysis on difficulty index for modified teacher-made Biology JCE Mock test items for learners with visual impairment. In addition, it explores the effectiveness of modified Biology JCE Mock test items in assessing learners with and without visual impairments using item characteristic curves; reliability of modified Biology JCE Mock test items for learners with Visual Impairment in relation to item difficult index; analysis of test item discrimination for modified Biology JCE Mock test items; and analyses the guessing level of the modified test items.

It further investigated the content validity of modified Biology JCE Mock test items for learners with Visual Impairment. Lastly, it provides the summary of the findings and discussions of the study.

4.1 Reliability Analysis of Difficulty Indices for Modified Biology JCE Mock Test Items for Learners with Visual Impairment

The comparison of modified questions (QM) and their original questions with diagrams (QD) resulted in the identification of distinct latent traits. Latent trait refers to an

unobservable characteristic of interest or ability and it is conventionally denoted by θ (StataCorp, 2023). The value of θ for a particular individual is commonly known as the person's location. The item properties of the latent trait are parameters, including difficulty and discrimination that are estimated in the IRT model (StataCorp, 2023). The analysis of latent traits provides an indication of the reliability of test items when they are not excessively difficult or too easy. Therefore, based on a *p-value* of 0.5, the latent traits of test items QM8, QM9, QM14, QD8, QD9, and QD14 are as follows in the table below:

Table 4.1 Reliability of Difficulty and Discrimination on Latent Trait for Items QM8, QM9, QM14, QD8, QD9 and QD14 at a 0.5 p-value

ITEM	LATENT TRAIT (θ)	ITEM	LATENT TRAIT (θ)
QM8	-3.492924	QD8	-.5323496
QM9	-.5323496	QD9	-.8817724
QM14	-1.95744	QD14	-2.174902

All the three modified items were answered correctly by learners with the latent trait below mean (0), meaning they were very easy items.

4.2 Exploring the effectiveness of modified Biology JCE Mock test items in assessing learners with and without visual impairments using item characteristic curves.

The analysis of the ICCs for test items QM8, QM9, QM14, QD8, QD9 and QD14 was as below:

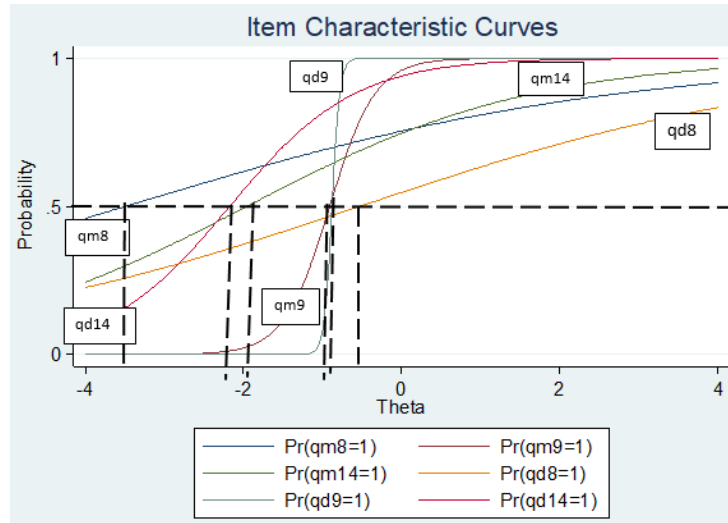


Figure 4.1 Reliability on difficulty and discrimination on ICCs for test items QM8, QM9, QM14, QD8, QD9 and QD14.

The Pr. in the item characteristics curves of item response theory (IRT) shows the probability that a person with the given level of the latent trait denoted by θ will correctly answer an item. One of its implications is that it provides test fairness by analyzing the discrimination parameter a . Psychometricians can ensure that items discriminate fairly across different levels of ability. In addition, they provide the item difficulty parameter b . (Embretson & Reise, 2013). For example, an item with a high b value is more suitable for assessing individuals with high ability levels while the low b value targets for those with low abilities.

The latent trait of the modified test item QM8 = -3.492924 demanded less learner ability than the original item QD8 = -.5323496. Item QM9 = -.5323496 needed a higher ability trait than the QD9 = -.8817724 while the QM14 = -1.95744 was also answered correctly by learners with the higher latent trait than the QD14 = -2.174902. Although, the three items were determined to be easy, the modified versions were not answered correctly by

learners with the same ability level as the original items, which had diagrams and pictures. On the other hand, more learners correctly answered the modified test item (QM8) than the original item (QD8) with a diagram. As a result, it was evident that the modified items did not accurately represent the ability index as compared to the original items thereby, rendering the modification to be invalid.

Table 4.2 Reliability of Difficulty and Discrimination on Latent Trait for Items QM6, QM13, QM18 and QM20 at a 0.5 p-value

ITEM	LATENT TRAIT (θ)	ITEM	LATENT TRAIT (θ)
QM6	-.3665743	QD6	1.472202
QM13	-.3523166	QD13	-.8199601
QM18	-.0729693	QD18	4.252257
QM20	.2181239	QD20	-.2859534

Test items QM6 = -.3665743, QM13 = -.3523166 and QM18 = -.0729693 were well responded by learners with the latent trait below an average theta (θ) while QM20 = .2181239 was correctly answered by learners with the ability level above theta (θ). Nevertheless, the latent trait for items QM6 and QM18 required a less theta to be answered correctly than the QD6 = 1.472202 and the QD18 = 4.252257 respectively. On the other hand, QM13 and QM20 required a higher ability index than the QD13 = -.8199601 and QD20 = -.2859534 respectively.

Items with a negative latent trait means were easy and were correctly answered by learners with low latent trait.

Test item QM20 showed it was well answered by learners with the latent trait nearly above the average theta (0).

Upon analysis, it became evident that the modified test items QM6 and QM18 were notably less challenging than the original items QD6 and QD18, which had a diagram. As a result, a substantial number of learners demonstrated improved accuracy in answering the modified items, indicating an apparent difference in difficulty levels between the modified and original items.

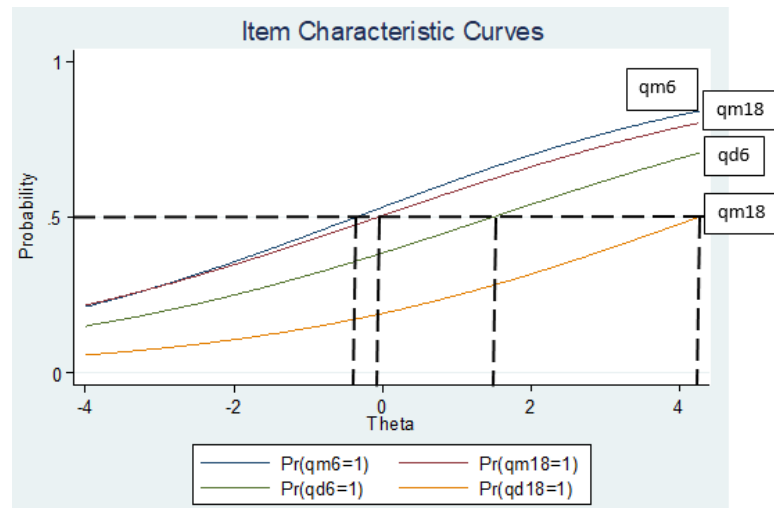


Figure 4.2 Reliability on difficulty and discrimination of ICCs for test items QM6, QM18, QD6, and QD18 at 0.5 p-value.

This suggests that the modifications to QM6 and QM18 had a notable impact on the learners' overall performance.

Test items QM13 and QM20 were found to be more challenging than the original test items with diagrams, QD13 and QD20, respectively according to the ICCs below.

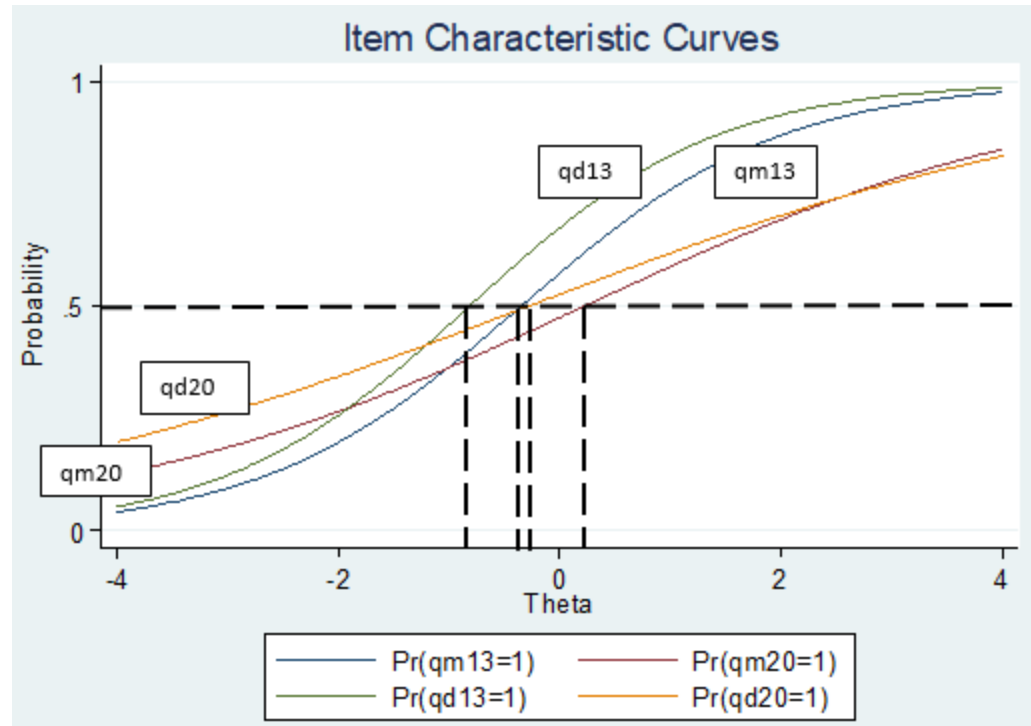


Figure 4.3 Reliability on difficulty and discrimination on ICCs for test items QM13, QM20, QD13, and QD20 at 0.5 p-value.

This indicated that learners with visual impairments might have created more complex items than their sighted peers, raising concerns about the reliability of these items.

