EXAMINING THE QUALITY OF FORM ONE SELECTION TEST FOR FAITH MISSION SECONDARY SCHOOLS: - CASE OF ANGLICAN DIOCESE OF UPPER SHIRE IN MALAWI

MASTER OF EDUCATION (TESTING, MEASUREMENT AND EVALUATION)

THESIS

LUKE JAMES KONALA

UNIVERSITY OF MALAWI

CHANCELLOR COLLEGE

FEBRUARY 2020

EXAMINING THE QUALITY OF FORM ONE SELECTION TEST FORFAITH MISSION SECONDARY SCHOOLS: - CASE OF ANGLICAN DIOCESE OF UPPER SHIRE IN MALAWI

MASTER OF EDUCATION (TESTING, MEASUREMENT AND EVALUATION)

THESIS

By

LUKE JAMES KONALA

B.Ed (Science) –Domasi College of Education [Malawi]

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

University of Malawi

Chancellor College

February 2020

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.

_____

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 our approval.

Signature: _____Date:_____

BOB W. CHULU, PhD (Lecturer)

Main Supervisor

Signature: _____Date:_____

FRANK MTEMANG'OMBE, PhD (Lecturer)

Co-Supervisor

DEDICATION

This research study is dedicated to my wife, Felecia Konala, my father, Mr. Chiudyakumalembe and my grandmother, Anna Nasineya Mbilikira (Ambuya) for bringing me up and taking education seriously, despite being uneducated.

To my four children Davie, Confidence, Queen and Dominic, this work should arouse your interest to wish to achieve more in education.

ACKNOWLEDGEMENTS

The success of this research study is deeply indebted to several peoples' wisdom, support, guidance and encouragement. I am grateful to **Dr. B.W. Chulu** my main supervisor and **Dr. F. Mtemang'ombe** my co-supervisor for their scholarly advice on my work. I thank them for their patience and sparing part of their time to supervise my work. I deeply give them a credit.

Secondly, I would like to thank my family members for giving me moral, social and financial support to complete my studies. The time taken to complete theory and the research work was so extensive that patience was required by all my family members.

Thirdly, I would like to thank the management of School of Education at Chancellor College, University of Malawi for making it possible for me to rise in the hierarchy of personal academic status.

Finally I thank the Almighty God for blessing me throughout my studies.

ABSTRACT

The study examined the quality of ADUS selection tests. Specifically, the study sought to examine the level of error of measurement in selection of faith mission selected students (FMSS) by computing and analyzing the item-difficulty parameter, item-discrimination parameter and reliability level of the test items. Due to scarcity of research studies on these parameters on faith selection (ADUS selection) tests, it was worthwhile to conduct the study. A total of 1003 Standard 8 examinees out of 8569 were randomly sampled and 2015 ADUS selection test was administered to generate data (scores). Descriptive research design with cohort longitudinal survey and quantitative approach were employed. Subject Matter professionals scored scripts of the test to obtain test scores of examinees. The responses of the examinees were computed using BILOG 3.0 to obtain item-difficulty, item-discrimination and test reliability. The study revealed that ADUS selection test items were from the syllabus of Malawi curriculum. However, examinees performed very low on the test. The findings on parameters indicated that most of the items (83%) were very difficult. As for the discrimination power most items (89%) were out of desirable range of +1.0 to +2.0. The large percentage of items being poor on discrimination (power) parameter decreased drastically the quality of test. On reliability the test produced statistical value of 0.65. The test indicated low reliability since failed to achieve minimum recommended statistic value of 0.70. The study implies that the majority of items in ADUS test had poor test quality parameters and could increase errors in selection process of form one examinees. The study recommended doing 'item analysis' as one way to improve the quality of ADUS selection tests to minimize the level of error of measurement in selection process of faith mission selected students (FMSS).