Table 4.3 Reliability of Difficulty and Discrimination on Latent Trait for Items QM5, QM10, and QM19 at a 0.5 p-value

ITEM	LATENT TRAIT (θ)	ITEM	LATENT TRAIT (θ)
QM5	1.836998	QD5	2.265567
QM10	8.033605	QD10	-47.94331
QM19	80.70332	QD19	1.634362

The latent trait of the modified test items $QM10 = 8.033605$ and $QM19 = 80.70332$ showed they were more difficult than original items $QD10 = -47.94331$ and $QD19 = 1.634362$ which had diagrams, pictures and tables.

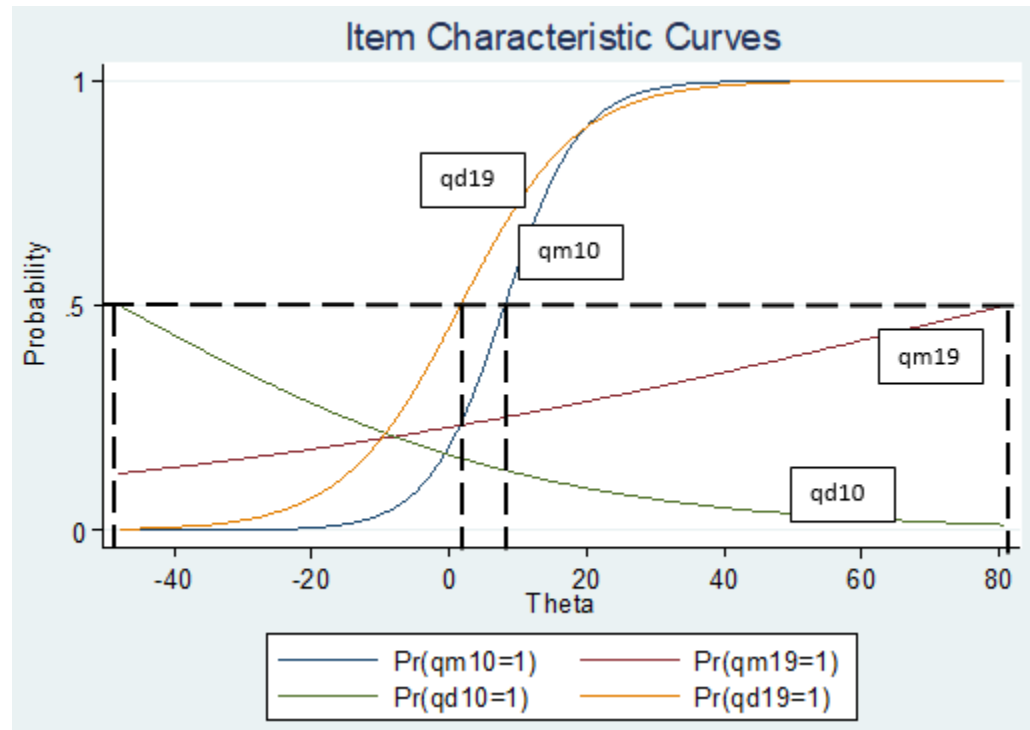


Figure 4.4 Reliability on difficulty and discrimination on ICCs for test items QM10, QM19, QD10, and QD19 at 0.5 p-value.

The latent traits of the modified test items QM10 and QM19 indicate that learners with a theta above the average (0) provided correct answers. However, item QM19 required a theta score of 80.70332, making it excessively challenging. It had a disproportionately negative impact on both visually impaired and non-impaired learners due to the low rate of correct responses. Consequently, item QM19 should have been entirely removed from the test, as it did not meet the criteria of a valid modified item match.

On the other hand, $QM5 = 1.836998$ was answered correctly by learners with a less latent characteristic than the $QD5 = 2.265567$ as it is shown in figure below.

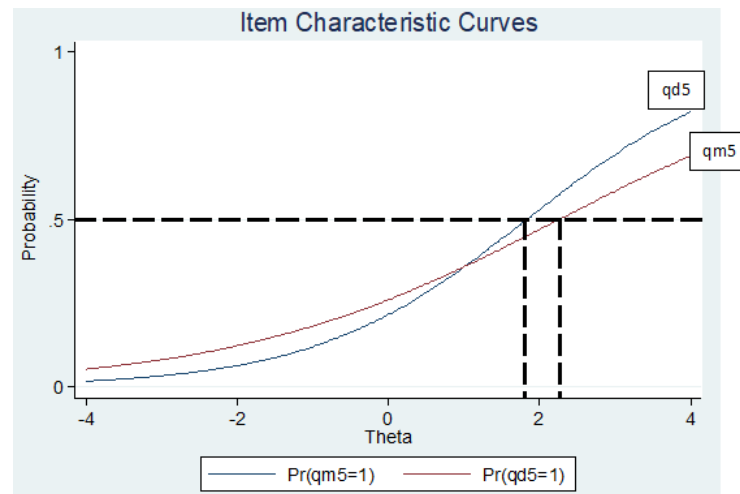


Figure 4.5 Reliability on difficulty and discrimination on ICCs for test items QM5 and QD5 at 0.5 p-value.

Modified test item QM5, although it was easier than the original item QD5, but it still demanded a learner with latent trait above average (0) therefore, it was a difficult test item.

In view of this, the modified item (QM5) did not have equal difficulty level to the original item with diagram (QD5).

4.3 Assessing the Reliability of Modified Biology JCE Mock Test Items for Learners with Visual Impairment: Analyzing Item Difficulty Index

The table below provides a comparison analysis of p-values for item difficulty between modified and original items with diagrams, pictures and tables. Below is the p-value analysis on the item difficulty for the modified items.

When modifying test items, the reliability of the items could be impacted by various test parameters and one of them is Item Difficulty (*P-value*) (Crocker & Algina, 2008). The item difficulty index, or *P-value*, which shows the proportion of examinees who answered an item correctly. Crocker & Algina (2008) further explain, if modified test items shift the distribution of difficulty too far in one direction i.e. too difficult or too easy, it can reduce test score variance, thereby lowering reliability.

The *p-value* difficulty index between 0.20 and 0.90 are considered as good and acceptable. The *P-values* above 0.90 indicate very easy items that they must not be taken up into the test because the item addresses a concept probably not worth testing or it may require to be restructured (Zahran & Mustafa, 2023).

P-values below 0.20 indicate very difficult items. This confirms that most learners responded incorrectly, either an item was flawed or learners did not understand the concept and as such the item can be revised or be removed from the test (Zahran & Mustafa, 2023). Therefore, the lower the *p-value*, the more difficult the particular test taker will get it correct. The higher the *p-value*, the more likely that the test taker would get it correct.

Below is the *p-value* analysis on the item difficulty for the modified items.

Table 4.4 A comparison analysis of p-values for Item Difficulty between Modified and Original Items

Modified item	Original item	Difficult level
QM5 = 0.24	QD5 = 0.27	Same
QM6 = 0.53	QD6 = 0.39	Less

QM8 = 0.75	QD8 = 0.55	Less
QM9 = 0.79	QD9 = 0.82	More
QM10 = 0.19	QD10 = 0.17	Same
QM13 = 0.56	QD13 = 0.65	More
QM14 = 0.73	QD14 = 0.90	More
QM18 = 0.51	QD18 = 0.20	Less
QM19 = 0.23	QD19 = 0.45	More
QM20 = 0.48	QD20 = 0.53	Same

Modified test items QM5, QM10 and QM20 had the same difficulty level to the original items QD5, QD10 and QD20 while modified items QM9, QM13, QM14 and QM19 were more difficult than the original items QD9, QD13, QD14 and QD19. On the other hand, modified items QM5, QM10 and QM20 maintained the difficulty level to the original items QD5, QD10 and QD20.

4.3.1 Analysing Less Difficult Modified Test Items with Diagrams, Pictures, and Tables

The modified test items QM6 = 0.53, QM8 = 0.75 and QM18 = 0.51 were less difficult than items with diagrams, pictures and tables QD6 = 0.39, QD8 = 0.55 and QD18 = 0.20 respectively. The modified item QM6 was in the moderate difficult range of 0.40 - 0.59 while QD6 was in the difficult range of 0.20 - 0.39. QM8 and QD8 were also in different difficult ranges 0.40 - 0.59 moderately difficult and 0.60 - 0.79 moderately easy

respectively. While QM18 and QD18 were both in the same difficulty range of 0.40 - 0.59.

Therefore, modification of items QM6 and QM8 was not the same because they were not in the same difficulty index so, they were not reliable while modified item QM18 was correctly modified and was reliable.

4.3.2 Analysing More Difficult Modified Test Items with Diagrams, Pictures, and Tables

Modified items QM9 = 0.79, QM13 = 0.56, QM14 = 0.73 and QM19 = 0.23 were more difficult than items QD9 = 0.82, QD13 = 0.65, QD14 = 0.90 and QD19 = 0.45 respectively. Item QM9 and QD9 were in different range 0.60 - 0.79 moderately easy and 0.80 - 0.89 easy respectively. Therefore, the modification was not reliable. Item QM13 and QD13 their difficult ranges were different 0.40 - 0.59 moderately difficult and 0.60 - 0.79 moderately easy respectively. So, the modification was not good. Item QM14 and QD14 had different difficult range 0.60 - 0.79 moderately easy and 0.90 the easiest. This meant the modification was unreliable. On the same items QM19 and QD19 were in different difficult ranges 0.20 - 0.39 meaning it was difficult and 0.40 - 0.59 moderately difficult. Therefore, the modification was not reliable.

The results from the test items reveal notable differences in participant performance. For the modified QM19 test item, only 230 out of 1,000 participants, approximately 22.8%, answered correctly. In contrast, the same QD19 test item with a diagram saw a significantly higher success rate, with 447 participants answering correctly, representing

44.7% of the total. This disparity is clearly illustrated in the accompanying frequency tables.

Tables 4.5 Frequency of QM19 and QD19 Scores

		QM19			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	770	76.5	77.0	77.0
	1	230	22.8	23.0	100.0
	Total	1000	99.3	100.0	

		QD19			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	553	54.9	55.3	55.3
	1	447	44.4	44.7	100.0
	Total	1000	99.3	100.0	

A particularly important observation emerges from the QM19 test item, where only one participant with visual impairment, out of a group of six VI learners, was able to answer correctly. This finding raises concerns about accessibility and comprehension for those with and without visually impaired learners.