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

ABBREVIATIONS AND ACRONYMS

1PLM	One Parameter Logistic Model
2PLM	Two Parameter Logistic Model
3PLM	Three Parameter Logistic Model
ACEM	Association of Christian Educator in Malawi
ACT	American College Test Administration.
ADUS	Anglican Diocese of Upper shire
AMAC	Australian Medical Assessment Collaboration
AT	Aptitude Test
CCAP	Church of Central Africa Presbytery
CDSS	Community Day Secondary School
CTB	Church Together in Britain
CTT	Classical Test Theory
DIF	Differential Item Functioning
EMIS	Education Management Information System
GRE	Graduate Record Examinations
ICC	Item Characteristics Curve
IRT	Item Response Theory
ITBS	Iowa Test of Basic Skills
KCPE	Kenya Certificate of Primary Education
MANEB	Malawi National Examination Board
MAT	Metropolitan Achievement Tests
MCC	Malawi Correspondence College
MCDE	Malawi college of distance education

MOEST	ministry of Education Science and Technology
MOU	Memorandum of Understanding
PLCE	Malawi Leaving Certificate of Education
SAT	Stanford Achievement Test
SDA	Seventh Day Adventist
SE	Standard Error of Estimation
SEED	South East Education Division
SPSS	Statistical Package for Social Scientists
ST	Selection Test
STDT	Standard Test
TCC	Test Characteristic Curve
TN	Terra Nova
TPL	Test Partnership Limited
US DLETA	United States-Department of Labor, Employment and Training Administration
USA	United States of America

CHAPTER 1

INTRODUCTION

1.1 Chapter overview

This chapter gives about accessibility level of secondary school education in Malawi, historical background of selection of students in Malawi, admission policies in education, history of ADUS and general quality of a good selection test. Furthermore, achievement levels of FMSS and GSS in faith mission schools, statement of the problem, purpose of the study, research questions, significance of study, limitations of study and attempts to minimize them and operational definition of terms .

1.2 Background

1.2.1 Access to secondary school education in Malawi

Different research and surveys conducted reveal that there is stiff competition for primary school (standard eight) pupils to be selected to secondary schools in Malawi due to limited space and increase in enrolment and high completion rate in primary schools. Chimombo, Meke, Zeitlyn, & Lewin (2014) report indicated the statistical increase in primary education enrolment from 1,795,451 in 1993 to 2,805,785 in 1998 and further to 4,034,220 in 2011. The increase in enrolment accelerated in primary schools demanded for space in secondary schools since the secondary school sector could not absorb all candidates who had passed standard (grade) 8 to start secondary education in Malawi. Woltjer (2006) found

that in 2006 over 80,000 pupils got primary School Leaving Certificate, and only about 8000 were selected to secondary schools which represented 10% of total pupils. That selection rate was both government and faith mission school "**selection tests**". Those who were left out on selection list for that year though legible were forced to repeat standard 8 or look for secondary school places in private schools and "open secondary school" and those that could not afford to pay they drop out of school, (World Bank, 2004; Makori, Cheboiwo, Yegon & Kandie, 2015).

1.2.2 Historical background of selection and admission policy of students in Malawi

The policy of admitting students into secondary school using selection test in Malawi was devised in 1971 to ease the growing demand for place in secondary schools, once students had finished standard (grade 8) eight in primary schools. The admission policy during the early years of independence was based on the relationship between a secondary school and the denomination of a student. The pupils from the Roman Catholic Church were selected into Roman Catholic Church (RC) built secondary schools while those from the protestant churches their children were admitted into the protestant built secondary schools. According to Sandikonda (2013) the policy promoted discrimination among students based on denominations and RC students had a lot of chances of being admitted to secondary schools. Until 1974 when the new policy was formulated and implemented which allowed all the children from religious background and non religious background to be selected into the secondary schools based on merit which focused on a child's ability and performance in the standard 8 examinations. The policy also met with challenge due to increase of

students since the policy had nothing to do with admitting students in Malawi College of Distance Education (MCDE). The situation led to Policy change in 1994 when government decided to convert Malawi MCDEs into CDSSs as one way of increasing access to secondary schools (Bisika, 1996). Banda (1996) suggested that Malawi, introduced the admission policy into secondary schools using selection test, as a way to control all the secondary schools in Malawi with the aim that the best students known as the 'cream' would be selected to the grant aided secondary schools, second to the 'cream' to the district conventional secondary schools and the rest who were composed of 'weak' and the average students would be left out to look for places at Malawi Correspondence College (MCC) now called the Community Day Secondary Schools (CDSS) (Sandikonda, 2013).

1.2.3 Admission policy in faith mission schools in Malawi

Later the faith mission organizations (churches) also fully adopted the system and started selecting their faithful (members of the church) into their faith mission secondary schools using selection tests as well which they started constructing locally themselves. But each faith mission organization which had schools produced and administered its own selection test locally for selecting faith (church) members into its secondary schools in different levels (grades).