The data strongly indicate that the modified QM19 test item presents greater challenges than the original QD19. Such a conclusion suggests potential flaws in the design of the modified item, which raises questions about its effectiveness. Ultimately, these findings demonstrate that QM19 lacks both validity and reliability, making it a less test item compared to its original item.

4.3.3 *Analysing Same Difficult Index Modified Test Items with Diagrams, Pictures, and Tables*

Modified items QM5 = 0.24, QM10 = 0.19 and QM20 = 0.48 were in the same difficult index to QD5 = 0.27, QD10 = 0.17 and QD20 = 0.53 respectively. QM5 and QD5 were in the range 0.20 - 0.39 and were difficult. While items QM10 and QD10 were in the difficult range less than 0.20 and were the most difficult. Items QM20 and QD20 were also in same difficult range 0.40 - 0.59. Therefore, these three items were correctly modified and were valid.

From the Table 4.5 above, three of the ten modified test items had the same difficult range to the original items. In addition, there were four modified items that were more difficult against the original items with diagrams, pictures and tables while the other three items were less difficult as compared to the original items and this summary is shown in the table below:

Table 4.6 Percentage Distribution of Items Across Same, More, and Less Difficult Ranges

Items	Total items	Percent (%)
Same difficult	3	30%
More difficult	4	40%
Less difficult	3	30%
TOATL	10	100%

Upon evaluating the reliability of the modified teacher-made Biology test items in comparison to the original items, the researcher noted that only 30% of them were of the

same difficulty level to original items with diagrams, pictures, and tables. The remaining 70% of the modified items, however, were found to be either more difficult or less difficult, indicating that the modification did not produce items of the same difficulty level as the original ones. The 30% of the modified items were less difficult while the 40% was more difficult. This raised concerns about the reliability of the modified test items in accurately measuring the ability of learners with visual impairment in Biology knowledge and skills in equal terms to their friends without visual impairment. The results were graphically presented as:

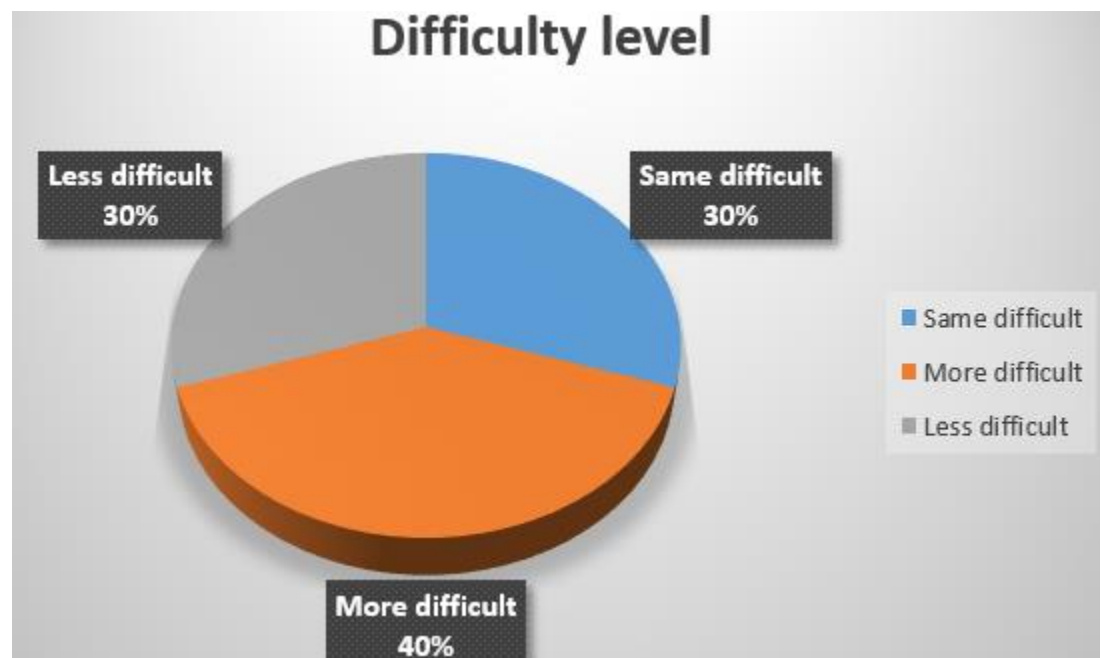


Figure 4.6 Test item difficult level

The modification of the test items did not maintain the difficult index as it was in the original test items with diagrams, pictures and tables. Therefore, the modified test items were not reliable.

4.4 Analysis of Item Discrimination in Modified Biology JCE Mock Test Questions

Test item discrimination is a technique to gauge the variation in item difficulty between groups of learners with high and low marks. It shows how much an item can differentiate learners with varying skill levels. Teachers can identify and fix problematic test items with the help of discrimination indices for each test item, which offer information about what learners have learnt (Azzopardi & Azzopardi, 2019). It offers a useful instrument for creating the exam.

The table of estimated IRT parameters is replayed and the reporting format is adjusted using the *estat* report. The *sort(a)* option is carried out to show items in ascending order of discrimination, and the by-parameter (*byparm*) option that arranges the output by parameter instead by item (Balov, 2016). This facilitated the discrimination observation of the modified test items.

Therefore, the discrimination of test items after item analysis were observed as in the table below:

Table 4.7 Item Discrimination Table Sorted in Ascending Order

MODIFIED TEST ITEMS (QM)		ORIGINAL TEST ITEMS WITH DIAGRAMS, PICTURES AND TABLES (QD)	
Test item	Discrimination index	Test item	Discrimination index
QM19	.0757455	QD10	-.1348761

QM10	.1940061	QD19	.1307631
QM8	.3037523	QD20	.3032057
QM18	.3737623	QD18	.4712237
QM9	.5418865	QD8	.5430468
QM6	.6717737	QD5	.588563
QM20	.8054764	QD6	.6495398
QM14	.865815	QD9	.8325726
QM13	.8812656	QD13	1.078227
QM5	1.108983	QD14	1.713783

According to the Table 4.7 above, the modified item that had the most discriminating index was item QM5 (Discrim = 1.11) whereas item QM19 (Discrim = 0.08) had the least discriminating index.

Therefore, items QM10 and QM19 were poorly discriminating and were supposed to be rejected.

Items that have a discrimination index between 0.3 – 0.39 have a good discrimination and are reasonably good but subject to improvement. Therefore, such items in the modified test items were QM8 = .30 and QM18 = 0.37 which needed to be improved.

Test items QM9 = 0.54, QM6 = 0.67, QM20 = 0.21, QM14 = 0.87, QM13 = 0.88 and QM5 = 1.11 had $D > 0.4$ therefore, they had an excellent discrimination representing

60% well. The other four items QM8, QM10, QM18 and QM19 did not discriminate well because they were less than 0.4 and represented 40% of the ten modified items.

Original item QD10 had a negative discrimination index of $-.1348761$ and was deemed to be very poor so it was supposed to be removed from the test. This is according to Mitra, (2009) states item with a negative discrimination index (D) is considered to be very poor and should be removed while an item with a D of $0.0 - 0.19$ is considered poor and should be revised. Discrimination index of $0.2 - 0.29$ is acceptable, $0.3 - 0.39$ is good, and >0.4 is excellent.

However, there was no modified item (QM) with a negative discrimination index but items QM10 = 0.08 and QM19 = 0.19 had a range of discrimination index D: $0.00 - 0.19$ so the items were poor and not reliable and they were supposed to be revised. Items QM8 and QM18 had the D: $0.3 - 0.39$ and were good while items QM5, QM6, QM9, QM13, QM14 and QM20 were in the discrimination index greater than 0.4 and were said to be excellent and were reliable.

4.5 Assessing the Guessing Level in Modified Test Items

The analysis found 0.0825487 as the guessing parameter with 0.0472796 Coefficient (Coef.) as it is in the table below:

Table 4.8 Guessing Parameter of Logistics Model

Three parameter logistics model		Number of obs		=	
1000					
Log likelihood = -5904.0541	Coef.	Std. Err	z	P> z	[95% Conf.
Interval]					

Guess 0.0825487	0.0472796	1.75	0.081	-0.0101175
0.175215				

The guessing parameter is below 0.35 therefore, according to Baker, (2001), explains a guessing parameter below 0.35 is considered acceptable. This affirms that the modified test items had minimal guessing probability. Therefore, the learners made very few guesses, which had no significant impact on the reliability of the scores.

4.6 Evaluating the Content Validity of Modified Biology JCE Mock Test Items for Learners with Visual Impairments

A further analysis was conducted on content validity of the modified test items and the original items that had diagrams, pictures and tables. The data was gathered from the Biology Subject Matter Experts (SMEs) and the Specialist Teachers for learners with visual impairment (STLWVI) who responded to the Likert Scale instrument. The understanding was to find out if the modified items were developed from the same topics and objectives to the original items that had diagrams, pictures, and tables in order to establish their validity.

Content validity based on comparative analysis (CA) between modified test items and the original test items with diagrams, pictures and tables QM5, QM6, QM8, QM9, QM10, QM13, QM14, QM18, QM19, and QM20 were compared to original items QD5, QD6, QD8, QD9, QD10, QD13, QD14, QD18, QD19, and QD20 to determine the content validity.

Additionally, the content validity was examined using the STLVI and SMEs' test item development and modification (TDM) experiences. It concentrated on their understanding of test development, test item modification if there was need for in-service training.

4.6.1 Content Validity Based on Comparative Analysis (CA) of Modified Test

Items and Original Test Items: An Examination with Diagrams, Pictures, and Tables

Data collected from Subject Matter Experts (SMEs) using Likert Scale was explored to find if it was normally distributed in order to use parametric method or if not normally distributed then non-parametric method could be used. It was then found that the data was not normally distributed (Ghasemi & Zahediasl, 2012).

Therefore, non-parametric analysis method of Ordinal Regression was used. So, data test of normality was tabulated as in the table below:

Table 4.9 Data Test of Normality

	Tests of Normality					
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
MTD	.188	23	.034	.872	23	.007
M						
MCA	.247	23	.001	.861	23	.004
Q						

The above table presents the results from two well-known tests of normality, namely the Kolmogorov-Smirnov Test and the Shapiro-Wilk Test. The Shapiro-Wilk Test is more appropriate for small sample sizes < 50 samples, but can also handle sample sizes as large

as 2000 (Ghasemi & Zahediasl, 2012). For this reason, the researcher used the Shapiro-Wilk test as the numerical means of assessing the data collected from SMEs and STLVI which was collected using the Likert Scale.

Shapiro Wilks W Test

$$W = \frac{\left(\sum_{i=1}^n a_i x_{(i)}\right)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

- W is the test statistic
- W is insignificant if the variable's distribution is not different from normal
- $W \approx$ the correlation between given data and ideal normal scores
- $W = 1$ when your sample-variable data are perfectly normal (perfect H_0)
- When W is significantly smaller than 1 = non-normal (H_a is accepted) (Ghasemi & Zahediasl, 2012).

Therefore, the Test for Normality on statements of experience of SMEs and STLVI on test item development and modification (TDM) were analyzed. It was observed that the *Sig. value* of the Shapiro-Wilk Test was less than 0.05 meaning the data significantly deviated from a normal distribution. This predetermined that the data was not normally distributed. If data was to be greater than 0.05, then it would be significant to a normal distribution.

Test of Goodness-of-Fit was conducted in order to find out if the model of Normality for the Shapiro-Wilk was fit to analyze the data and the results were as in table below:

Table 4.10 Goodness-of-Fit Analysis

Goodness-of-Fit

	Chi-Square	Df	Sig.
Pearson	125.396	109	.135
Deviance	69.492	109	.999

To ascertain whether a given sample originates from a population with a certain theoretical distribution, one can employ the goodness-of-fit tests. Goodness of Fit (GOF) is a statistical model that describes how well it conforms to a set of data (Maydeu-Olivares & Forero, 2010). The difference between the values observed and the values predicted by a statistical model is summarized by GOF indices. The Goodness-of-Fit determines whether the model fits the data.

Pearson and Deviance Chi-Square Tests are statistical techniques utilized to determine if a model is a suitable fit for the data. In this case, the significance of Pearson and Deviance was found to be 0.135 and 0.999, respectively, both exceeding the accepted value of 0.05. These outcomes suggested that the data sets were not statistically significant. Nevertheless, the test for normality demonstrated that the model was a good-fit for the data sets.

As a result, the researcher utilized a non-parametric analysis technique called Ordinal Regression to investigate the validity of the modified test items. This method is used when the assumptions of standard parametric tests are not met, or when the data is not normally distributed (Maydeu-Olivares & Forero, 2010). By using Ordinal Regression,

the researcher evaluated the connection between the modified test items and the SMEs' responses in a more precise and dependable manner.

Then parameter estimates were analyzed and results were as below:

Table 4.11 Parameter Estimates for the Statistical Model

Parameter Estimates								
		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[TDM = 2.00]	-6.364	2.682	5.629	1	.018	-11.621	-1.107
	[TDM = 2.38]	-5.153	2.544	4.103	1	.043	-10.140	-.167
	[TDM = 2.50]	-3.884	2.469	2.474	1	.116	-8.723	.956
	[TDM = 2.63]	-3.506	2.449	2.050	1	.152	-8.306	1.293
	[TDM = 2.75]	-2.945	2.418	1.483	1	.223	-7.685	1.795
	[TDM = 2.88]	-2.744	2.408	1.298	1	.255	-7.463	1.976
	[TDM = 3.00]	-2.529	2.397	1.113	1	.291	-7.227	2.170
	[TDM = 3.13]	-2.065	2.378	.754	1	.385	-6.725	2.596
	[TDM = 3.25]	-1.506	2.366	.405	1	.524	-6.144	3.131
	[TDM = 3.63]	-1.153	2.369	.237	1	.626	-5.797	3.490
Location	[TDM = 4.13]	-.701	2.389	.086	1	.769	-5.384	3.981
	CAQ	-.959	.717	1.788	1	.181	-2.364	.446

In the parameter estimate, for every one unit decrease on the independent variable, there is a predicted increase of a certain amount in the *log odds* of being in a higher level on the dependent variable (Agresti, 2010)). This simply means that as the values of independent variable TDM decreases there is also a decreased probability of rising at on the dependent variable CAQ.

Therefore, the little experience of Biology teachers in the test item development and test item modification was a negative significant predictor of dependent variable of

comparative analysis (CAQ) between modified test items and the original test items with diagrams, pictures and tables based on content validity.

The negative value of comparative analysis (CAQ) of test item -0.959 in the table showed that for every one unit increase of the little experience in test item development and item modification there was a predicted decrease in the *log odds* of being on the higher level of being competent in developing valid modified Biology test items.

Table 4.12 Mean Statistics for Comparative Analysis (CA) of Modified and Original Test Items

		Statistics									
		CAQ 5	CAQ 6	CAQ 8	CAQ9	CAQ1 0	CAQ1 3	CAQ1 4	CA1 8	CAQ1 9	CAQ20
N	Valid	23	23	23	23	23	23	23	23	23	23
	Missing	0	0	0	0	0	0	0	0	0	0
Mean		3.87	2.04	2.48	4.04	4.22	4.22	3.74	4.00	3.48	1.22

Based on the information provided in Table 4.11, it is evident that the comparative analysis (CA) of item CAQ20 resulted in a mean score of 1.22 which demonstrates a Strong Disagree. This analysis involved input from subject matter experts and specialist teachers for learners with visual impairment, who used a Likert Scale to compare a modified test item to the original version. The strong disagreement expressed by the respondents indicates that the modified test item differed significantly in difficulty level and was not aligned with the same topic and objective as the original item. Notably, the respondents strongly disagreed, leading to the conclusion that the modified test item CAQ20 could not be considered valid. Therefore, it represented a 10% of the ten modified test items.

Test items CAQ6 and CAQ8 had the mean of 2.04 and 2.48 respectively. The respondents expressed them to be in the category of Disagree. This meant that the modified test items were not valid as compared to the original items with diagrams, pictures and tables. Respondents demonstrated that 20% of the test items were not valid.

The analysis confirmed that the subject matter experts and specialist teachers' responses of test item CAQ19 on the Likert Scale provided an average score of 3.48. This suggested a neutral overall response, reflecting their uncertainty about the item's validity and its alignment with the intended topic and objective to the original items with diagrams, pictures and tables. It represented a 10% of all the ten modified test items.

The average scores for test items CAQ5, CAQ9, CAQ10, CAQ13, CAQ14, and CAQ18 ranged between 3.74 and 4.22, falling within the "Agree" range. These results indicate that the respondents generally agreed that 60% of the modified test items were valid. However, for the remaining 40% of the items, the respondents disagreed, suggesting discrepancies in difficulty levels and alignment with the original items' topics and objectives.

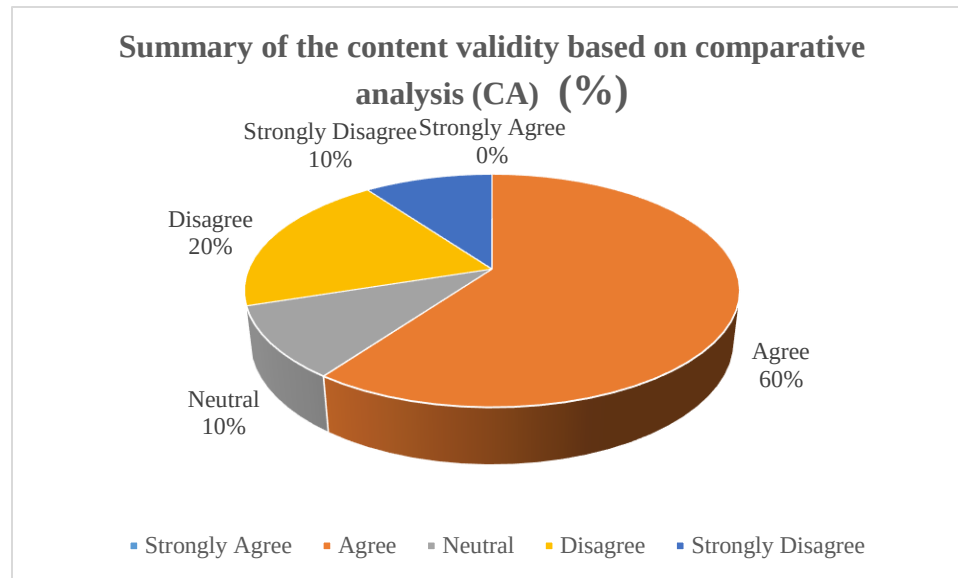


Figure 4.7 Summary of the content validity based on comparative analysis (CA)

The 60% of the modified test items were agreed to be valid, 20% of the items were disagreed, 10% of the modified test items were neutral, 10% were strongly disagreed to be valid while 0% of the items was said to be strongly agreed.

In this perspective, it showed that the 40% of the test items were not developed from the same topics and objectives and were invalid. In relation to the learners' responses from the modified and unmodified test items indicated that 40% of the modified items were more difficult than the original items. Apparently, they had come from the 40% of the items that did not comply to have been developed from the same topics and objectives or having the same difficulty level with the original items. Therefore, those invalid items were supposed to be restructured.

4.6.2 Evaluating Content Validity through Expert Review: Insights from Subject Matter Experts and Specialist Teachers for Learners with Visual Impairment.

The responses for the Subject Matter Experts (SMEs) and the Specialist Teachers for learners with visual impairment (STLVI) on their knowledge in modifying test items that had diagrams, pictures and tables into text format for learners with visual impairment was as below:

Table 4.13 Content Validity Analysis Based on the SMEs and the STLVI

		Statistics							
		1. Learnt test item development in college.	2. Have knowledge in test item modification.	3. I learnt test item modification for learners with visual impairment in college.	4. I know the guidelines for test item modification for learners with visual impairment.	5. It is important to have knowledge in test item modification.	6. I have knowledge in test item modification for items with diagrams, pictures, and tables for learners with visual impairment.	7. I did in-service training on modification of test items with diagrams, pictures and tables for learners with visual impairment.	8. It is important to have in-service training on modification of test items with diagrams, pictures and tables for learners with blindness.
N	Valid	23	23	23	23	23	23	23	23
	Missing	0	0	0	0	0	0	0	0
Mean		3.39	2.91	1.52	1.91	5.00	2.26	1.65	4.96
Median		4.00	3.00	1.00	2.00	5.00	2.00	1.00	5.00
Mode		4	4	1	1	5	1	1	5
Range		4	4	4	4	0	4	4	1
Minimum		1	1	1	1	5	1	1	4
Maximum		5	5	5	5	5	5	5	5

In the above 4.13 Table, the statistical results showed that statements 3, 4, 6 and 7 had the mean between 1.52 and 1.91 disagree range. It showed that the SMEs and STLVI expressed that they did not learn test item modification for learners with visual impairment in college. In addition, in statement 4 they expressed that they did not know the guidelines for test item modification for learners with visual impairment.