Anglican Diocese of Upper Shire (ADUS) was no exception according to Association of Christian Educators in Malawi (ACEM, 2007). ADUS is one of the faith mission organizations along with other groupings like Roman Catholic (RC), Church of Central African Presbytery (CCAP), Seventh Day Adventist Church (SDA) and many more were

also involved in selecting faithful students into their respective primary schools, secondary schools and colleges as benefits of proprietors (ACEM, 2007).

Sandikonda (2013) revealed that over 80% of schools in Malawi were faith mission schools such that a number of students are admitted into schools by faith school governing bodies. Almost all faith based mission organizations, therefore, use selection test as means for admitting/enrolling students into their faith mission schools. "Church-*commitment criteria*" which are about church membership activeness in the churches is used as an "inclusion and exclusion criteria" to be selected to their schools, (Sandikonda, 2013; CTB & Ireland, 2015).

Therefore, the faith mission governing bodies (churches) have great impact toward enrolment of students in secondary schools in Malawi hence their selection tests they use. Although Psychometrians are not involved in the development of these tests, there is need for the tests to be of high quality if they are to be fair to the examinees.

1.2.4 Anglican Diocese of Upper shire

There are so many churches and faith organizations in Malawi that act as government partners in education. One of them is Anglican Diocese of Upper Shire (ADUS).

Anglican Diocese of Upper Shire is a forth Anglican church mission established in Malawi on 3rd May 2002. It is located to the southern region of Malawi with its headquarters (cathedral) at Mpondas in Mangochi. ADUS had over 44 primary schools, 12 secondary schools and 4 tertiary institutions in south east education division (SEED). The mission

covers over nine districts in Malawi including Mangochi, Machinga, Balaka, Zomba, part of Mwanza, Ntcheu, Dedza, Chiladzuru and Neno.

ADUS faith mission allocation from government was 40% boys and 35% girls annually per its faith secondary school by Malawi government. For colleges, enrolment space was 20% girls and 15% boys per college per annum (ACEM & government 2007 MOU; Tengatenga, 2006; Tengatenga, 2010; Litereko, 2017; Jailos, 2017).

1.2.5 Selection test and quality

"Selection test" (aptitude test) has several definitions and one of them is that it is a high stakes tests which is commonly used as a placement tool for students in schools worldwide. Psychometricians argue that a good quality selection test items should attempt to indicate what a candidate could learn if opportunity and motivation are present. Furthermore, it should have consistent and acceptable statistics or indices such that candidates with more desired characteristics and less or no characteristics are identified and should attempt to provide objective data that are better, more defensible and generate fair judgment by authorities (Hopkins, 1998; Nitko 1983; Chakwera, 2002). All procedures for test development should not be ignored for robust psychometrical indices as pre-condition for reliability and validity of any selection test. Selection test needs to be of high quality (more reliable) than "classroom test" since results from this selection test normally stand alone when making judgment and interpretation whereas "classroom scores" are aggregated into a composite "score", where judgment is based on aggregate of the outcome (Hopkins, 1998; Mazzeo, Schmitt & Bleistein, 1993).

1.2.6 Achievement levels between faith mission selected and government selected students

Safuli (1996) and World Bank (2004) reports indicated that faith mission (Grant aided) secondary schools students generally perform much far better academically than Community Day Secondary School (CDSS) students at Malawi School Certificate of Education (MSCE)in Malawi. Among so many elements the World Bank (2004) cited good resources, qualified staff and environment as contributing factors for better performance of faith schools over CDSS. For instance, the table 1shows how those two types of schools performed at MSCE from 1994 to 2000 which indicated that faith mission schools (grant aided) performed above national average and better than CDSS in all the years.

Table 1: Showing pass rates of Grant Aided and CDSS schools in 7 years

Year	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**Average**
National Pass rate at MSCE (%)	48	32	33	23	16	14	20	25.29
Grant aided pass rate (%)	65+	55+	52+	36	32	30	33	43.29+
CDSS Pass rate (%)	19	37	12	8	5	4	9	13.43

Source: World Bank, 2004, page, 50

The same report revealed that most girls achieve less academically compared to boys where it cited the reason among others being the "selection policy" which deliberately enroll (choose) a girl when a boy and a girl have achieved equally at PSLCE in Malawi. The policy allowed girls to enter secondary schools more easily than boys for the sake of increasing equity in schools.

Malambo (2012) found that secondary grant aided secondary schools performed much better (over 90%) than non-grant Aided schools in western province of Zambia as he was

trying to find factors that affect performance of pupils in those schools. He cited reasons like grant-aided secondary schools had appropriate and suitable teaching and learning materials, frequently homework, schools INSET activities for regular visits by supervisors, lower enrolment levels low absenteeism, high discipline, clear academic policies and effective school administration. Another study by Sandikonda (2013), found that grant aided schools perform generally better academically than CDSS and also convention boarding secondary schools.

Literature so far reviewed in my study indicated that there was no study about performance between government selected students (GSS) and faith mission selected students (FMSS) within same faith mission schools (intra- school study).There had been so many studies on academic performance related work worldwide and Malawi in particular. A Very close study on intra-school on performance was done by Manjombe (2018) who tried to equate scores of standard test (ST) for regular candidates and modified test (MT) scores for Total Visual Impaired (TVI)candidates in Geography at MSCE examinations which MANEB does. Manjombe (2018) found that ST was not parallel examinations to MT and that it was unfair to treat scores and examinees from those two different tests equally and proposed introducing equating system at MANEB for fair treatment.

This could be similar story to what happens in faith mission secondary schools, in the sense that two different examinations were used to select students into the same faith mission secondary schools and expecting them to perform in the same way which could be a farfetched dream to achieve. Further scrutiny and experience on academic performance

of students between FMSS and GSS within the faith mission schools was very different. The general performance was that GSS perform better than FMSS in Malawi and there was very little literature to shade more light on the same as to why that had been the case.

Report by Jailos (2017) during Diocesan Synod of Upper Shire, indicated that the general performance of students enrolled by faith mission schools in the diocese of Upper Shire in all secondary schools (over 12 faith schools) was lower at MSCE compared to those selected by government. The report indicated statistics on how some schools performed within the ADUS diocese per student type as given in table 2which were labelled P, Q and R (pseudo names) as examples:

Table 2: Showing performance of Government Selected students and Faith Mission Selected students at MSCE in 2016 in three different faith mission schools of ADUS

Name of school	Type of students	Sat MSCE	Passed MSCE	Pass rate (%)
P	GMSS	35	27	77
	FMSS	17	8	47
Q	GMSS	30	21	67
	FMSS	16	2	13
R	GMSS	61	61	100
	FMSS	32	26	81

Source: Researcher

Both faith missions selected and government selected students start form one at the same time and treated with matched conditions academically. During the study, information on EMIS data analysis for MSCE academic results indicated that there was no any moment when best performing students came from FMSS type of students in all the schools since selection test was introduced in the ADUS Diocese 2002. That information gave more questions than answers knowing that both types of students were treated in the same way

and the only difference was how they were selected. In others words, students were selected by two different types of selection tests government and mission selection tests.

The challenge was that there was very little information known about the quality of faith mission selection tests administered by different faith missions in Malawi. The study found a knowledge gap in the literature about the quality of selection tests which ADUS was using and extent of error the tests were making in selection process in relation to the performance of FMSS. Therefore, the study wanted to establish to what extent do these faith selection tests function in differentiating examinees with high ability from those with low ability and how much reliable were those tests as precondition for validity for any quality selection test.

There were so many studies on performance related to gender, school type, school location or mode of learning and so on. For instance, studies by World Bank (2004) on gender which found that boys perform better than girls. Sandikonda (2013) found that grant aided (faith mission) schools perform better than CDSS. Most of these studies generally concentrated on the aggregate performance of the whole school without an analysis on each of the two types of students who were selected by two different selection tests in the case of grant aided (faith mission) schools which show that their performance is quite different.
Therefore this research wanted to examine the quality of selection test which ADUS used to enroll students in terms of test parameters and estimate measurement error the test make in selection process.