According to Section A of Likert Scale Tool on Test item development and modification (TDM) on statement 6 “*I had no knowledge in test item modification for test items with*

diagrams, pictures, and tables for learners with visual impairment” the respondents indicated as in the table below:

Table 4.14 Knowledge on Test Item Development and Modification (TDM)

Analysis	Score (Out of 23)	Percent (%)
Strongly Disagree	9	39.13
Disagree	4	17.39
Neutral	6	26.09
Agree	3	13.04
Strongly Agree	1	4.35
TOTAL	23	100

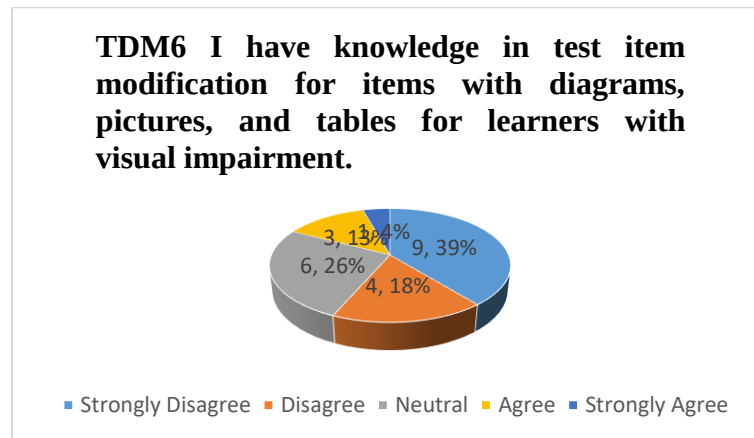


Figure 4.8 Knowledge in test item modification for test items with DPTs for learners with VI.

The pie graph for the statement 6 of the Likert Scale showed that the sum proportion of the respondents who strongly disagreed, disagreed and neutral was too high. That revealed SMEs and STLVI had no knowledge about item modification of items with

diagrams, pictures and tables. Therefore, it was difficult for them to modify items and be of the same difficult level index and discrimination to the original items.

On the statement 7 the SMEs and STLVI expressed that they did not have had in-service training on modification of test items with diagrams, pictures and tables for learners with visual impairment.

The frequency distribution of the Strongly Agree, Agree, Neutral, Disagree and Strongly Disagree responses for statement 3 was represented as it is in graph below:

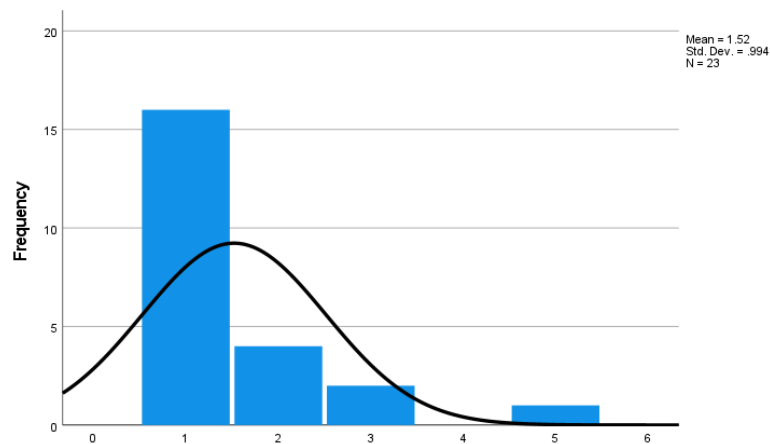


Figure 4.9 I learnt test item modification for learners with visual impairment in college.

The responses had a high frequency of 18 out of 23 on Strongly Disagree to the statement which was finding out if the respondents learnt test item modification for learners with visual impairment in college. In addition, there was no normal distribution of the responses on statement 3. It represented 69.6% of respondents who strongly disagreed.

On the other hand the responses for statements 1 and 2 on the Likert Scale;

1. *Learnt test item development in college.*
2. *Have knowledge in test item modification.*

... had the mean range between 2.91 and 3.39 meaning the SMEs and STLVI demonstrated to be neutral. On statement 1, the respondents demonstrated that they neither agreed nor disagreed whether they learnt test item development in college. Their neutrality meant that they had limited knowledge in test item modification. The analysis on statement 2 further indicated that their knowledge in test item modification was as well neither to agree nor disagree.

The frequency graphical representation of statement 2 was as below:

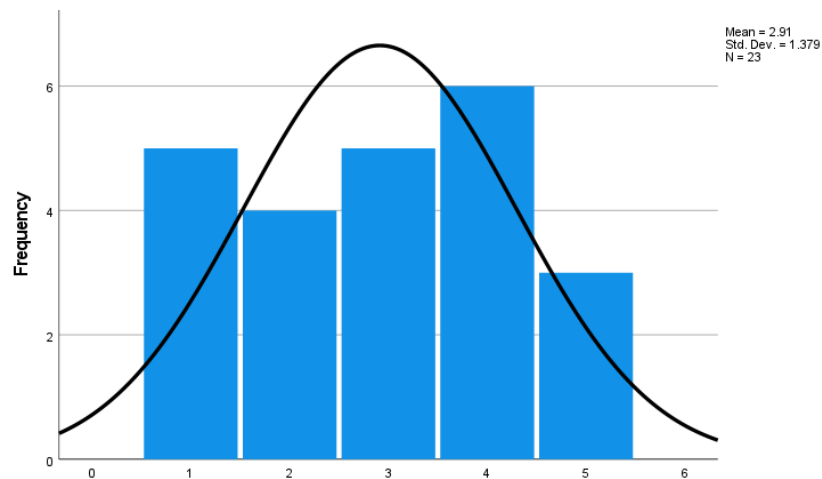


Figure 4.10 Have knowledge in test item modification.

The Agree had a higher frequency response of 6 out of 23 representing 26.1%. The Agree response had a higher response than the rest of other responses.

The mean of statements 5 and 8 responses was in the range between 4.96 and 5.00 which indicated that the SMEs and STLVI strongly agreed that it was really important to have knowledge in test item modification. On statement 8 the respondents strongly agreed that it was important to have in-service training on modification of test items with diagrams, pictures and tables for learners with visual impairment.

The graph below illustrates the frequency for statement 5:

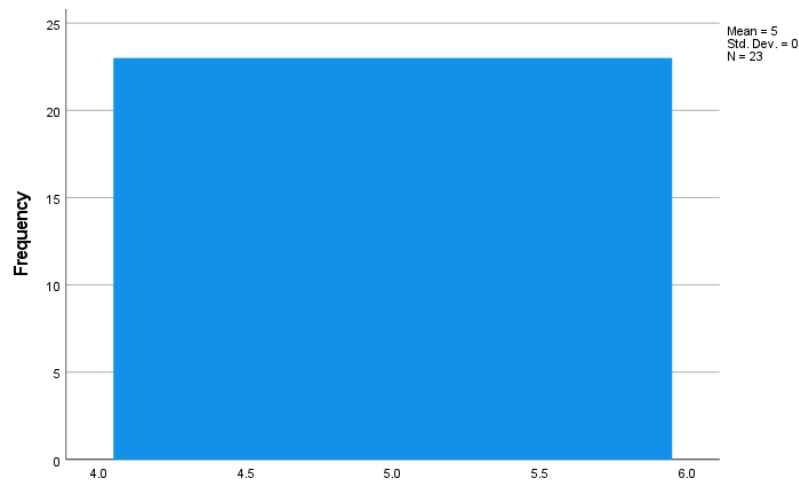


Figure 4.11 it is important to have knowledge in test item modification.

All the respondents expressed for the Strongly Agree which had the mean, median, mode, minimum and maximum of 5 with the frequency of all 23 respondents.

Therefore, the experience of SMEs and STLVI revealed that they did not have the good level of understanding of test item development and modification of test items that had diagrams, pictures and tables for learners with visual impairment. In addition, they demonstrated that there was a great need to provide them with skills in test item modification through in-service training. Furthermore, there was need to have guidelines in modification of test items that had diagrams, pictures and tables for learners with visual impairment.

In addition, the SMEs and STLVI showed that they did not have the good level of understanding of test item development and modification of test items that had diagrams, pictures and tables for learners with visual impairment.

Furthermore, they demonstrated that there was a great need to provide them with skills in test item modification through in-service trainings. On the same, there was need to have guidelines in modification of test items that had diagrams, pictures and tables for learners with visual impairment.

4.7 Chapter summary

The Chapter presented and discussed the analysis of the results based on the main objective and the specific objectives in terms of the validity and reliability of modified items with diagrams, pictures and tables. It presented the item characteristic curves, discrimination and guessing parameters. The analysis and discussion showed that the modified teacher-made Biology JCE Mock test items for learners with visual impairment were not reliable and were invalid.

CHAPTER FIVE

CONCLUSION AND RECOMMENDATIONS FOR FURTHER STUDIES

5.0 Chapter Overview

This Chapter presents findings of the study, conclusions and study's contribution to knowledge. The Chapter also presents proposed recommendations for further studies.

5.1 Findings of the Study

The study aimed to investigate the reliability and validity of modified teacher-made Biology test items for learners with visual impairment. The results found that 40% of the modified test items were more difficult than the original items that had diagrams, pictures and tables while 30% of the modified test items were less difficult in relation to the original items with diagrams, pictures and tables thereby making a total of 70% of the modified test items that were not reliable or valid.

The following modified items $QM9 = 0.79$, $QM13 = 0.56$, $QM14 = 0.73$ and $QM19 = 0.23$ were more difficult than the related items $QD9 = 0.82$, $QD13 = 0.65$, $QD14 = 0.90$ and $QD19 = 0.45$ that had diagrams, pictures and tables respectively. For example, the modified item $QM19$, only 230 participants out of 1,000 got it correct representing approximately 22.8% while 447 participants correctly answered the original item with

diagram QD19 representing 44.7% of the total. On the same, only one participant with visual impairments out of six participants got the item QM19 correctly.