1.3 Statement of the Problem

For more than two decades, faith mission institutions (churches) have been enrolling students into their faith secondary schools in Malawi using locally made tests (teacher made tests). On other hand, government also has been selecting students to the same faith mission secondary schools to start form one at the same. There had been two types (categories) of students in faith mission secondary schools: students selected by government (GSS) and students selected by mission schools (FMSS) every year. These grant aided schools (faith mission) have almost everything it takes for students to promote and achieve high standards of education for both types of students (World Bank, 2004). The experience showed that there is disparity in performance between FMSS and GSS despite being treated equally academically in those schools. Jailos (2017)'s report indicated there was less academic achievement of FMSS compared to GSS in all ADUS secondary schools. The report contained such MSCE statistical information of academic performance results of three sampled secondary schools labelled P, Q and R(pseudo names) as shown in table 3of two categories of candidates(FMSS and GSS) in Diocese of Upper Shire.

Table 3: Showing performance of Government Selected students (GSS) and Faith Mission Selected students (FMSS) at MSCE in 2016 in three different faith mission schools of ADUS.

Name of school	Type of students	Sat MSCE	Passed MSCE	Pass rate (%)
P	GMSS	35	27	77
	FMSS	17	8	47
Q	GMSS	30	21	67
	FMSS	16	2	13
R	GMSS	61	61	100
	FMSS	32	26	81

Source: Researcher

Furthermore, the EMIS data summarized and analyzed during the study reviewed on MSCE academic results indicated that there was no any moment when best performing students came from FMSS type of students in all the schools across years since selection test was introduced in the ADUS Diocese in 2002. The only notable difference between the two types of students was the selection tests they were used to enroll them in schools. Since the students were selected using two different selection tests; mission and government. The study conducted by Kadzitche (2018) in Malawi around Zomba district primary schools claimed that most of 'teacher –made tests' were flawed, not reliable and valid and faith mission selection tests are more less 'teacher made tests' in the way they were constructed and administered (Kadzitche, 2018, P.5). Literature in Malawi and beyond says little about the quality of 'local teacher made- tests' (faith selection tests) in Malawi (Kadzitche, 2018).The assertion prompted the study to establish disparity in performance of the two categories of students in faith mission secondary schools. Kadzitche (2018) did much on reliability and validity on primary school 'teacher made-

tests' as test quality parameters but this study concentrated on other three parameters namely: item-difficulty, item-discrimination and test information (reliability).

The study wanted to find out whether selection tests contribute to low performance of FMSS in faith secondary schools as alluded to by Kadzitche (2018) study findings. In other words, the study wanted to establish to what extent was the quality of faith mission selection tools had. Psychometrically, the research study wanted to establish how much error did those selection tests make when selecting students into their faith secondary schools in Malawi by focusing on three test quality parameters? The ADUS selection test was used as a case study due to convenience, time limit and geographical location.

1.4 Purpose of study

The purpose of this study was to analyze quality of selection tests and test items administered by faith mission secondary schools during selection and enrolment of form 1 students into their faith secondary schools in terms of item difficulty, item discrimination and test reliability.

1.5 Main research question

The researcher aimed at understanding what quality properties of selection tests ADUS selection tests had, in terms of level of item-difficult, item-discrimination and test reliability?

1.6 Specific research questions

To address that main research question, the following specific questions guided the study:

1. What are the item difficulty parameters for the ADUS selection test items?
2. What are the discrimination indices for the ADUS selection test items?
3. What is the reliability of the ADUS selection test?

A research study was conducted to focus on the range of evidence available so that an important contribution was made towards a profound understanding of the phenomena under consideration.

1.7 Significance of study

The study will inform the ADUS about the caliber of staff it had in the education department and realize potential areas where training or orientation might be needed for profession development for quality service delivery in item development. The study would inform ADUS on the need of developing a table of specifications as an important tool that guides and ensures development of reliable and valid selection tests. The study would also inform ADUS education department on the importance of conducting item analysis as a way of increasing the reliability of the test.

The study would also inform the Anglican Dioceses of Upper Shire authority on possible review of its selection policy to ensure that only deserving standard 8 examinees are enrolled at transparent way. The ADUS authority would also benefit in identifying areas where examination developers require in- service raining regarding quality and credibility

of selection tests. The knowledge about the quality of ADUS selection tests may influence policy making in the organization pertaining to the administration and selection of examinees in the education department improve credibility, effectiveness and trustworthy of the selection processes to gain public trust and confidence among the faithful.