On the other hand modified items QM6 = 0.53, QM8 = 0.75 and QM18 = 0.51 were less difficult than items with diagrams, pictures and tables QD6 = 0.39, QD8 = 0.55 and QD18 = 0.20 respectively.

The findings suggest that learners with visual impairment face unique challenges when taking examinations and require a better way of modifying items with diagrams, pictures, and tables by the subject matter experts and specialist teachers for learners with visual impairment to ensure fairness.

The content validity of modified Biology JCE Mock test items for learners with Visual Impairment found that they were not valid. Some modified items could not measure the same concept as the original items. An example of this was the modified item QM20. It was not developed from the same topic and objective as was the original item.

5.2 Conclusions

The results of the Study provided valuable response to the research topic. It became clear that most of the modified teacher-made Biology Mock test items did not have the same range of difficulty level as the original items with diagrams, pictures and tables. Fewer modified items had the same difficult range as compared to the original items, and some modified items were more difficult than the original items while others were less difficult. As such, the modified teacher-made Biology Mock test items were not reliable.

Furthermore, it was discovered that the subject matter experts and teachers specialized in teaching learners with visual impairments lacked sufficient knowledge to modify the

Biology test items with diagrams, pictures and tables. This was also observed when the majority of the SMEs expressed lacked knowledge in making modifications of items with diagrams, pictures and tables through analysis of the Likert Scale data. They lacked the guidelines for modification of test items that have diagrams, pictures and tables.

Additionally, some modified test items were not developed from the same topics or objectives as were the original items with diagrams, pictures and tables. Therefore, the modified items were not valid.

5.3 Recommendations for the Study's contribution to knowledge

1. The study's findings provide valuable insights for Subject Matter Experts (SMEs) and the Specialist Teachers for Learners with Visual Impairments (STLVI) to enhance their approach to modifying Biology test items that include diagrams, pictures, and tables. These modifications will help to ensure that learners with visual impairments are assessed equitably alongside their sighted peers.
2. Education institutions, examination bodies, and policymakers should consider developing comprehensive guidelines for modifying test items with visual diagrams, pictures, and tables to support inclusive assessments. This will enable learners with visual impairments to participate in tests and examinations that maintain the same level of difficulty and discrimination as those taken by sighted learners, ensuring fair and consistent grading.
3. Moreover, teacher training institutions should incorporate test item development and modification into their Testing, Measurement, and Evaluation curricula. This

will equip future educators with the necessary skills and knowledge to adapt assessments for learners with visual impairments and other disabilities.

4. Additionally, Montfort SNE College should offer a course in Testing, Measurement, and Evaluation for specialist teachers working with learners with visual impairments and other disabilities. This will provide them with the essential expertise in test development and modification, ensuring that they can create accessible assessments for all learners including those with visual impairments.

5.4 Proposed Recommendations for Further Research Studies

The Study recommends that further studies be conducted in the following areas:

1. An investigation on how continuous assessment of learners with visual impairment is done as they are preparing for national examinations.
2. An investigation on test item modifications for learners with cerebral palsy with respect to items for regular learners.
3. An exploration on Differential Item Functioning (DIF) between modified test items and the MANEB items in Biology.
4. An investigation on appropriate methods for modifying test items while maintaining the same level of difficulty and discrimination as the original items.

5.5 Chapter Summary

This Chapter has discussed findings of the study, the conclusions, study's contribution to knowledge and proposed recommendations for further studies.

REFERENCES

- Abdulmalik, R. (2013). Measurement and Evaluation in Education (PDE 105) Unit One: The Concepts of Test, Measurement, Assessment and Evaluation in Education. *Journal of Education and Training Studies*, 1(2), 54-67
- Academy for Educational Development, (2006). *Introduction to Data Analysis Handbook*. AED/TAC-12 Spring.
- Adedoyin, O.O. & Mokobi, T. (2013). 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.
- Agresti, A. (2010). *Analysis of Ordinal Categorical Data* (2nd ed.). Wiley.
- Ajeng, D. H. (2009), *The Analysis of Validity, Reliability, Discrimination Power and Level of Difficulty of First Mid-Term Test in the Case of the Eighth Grade Students of SMP 33 Semarang*. Semarang State University.
- Allman, C. (2009). *Making Tests Accessible for Students with Visual Impairments: A Guide for Test Publishers, Test Developers, and State Assessment Personnel* (4th edition.) American Printing House for the Blind.
- Ayanwale, M.A., & Ndlovu, M. (2021). Ensuring Scalability of a Cognitive Multiple-Choice Test through the Mokken Package in R Programming Language. *Educ. Sci.*, 11(1). <https://doi.org/10.3390/educsci11120794>

- Azzopardi, M. & Azzopardi, C. (2019, April 19-21). *The statistical analysis and evaluation of final advanced level biology examinations*. Paper presented at the Multidisciplinary Academic Conference in Prague.
- Baker, F. B. (2001). *The Basics of Item Response Theory* (2nd ed.). Wisconsin, ERIC Clearinghouse on Assessment and Evaluation.
- Balov, N. (2016). *Bayesian Binary Item Response Theory Models Using Bayesmh*.
<http://blog.stata.com/2016/01/18/bayesian-binary-item-response-theory-models-using-bayesmh/>.
- Bichi, A. A. (2016). Classical test theory: An Introduction to Linear Modeling Approach to Test and Item Analysis. *International Journal for Social Studies*, 2(9), 27-36.
- Borghi J, et.al. (2018) Support Your Data: A Research Data Management Guide for Researchers. *Research Ideas and Outcomes* 4(1), e26439.
- Camilli, G. (2017) The Scaling Constant D in Item Response Theory. *Open Journal of Statistics*, 7(1), 780-785. <http://doi.org/10.4236/ojs.2017.75055>.
- Coaley, K. (2010). *An Introduction to Psychological Assessment and Psychometrics*. SAGE Publications Ltd.
- Committee on Special Education (2016). *Test Modifications*. Author
- Creswell, J. W. (2014). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (4th ed.). SAGE Publications, Inc.
- Creswell, J.W. & Creswell, J.D. (2017) *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (4th ed.). Sage

- Crocker, L. & Algina, J. (2008). *Introduction to Classical and Modern Test Theory*. Cengage Learning .
- Cronbach, L.J. (2004). My Current Thoughts on Coefficient Alpha and Successor Procedures. *Educational and Psychological Measurement*, 64(1), 391–418.
- Curtin, R., Presser, S., & Singer, E. (2005). Changes in Telephone Survey Nonresponse Over the Past Quarter Century. *Public Opinion Quarterly*, 69(1), 87-98.
- David, M. (2013). A note on the item information function of the four-parameter logistic model. *Applied British Journal of Mathematical and Statistical Psychology*, 63(3), 509–525.
- De Ayala R. J., (2013). *The Theory and Practice of Item Response Theory*. The Guilford Press.
- Elliott, S. N., Kettler, R. J., Beddow, P. A., Kurz, A., Compton, E., McGrath, D., Bruen, C., Hinton, K., Palmer, P., Rodriguez, M. C., Bolt, D. & Roach, A. T. (2010). Effects of Using Modified Items to Test Students with Persistent Academic Difficulties. *Exceptional Children*, 76(4), 475–495.
<https://doi.org/10.1177/001440291007600406> 08/06/2023.
- Embretson, S. E., & Reise, S. P. (2013). *Item Response Theory* (2nd ed.). Psychology Press.
- Fowler, F.J (2009). *Sampling. Survey Research Methods* (4th ed.). Sage Publications

- Ghasemi , A. & Zahediasl, S. (2012). Normality Tests for Statistical Analysis: A Guide for Non-Statisticians. *Int J Endocrinol Metab.*, 10(2), 486-9.
<http://doi.org/10.5812/ijem.3505>.
- Haberman, S. J. (2016), *Statistics in the Social and Behavioral Sciences Series: Handbook of Item Response Theory Volume Two Statistical Tools*. CRC Press.
- Habulezi, J., Kefilwe, P. J. Batsalelwang & Malatsi, N. M. (2017). Factors Influencing the Poor Academic Performance of Learners with Vision Impairment in Science Subjects in Kgatleng District in Botswana. *International Journal of Learning, Teaching and Educational Research*, 16(11), 28-44.
- Hambleton, R. K. & Slater, S. C. (1997). Item Response Theory Models and Testing Practices: Current International Status and Future Directions. *European Journal of Psychological Assessment*, 13(1), 21–28.
- Hopkins P. & Unger M. (2017). What is a Subject Matter Expert? *The Journal of Pipeline Engineering*, 16(4), 227-130.
- Hotiu, A. (2006). *The relationship between item difficulty and discrimination indices in multiple-choice tests in a Physical science course (MSc thesis)*. Florida, Atlantic University.
- Israel, M., & Hay, I. (2006). *Research Ethics for Social Scientists: Between Ethical Conduct and Regulatory Compliance*. Sage.

- Johari, J., Sahari, J., Abd Wahab, D., & Abdullah, S. (2011) Difficulty index of examinations and their relation to the achievement of programme outcomes. *Procedia - Social and Behavioral Sciences*, 18(1), 71-80.
- Konala, L. J. (2020). *Examining the Quality of Form One Selection Test for Faith Mission Secondary Schools: - Case of Anglican Diocese of Upper Shire in Malawi*. Zomba, University of Malawi.
- Livingston, S. A. (2018). *Test reliability—Basic concepts* (Research Memorandum No. RM-18-01). Educational Testing Service.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Lawrence Erlbaum.
- Macdonald, P. & Paunonen, S. V. (2002) A Monte Carlo comparison of item and person statistics based on item response theory versus classical test theory. *Educational and Psychological Measurement*. 62(1), 921-943.
- Manichander, T. (2016), *Evaluation in Education- Educational Evaluation*. Laxmi Book Publication.
- Maydeu-Olivares, A. & Forero, C. (2010). Goodness-of-Fit Testing. *Journal of International Encyclopedia of Education*, 7(1), 190-196.
- McLinden, M., Ravenscroft, J., Cobb, R., Douglas, G., & Hewett, R., (2017). The Significance of Specialist Teachers of Learners with Visual Impairments as Agents of Change: Examining Personnel Preparation in the United Kingdom

- through a Bioecological Systems Theory. *Journal of Visual Impairment & Blindness*, 111(6), 569-584.
- Middleton, F. (2022). *The 4 Types of Validity in Research | Definitions & Examples*. Scribbr. <https://www.scribbr.com/methodology/types-of-validity/>
- Mitra, N. K., Nagaraja, H. S, Ponnudurai, G., & Judson, J. P. (2009). The Levels of Difficulty and Discrimination Indices in Type A Multiple Choice Questions Of Pre-clinical Semester 1 Multidisciplinary Summative Tests. *International e-Journal of Science, Medicine & Education*, 3 (1), 2-7.
- Morgan, D.L. (2007). Paradigms Lost and Pragmatism Regained: Methodological Implications of Combining Qualitative and Quantitative Methods. *Journal of Mixed Methods Research*, 1(1), 48-76.
- Muzingo, L. (2018). *A Measurement of Inter-Rater Agreement of Modified Examination Scores Using Generalizability Theory. A Case Study of Malawi School Certificate of Education Biology Paper 1*. Zomba, University of Malawi.
- Ogunsakini, I. B. & Shogbesani, Y. O. (2018). Item Response Theory (IRT): A Modern Statistical Theory for Solving Measurement Problem in 21st Century. *International Journal of Scientific Research in Education*, 11(3B), 627-635.
- Osterlind, S. J. (2002). *Constructing Test Items: Multiple-Choice, Constructed-Response, Performance, and Other Formats* (2nd ed.). Kluwer Academic Publishers.
- Osuji, U.S.A & Okonkwo, C.A. (2006). *Course Guide Edu 403: Measurement and Evaluation*. National Open University of Nigeria

- Pena, C. S., Costa, M. A., & Braga, R. P. (2018). A New Item Response Theory Model to Adjust Data Allowing Examinee Choice. *PLoS ONE* 13(2), e0191600. <https://doi.org/10.1371/journal.pone.0191600>.
- Philip, A. & Ojo, B. O. (2017). Application of Item Characteristic Curve (ICC) in the Selection of Test Items. *British Journal of Education*, 5(2), 21-41.
- Punch, K. F. (2005). *Introduction to social research: Quantitative and qualitative approaches* (2nded.). Sage.
- Rezigalla, A. & Eleragi, A. & Ishag, M. & Mohammed, A. (2020). Comparison between students' perception and examination validity, reliability and items difficulty: a cross sectional study. *Sudan Journal of Medical Sciences*, 15(2), 114-23. <http://doi.org/10.18502/sjms.v15i2.5503>.
- Singh, A. S. & Masuku, M. B. (2014). Sampling Techniques and Determination of Sample Size in Applied Statistics Research: An Overview. *International Journal of Economics, Commerce and Management United Kingdom*, II(11), 87-96.
- Siri, A. & Freddano, M. (2011). The Use of Item Analysis for the Improvement of Objective Examinations. *Social and Behavioral Sciences*, 29 (2011), 188–197.
- StataCorp. (2023). *Stata 18 Item Response Theory: Reference Manual*. Stata Press.
- Stone, E., Cook, L., Laitusis, C., & Cline, F. (2010). Using Differential Item Functioning to Investigate the Impact of Testing Accommodations on an English-Language Arts Assessment for Students who are Blind or Visually Impaired, *Journal of Applied Measurement in Education*, 23(2), 132-152.

- Thissen, D. (1991). *Multilog: Multiple category item analysis and test scoring using item response theory [computer software]*. Scientific Software International.
- UNESCO (2005). *Guidelines for Inclusion: Ensuring Access to Education for All*. Author
- Watering, G. van de. (2006). Teachers' and students' perceptions of assessments: A review and a study into the ability and accuracy of estimating the difficulty levels of assessment items. *Educational Research Review*, 2(1), 133-47.
- Webb, N.M., Shavelson, R. J., & Haertel, E. (2006). Reality Coefficient and Generalizability Theory. *Handbook of Statistics*, 26(1), 81-124.
- Wilson, P.M., et al. (2010), Disseminating Research Findings: what Should Researchers Do? A Systematic Scoping Review of Conceptual Frameworks. *Implementation Science*, 5(1). <https://doi.org/10.1186/1748-5908-5-91>.
- Winter, P. C., Hansen, M., & McCoy, M. (2019). Ensuring the comparability of modified tests administered to special populations. *CRESST Report 864*.
- Wu, C., Thompson, M.E. (2020). Stratified Sampling and Cluster Sampling. In *Sampling Theory and Practice. ICSA Book Series in Statistics*. Springer, Cham. https://doi.org/10.1007/978-3-030-44246-0_3.
- Xinming, A. & Yung, Y.F. (2014). *Item Response Theory: What It Is and How You Can Use the IRT Procedure to Apply It*. SAS Institute Inc.
- Yaghmale, F. (2003). Content Validity and Its Estimation. *Journal of Medical Education*, 3(1), 76-85.

- Zahran, E. & Mustafa, A. (2023). Item analysis of multiple choice questions of a genetic term exam in an Egyptian Veterinary college for a viable question bank. *Educational Evaluation and Learning Management Journal*, 1(1), 1-5.
- Zhao, Y. (2006). How to Design and Interpret a Multiple-Choice-Question Test: A Probabilistic Approach. *International Journal Engaging Education*, 22(6), 1281 – 1286.

APPENDICES

APPENDIX 1: Modified Teacher-made Biology JCE Mock Test Items.

BLANTYRE SECONDARY SCHOOL CLUSTER

2020 JCE MOCK EXAMINATIONS

BIOLOGY (100 MARKS)

SUBJECT NUMBER: J022

Time Allowed: 30 ~~mins~~

Date: Monday, 2nd August

08:30 – 09:00 am

INSTRUCTIONS:

1. This paper has 2 pages. Please check.
2. Answer **all** multiple choice questions
3. Write your **Name** and **Sex** on top of every page.

Section A: Multiple choice (20 marks)

Answer ALL questions:

1. Phloem is one of the plant tissues involved in transportation. One of the following is an adaptation of the tissue for its functions.
A. Companion cell fails to carry out some life processes of the phloem
B. Pores in sieve plate allow sugars to pass from one cell to the other
C. Pores in sieve plate deny sugars passage from one cell to the other
D. Absence of sieve plate and companion cell
2. By what process does a nitrate ion move into a root hair cell when its concentration is higher in the root hair cell than in the soil?
A. Diffusion
B. Absorber
C. Osmosis
D. Active transport
3. The following is not true about the differences between disaccharides and polysaccharides.
A. Disaccharides are sweet while polysaccharides are not sweet
B. Disaccharides are insoluble in water while polysaccharides are soluble in water
C. Disaccharides are white crystalline solids while polysaccharides are not crystalline
D. Disaccharides are soluble in water while polysaccharides are insoluble in water
4. Which of the following is correct about a villus being suitable for food absorption?
A. Has cartilage which keeps it firm and open
B. Has a dense network of blood capillaries for warmth
C. Has a thick epithelium for fast diffusion
D. Is neutralized for easy gaseous exchange
5. Name the valve that is found at the left side of the human heart.
A. Bicuspid
B. Auricle
C. Tricuspid
D. Ventricle
6. Which of the following statements is true about blood flowing through pulmonary artery and Aorta.
A. **Aorta** has more oxygen than pulmonary artery
B. Aorta has more carbon dioxide than pulmonary artery
C. Pulmonary artery has more oxygen than Aorta
D. Pulmonary artery is faster than Aorta.
7. Which of the following is the best condition for blood transfusion?
A. match in blood ~~colour~~

- B. match in blood serum
C. match in blood thickness
D. match in blood groups
8. Which of the following characteristics could be used to classify living things.
A. locomotion
B. Pirious
C. Cells
D. Cytoplasm
9. To which of the following groups of animals does a lizard belong?
A. Fish
B. Reptile
C. Mammal
D. Amphibian
10. What type of leaf arrangement belong to compound leaf?
A. Digitate
B. Bipinnate
C. Palmate
D. Trifoliate
11. Why should living things respond to changes in their environment?
A. For survival in the environment
B. To avoid making similar mistakes
C. For adaptation in their environment
D. To be protected from their enemies
12. What is the function of a nucleus in a cell?
A. To select substances that enter or leave the cell
B. To determine the size and shape of the cell
C. To allow gases and water to enter the cell
D. To control activities of the cell
13. What type of food is eaten by sunbird?
A. Seeds
B. Nectar
C. Flesh
D. Grass
14. How is an eagle adapted to its mode of feeding? It has _____.
A. curved and strong beak
B. long and slender beak
C. conical and long beak
D. long and strong beak
15. The best way of preventing worm infestation is by proper _____.
A. Waste disposal
B. Sanitation and sewage disposal
C. Cooking of meat
D. Personal cleanliness
16. What is used to make a scientific name?
A. Genus and phylum
B. Genus and class
C. Genus and species
D. Species and kingdom
17. How does high fibre content in diet prevent constipation?
A. It produces water in large intestines
B. It promotes peristalsis
C. It mixes up to form chime
D. It stimulates action of enzymes in intestine
18. Which of the following micro-organisms grows inside a living cell only?
A. Yeast
B. Bacteria
C. Virus
D. Euglena
19. Which of the following micro-organisms are able to move through liquids like water?
A. Yeast and virus
B. bacteria and Euglena
C. virus and Euglena
D. Yeast and bacteria
20. Which part of the brain is used for body balance and posture?
A. Cerebellum
B. Medulla Oblongata
C. Cerebrum
D. Pons
- END OF QUESTION PAPER**

APPENDIX 2: Original Biology JCE Mock Test Items with diagrams, pictures and tables.

BLANTYRE SECONDARY SCHOOL CLUSTER

2020 JCE MOCK EXAMINATIONS

BIOLOGY (100 MARKS)

SUBJECT NUMBER: J022

Date: Monday, 2nd August

Time Allowed: 30 ~~mins~~

08:30 – 09:00 am

INSTRUCTIONS:

1. This paper has 2 pages. Please check.
2. Answer **all** multiple choice questions
3. Write your **Name** and **Sex** on top of every page.

Section A: Multiple choice (10 marks)

Answer **ALL** questions.

Figure 1 is a diagram of a human heart. Use it to answer **questions 1 and 2** that follow



Figure 1

1. Part labeled **S** is _____.
 - A. Bicuspid
 - B. Auricle
 - C. Tricuspid
 - D. Ventricle
2. Which blood vessel is an Aorta?
 - A. U
 - B. S
 - C. T
 - D. U and T

Figure 2 is a diagram of vertebrates. Use it to answer the **questions 3 and 4**.