1.8 Limitations and attempts to minimize

This study was conducted in one of the four Anglican dioceses in Malawi which represented 25% of the study area due to accessibility and financial limitation. This could be regarded as small area. Samples of examinees were drawn from primary schools that offer Bible Knowledge only leaving out those that offer Religious and moral Education (RME) due to the fact that the ADUS selection tests had Bible Knowledge items part and not RME. One selection test (ADUS -2015 Test) was administered in the study instead of two or more ADUS selection tests due to time and financial limitation.

But all these limitations were taken care of by the use of systematic random sampling in getting representative sample. The study employed primary school experts/professional practitioners during admission of test, scoring and interpretation of test scores (results) of examinees. The researcher analyzed more (three) main parameters that account for more test qualities in both CTT and IRT theoretical framework. The study drew a very large sample of examinees to increase viability and consistence of test parameters (Lord, 1980).The efforts used in testing process of the examinees increased reliability and validity of the study and maintained rigour of research study.

1.9 Operational Definition of Terms

a-parameters: statistic measure that gives an ability level of an item to discriminate students with high ability levels from those with lower ability levels.

b-parameters: statistic measure that gives a level of difficult of an item.

Difficult level: the percentage of examinees that answered the item correctly (Boupathiraj, & Chellamani, 2013, p.190).

Discrimination power: the ability of an item to differentiate among examinees based on how well they know the material being tested.

Education zone: a group of schools at a particular place under one primary Education Advisor (PEA).

EMIS: files in which government keeps education information in Malawi.

Examinees: candidates who sit for a particular selection test.

Experts: are primary teachers who are well knowledgeable about a particular subject matter.

Faith Mission secondary schools or (grant aided): Secondary schools which are owned by faith mission organization and receive grants from government of Malawi.

Faith mission selected students (FMSS): students selected by faith mission secondary school.

Government selected students (GSS): students selected by government of Malawi.

Grade 8 examinees: candidates who are about to sit for standard eight at primary school in Malawi.

Scores: Marks obtained by examinees on a particular selection test.

Higher order abilities: the upper more complex categories of cognitive activity that covers application, analysis, and synthesis in the Bloom's taxonomy. (Mbunge, 1986, p.12).

Item analysis: a process which evaluates responses of students to individual test items in order to assess their quality and the quality of the test as a whole.

Lower order abilities: refers to knowledge and comprehension in the Bloom's taxonomy. (Mbunge, 1986, p.12).

Test: an instrument used to judge achievement among examinees (Gronlund, 1993; Nitko, 1996).

Test items: questions that make up an examination.

Test reliability: refers to the internal consistency of the scores whereby those who have performed better on the test also perform better on individual items.

Validity: the degree to which the item/test measures what intends to measure and abilities that a course of instruction has aimed to teach. (Ebel, 1979.).

CHAPTER 2

LITERATURE REVIEW

2.1 Chapter overview

This chapter provides definitions of a test, description of standard test, selection test, and types of selection tests, quality of a test and statistical quality of a test. Furthermore, outlines views about test as a selection tool, ways of examining quality of test and test and ends up with theories in the field of testing measurement and evaluation

2.2 Definition of a test

Literature shows that there are several definitions of a test. Some are according to context while others are according to purpose or type. Linn & Gronlund (2000) define a test as an instrument or a systematic procedure for measuring a sample of behavior by posing a set of questions in uniform manner. According to Nitko (1996), a test is an instrument or systematic procedure for observing one or more characteristics of a student using either a numerical scale or a classification scheme. Robertson & Mike (1986) view a test as a procedure for critical evaluation; a means for determining the presence, quality, or truth of something and a trial. According to "free dictionary online", a test is a basis for evaluation or judgment. Hughes (2005) defines a test as a carefully chosen, systematic and standardized procedure for evolving a sample of responses from candidates which can be used to assess one or more of their psychological characteristics with those of a

representative sample of an appropriate population. Therefore, broadly speaking, a ‘test’ is a standard procedure for obtaining a sample of behaviours for a specific domain.

2.3 A standard test

A test needs to be standardized as one of the very important characteristics of a good test. Several scholars have tried to describe the term “standardized test” as follows: Kaira (2002) refers Standardized test any form of a test that requires all test takers to answer the same questions or a selection of questions from common bank of questions. She further added that it is scored in standard or constant manner which makes it possible to compare the relative performance of individuals. According to Weaver (2011), standardized test is a test designed by people with specialized knowledge and trained in the test construction, test takers respond to the same items under the same conditions, answers from respondents are evaluated according to the same scoring standards and the scores are interpreted through comparison to the same scores obtained from the group that took the same test under the same conditions. Therefore, the researcher, in this context, defines a standard test as a test developed by people with appropriate expertise and is administered with matched conditions across the examinees.