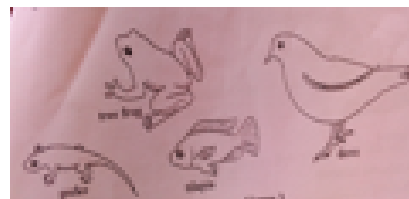


Figure 2

3. Which of the following characteristics could be used to classify the organisms.
 1. Fins
 2. Legs
 3. Toes
 4. Size
 - A. 1, 2 and 3
 - B. 1, 2 and 4
 - C. 1, 3 and 4
 - D. 2, 3 and 4
4. To which of the following groups of animals does a lizard belong?
 - A. Fish
 - B. Reptile
 - C. Mammal
 - D. Amphibian

Figure 3 is a cassava leaf. Use it to answer question 5 that follows.

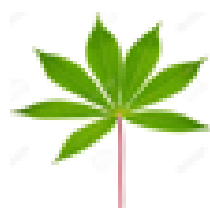


Figure 3

5. What type of leaf arrangement is shown in the diagram?
- Digitate
 - Bininate
 - Palmate
 - Trifoliate

Figure 4 shows beaks of birds **Y** and **Z**. Use it to answer questions 6 and 7.



Figure 4

6. What type of food is eaten by **Y**?
- Seeds
 - Nectar
 - Flesh
 - Grass
7. How is **Z** adapted to its mode of feeding? It has _____
- curved and strong beak
 - long and slender beak
 - conical and long beak
 - long and strong beak

Figure 5 Shows diagrams of different micro-organisms. Use it to answer questions 8 and 9.



Figure 5

8. Which of the following micro-organisms grows inside a living cell only?
- W**
 - X**
 - Y**
 - Z**
9. Which of the following micro-organisms are able to move through liquids like water?
- W and Y**
 - X and Z**
 - Y and Z**
 - W and X**
10. A specimen, 13cm is magnified 2.5 times its size. What is the size of the image or drawing?
- 5.2cm
 - 11.5 cm
 - 32.5 cm
 - 15.5 cm

END OF QUESTION PAPER

APPENDIX 3: Likert Scale Instrument

UNIVERSITY OF MALAWI



TEST ITEM EVALUATION FORM FOR BIOLOGY SUBJECT MATTER EXPERTS

ROBIN FRANCIS CHATAIKA

Dear ... I am a student doing Master of Education (Testing, Measurement and Evaluation) at the University of Malawi. I am conducting a study on: **"Investigating Validity and Reliability of Modified Teacher Made Biology Test Items for Learners with Visual Impairment"**. You are kindly asked to participate in this study. I appreciate your participation. The information you provide will be treated with confidentiality. Please be as transparent as possible.

- a. SEX: Male (☐) Female (☐)
- b. Specialist in:
 - (i) Biology: (☐)
 - (ii) Other subject areas (☐) Please specify _____
 - (iii) Learners with visual impairment: (☐)
- c. Teaching experience:
 - (i) Less than 5 years: (☐)
 - (ii) Between 6 and 10 years: (☐)
 - (iii) More than 10 years: (☐)
- d. The highest professional qualification:
 - (i) Diploma (☐)
 - (ii) Degree (☐)
 - (iii) Master of Education (☐)
 - (iv) Other (☐) Specify _____

In Inclusive Education (IE), Biology test items with diagrams, pictures and tables are modified into text format in order to be converted into Braille for learners with visual impairment. The modification must conform to issues of reliability and validity as it is in the original print copy of the test items.

Instructions

1. The rating key for the level of your understanding: (1 – Strongly Disagree; 2 – Disagree; 3 – Neutral; 4 – Agree; 5 – Strongly Agree)
2. Please select how much you agree or disagree with the following statements.
3. Do NOT write your name as it is confidential.

Focus areas	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
SECTION A: Test item development and modification (TDM)					
1. I learnt test item development in college.	1	2	3	4	5
2. I have knowledge in test item modification.	1	2	3	4	5
3. I learnt test item modification for learners with visual impairment in college.	1	2	3	4	5
4. I know the guidelines for test item modification for learners with visual impairment.	1	2	3	4	5
5. It is important to have knowledge in test item modification.	1	2	3	4	5
6. I have knowledge in test item modification for items with diagrams, pictures, and tables for learners with visual impairment.	1	2	3	4	5
7. I did in-service training on modification of test items with diagrams, pictures and tables for learners with visual impairment.	1	2	3	4	5
8. It is important to have in-service training on modification of test items with diagrams, pictures and tables for learners with blindness.	1	2	3	4	5
SECTION B: The comparative analysis (CA) between modified and original test items that demonstrates validity, same difficulty level, same objectives and topic.					
9. Question 5	1	2	3	4	5
10. Question 6	1	2	3	4	5
11. Question 8	1	2	3	4	5
12. Question 9	1	2	3	4	5
13. Question 10	1	2	3	4	5
14. Question 13	1	2	3	4	5
15. Question 14	1	2	3	4	5
16. Question 18	1	2	3	4	5
17. Question 19	1	2	3	4	5
18. Question 20	1	2	3	4	5

END OF EVALUATION

**APPENDIX 4: University of Malawi Research Ethics and Regulatory Committee
Approval and Permit for Protocol No. P.07/23/276**



VICE-CHANCELLOR
Prof. Samson Sajida, BSc MSc, MPhil Cantab, PhD MEd

CHANCELLOR COLLEGE
P.O. Box 280, Zomba, Malawi

Our Ref: P.07/23/276

Telephone: (265) 1 526 622
Fax: (265) 1 524 031
E-mail: vc@unimarec.unimw.ac.mw

Your Ref.:

29th September 2023

Mr Robin Francis Chataika
Master of Education (TME)
University of Malawi
P.O. Box 280
Zomba

Dear Mr. Chataika,

**RESEARCH ETHICS AND REGULATORY APPROVAL AND PERMIT
FOR PROTOCOL NO. P.07/23/276: INVESTIGATING CONTENT
VALIDITY AND RELIABILITY OF MODIFIED TEACHER MADE
BIOLOGY JCE MOCK TEST ITEMS FOR LEARNERS WITH VISUAL
IMPAIRMENT: THE CASE OF SELECTED SECONDARY SCHOOLS
IN SOUTH WEST EDUCATION DIVISION, MALAWI.**

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
The Registrar
Director of Finance and Investments
Acting Head of Research
Chairperson UNIMAREC
UNIMAREC Compliance Officer

APPENDIX 5: Department of Education Foundation Letter of Introduction: Robin F. Chataika

	VICE-CHANCELLOR Prof. Samson M.I. Sijida, BSc: MSc, MPhil, DPhil, PhD MSc » Connect with Excellence	UNIVERSITY OF MALAWI P.O. Box 280, Zomba, Malawi TEL: (265) 1 524 222 FAX: (265) 1 524 046 EMAIL: vc@unima.ac.mw
---	---	---

Our Ref: _____
Your Ref: _____

2nd October, 2023

TO WHOM IT MAY CONCERN

LETTER OF INTRODUCTION: ROBIN CHATAIKA

This letter serves to confirm that **Mr. Robin Chataika** is a registered postgraduate student in the Education Foundations Department, of the School of Education, in the University of Malawi. He is studying under the Master of Education (Testing, Measurement & Evaluation) program. His registration Number is MED/MEV/08/21.

Mr. Chataika has completed his first year of studies which mainly involves coursework. As a requirement for completion of his study program, he is conducting a research titled: ***“Investigating the validity and reliability of modified teacher made test items for Biology Form Two learners with visual impairment: The case of visual impairment learners in selected secondary schools in the South West Education Division (SWED) in Malawi”***. This letter therefore, serves to request your institution to assist our student to collect the required data.

For any inquiries please contact the undersigned via the following email address: med@cc.ac.mw

Sincerely yours,



Symon Winiko, PhD.
HEAD OF DEPARTMENT – EDUCATION FOUNDATIONS

UNIVERSITY OF MALAWI

02 OCT 2023

SCHOOL OF EDUCATION
P.O. BOX 280, ZOMBA

APPENDIX 6: A letter requesting permission for data collection to the South West Education Division Manager (EDM)

Robin Francis Chataika

Montfort SNE College

P.O. Box 5554

LIMBE

2nd October, 2023

The Education Division Manager

South West Education Division (SWED)

Private Bag 386

Chichiri

Blantyre 3

Dear Sir,

REQUEST FOR A PERMISSION TO CARRY OUT RESEARCH IN SOUTH WEST EDUCATION DIVISION SECONDARY SCHOOLS

I write to request for the permission to carry out research study in South West Education Division Secondary Schools on topic titled: INVESTIGATING CONTENT VALIDITY AND RELIABILITY OF MODIFIED TEACHER MADE BIOLOGY JCE MOCK TEST ITEMS FOR LEARNERS WITH VISUAL IMPAIRMENT: THE CASE OF SELECTED SECONDARY SCHOOLS IN SOUTH WEST EDUCATION DIVISION, MALAWI.

Attached is the letter of approval from the University of Malawi Research Ethics Committee and Regulatory Approval and Permit for Protocol No. P.07/23/276.

I am looking forward to your approval.



Robin Francis Chataika

APPENDIX 7: Authority to Conduct Research in South West Education Division (SWED)

All correspondences should be addressed to:
The Education Division Manager



In reply please quote ref no. SWED/L/3

SOUTH WEST EDUCATION DIVISION
PRIVATE BAG 386
CHICHIRI
BLANTYRE 3
MALAWI

3RD OCTOBER, 2023

TO WHOM IT MAY CONCERN

AUTHORITY TO CONDUCT RESEARCH

I write to kindly request your office to allow ROBIN CHATAIKA to undertake research activities at your institution.

He is a postgraduate student at University of Malawi pursuing a Master of Education in Testing, Measurement and Evaluation and is conducting research titled: **Investigating the validity and reliability of modified teacher made test items for Biology Form 2 learners with visual impairment: The case of visual impairment learners in selected secondary schools in South West Education Division(SWED) in Malawi.**

I would be most grateful if he is given all the necessary support and guidance so that his research activities are carried out successfully.

I look forward to your usual support and hoping at the same time that you will accord this request all the attention and urgency that it deserves.

LAWRENCE UDEDI -SQAO

FOR: EDUCATION DIVISION MANAGER (SWED)