2.4 Selection test

This document defines a selection test according to Robertson & Mike (1986), as a carefully chosen, systematic and standardized procedure for evolving a sample of responses from candidates which can be used to assess one or more of their psychological characteristics with those of a representative sample of an appropriate population.

Furthermore, selection tests are measuring instruments that are often referred to as psychometric tests. The purpose of a selection test is to provide an objective means of measuring individual abilities or characteristics. They are used to enable selectors to gain a greater understanding of individuals so that they can predict the extent to which they will be successful in a job or academic performance. Robertson & Mike (1986) say the selection tests are used to provide more valid and reliable evidence of levels of intelligence, personality characteristics, abilities, aptitudes and attainments than can be obtained from an interview.

2.5 Types of selection test

Selection tests can be categorized into different types depending on the nature, use and purpose. Norman et al. (1987) and Saville & Sik, (1992) gave main types of selection test as intelligence, personality, ability, aptitude and attainment tests. A distinction can be made between psychometric tests and psychometric questionnaires. As explained by Norman et al (1987), a psychometric test such as one on mental ability has correct answers so that the higher the score, the better the performance. Psychometric questionnaires such as personality tests assess habitual performance. Such selection tests measure personality characteristics, interests, values or behavior. With questionnaires, a high or low score portrays the extent to which a person has a certain quality and the appropriateness of the replies depending on the particular qualities required in the job to be filled. For general selection purposes, an intelligence test that can be administered to a group of examinees is the best, especially if it has been properly validated, and it is possible to relate test scores to ‘norms’ in such a way as to show how the individual taking the test compares

with the rest of the population, in general or in a specific area (Ward, 1982).

For some types of tests out lined in this document, there are comments and description about them. For instance, personality tests (questionnaires) were shown to have the low validity coefficient of 0.15 on the basis of research conducted by Schmitt, Gooding, Noe & Kirsch (1984). But as Barrett, Kline, Paltiel & Eysenck (1996) point out, Schmitt et al. (1984) used inappropriate tests, many developed for clinical use and some using 'projective' techniques such as the Rorschach inkblots test, the interpretation of which relies on a clinician's judgment and is, therefore, quite out of place in a modern selection procedure. Smith's (1988) studies, based on modern self-report questionnaires, revealed an average validity coefficient of 0.39, which is reasonably high.

The persistent attack was launched on use of personality tests by Blinkhorn & Johnson (1990), where commented: 'We see precious little evidence of personality tests predicting job performance' (p. 465). In contrary view, Fletcher (1991) backed up the idea that: "Like any other selection procedure, they (psychometric tests) can be used effectively or badly. But it would be not good at all to dismiss the evidence of the value of personality assessment in selection on the basis of some misuse. Certainly the majority of applied psychologists feel the balance of the evidence supports the use of personality inventories" (p.38). Personality tests can provide interesting supplementary information about candidates that is free from the biased reactions that frequently occur in face-to-face interviews. But they have to be used with great care.

Ability test measures job-related characteristics such as number, verbal, perceptual or mechanical ability of an individual pertaining to his or her future performance. Aptitude test is job-specific test that is designed to predict the potential an individual has to perform tasks within a job or academic performance. They can cover such areas as clerical aptitude, numerical aptitude, mechanical aptitude and dexterity. Aptitude tests should be properly validated. The common procedure is to determine the aptitudes required by means of job and skills analysis. A standardized test or a test battery is then obtained from a test agency.

Attainment tests measure abilities or skills that have already been acquired by training or experience.

2.6 Quality of a test

Quality of a test is described by a number of different elements or characteristics. Across literature, there are diverse perceptions or views on what makes a test item high quality or not that constitute high quality test (Schuwirth, Bosman, Henning, Rinkel & Wenink, 2010; Kline, 2015). This shows that there is no single definition about 'high quality test' in education, since literature does not provide clear-cut answers to questions concerning quality of tests. The role of this document is, therefore, to provide a framework for quality of a test in terms of characteristics and statistical values. High quality test is characterized by high – quality items that have minimal false –positive, that means candidates can answer the item correctly without having the necessary knowledge (abilities)characteristics or false

– negative responses that means candidates answer the item incorrectly despite having sufficient relevant knowledge or competence (abilities).

A high-quality item is more than an item that just does not have any violations against agreed-upon item construction rules; the item must also be creative, relevant for the discipline and with appropriate difficulty level. It is clear that these are expert judgments and therefore require communication and agreement between partners. High quality item constitute high quality test.

AMAC (2014) gave five characteristics of a quality test in terms of the following: it should show indicators for knowledge/ability, creativity, relevance, format versus content and difficulty. Among so many scholars, Robertson & Mike (1986) cited other characteristics of good selection test; it should be a sensitive measuring instrument that discriminates well between subjects. It has to be standardized on a representative and sizeable sample of the population for which it is intended so that any individual's score can be interpreted in relation to that of others. It should be reliable in the sense that it always measures the same thing. A test should measure specific abilities; test aimed at measuring a particular characteristic, such as intelligence, should measure the same characteristic when applied to different people at the same or a different time, or to the same person at different times. It should be valid in the sense that it measures the characteristic that the test is intended to measure. Thus, an intelligence test should measure intelligence (however defined) and not simply verbal facility. A test meant to predict success in a job or in passing examinations should produce reasonably convincing (statistically significant) predictions.

Reasoning behind, is that what it the selection test measures?

Validity can be expressed as a coefficient of correlation in which 1.0 would be equal to perfect correlation between test results and subsequent behavior, while 0.0 would-be equal to no relationship between the test and performance. The following rule of thumb guide on whether a validity coefficient is big enough was produced by Smith & Smith (2004) as follows: over 0.5 is excellent, 0.40-0.49 is good, 0.30-0.39 is acceptable, and less than 0.30 is poor. On this basis, only ability tests, bio data and according to Saville & Sik's (1992) personality questionnaires reach acceptable levels of validity.

2.7 Statistical quality of a test

Quality of a selection test can be described by a number of statistical variables or parameters. Some of the factors are conditions while others are statistical indices which can be calculated, analyzed and interpreted. Some of the test parameters which can give or describe the level of quality of a test that can be statistically calculated are item-difficult, item-discrimination, test-reliability, differential item function (DIF) and validity while those which are conditions are test fairness, use of test and many more.

A good quality selection test should have acceptable statistical indices found and set by research studies using recommended procedures and practices. Therefore, a test is valid only if it is rightly used and its scores are well interpreted using recommended theoretical or conceptual frameworks. Razak, Khairanib & Thien (2012) found that a test is regarded as a good quality test for use, if it satisfies the following characteristics in addition to its statistical indices;

(i) Test items must be fair, valid and reliable in order to create fair, valid and reliable tests;

(ii) A test is only as good as each item on it. If items don't really measure the standard, the test results will not be useful. This means that every item of the test must be valid, reliable and dependable on its own.

Razaket al. (2012) concurred with Messick (1989) that selection test should have reliable statistical values ifit to be fair specifically criterion validity since this validity deals with how instrument (test) relates to some external criterion of which the instrument is expected to give information for inferences.US DLETA (2000) research findings indicated that only valid and reliable selection tests do more benefit than harm. Therefore, a test must measure what it intends to and should produce consistent measure (validity and reliable) before use. This implies that a selection test should be only administered to candidates /examinees if and only if it has reliable and convincing invariant parameters that will give equal chances to spell out true differences in ability of examinee (Kline, 2015; Robertson & Mike, 1986). In short, a selection test must be "standard test" with acceptable conditions and statistical values in question.

2.7.1 Item-difficulty and quality of test

It is desirable to have test items which vary in their item-difficult (b-parameters in IRT and P-values in CTT), so that all points of the ability stratum may be fully tested. However, according to McAlpine (2002) research findings, it is undesirable to have facility values (p-value) of a test above 0.85 or below 0.15. That meant that a good quality test should have items with item-difficulty within the range of 0.15 to 0.85.

Items with p-values below or above the range should not be included since they are regarded as too difficult or too easy items respectively for the examinees to get the correct responses hence, compromise the quality of the test and produce low reliability and erroneous predictive validity of the test.

Research by Adedoyin & Mokobi (2013),which examined three statistical parameters that constitute quality of test items; item difficulty, item discrimination and pseudo guessing parameters using 3 parameter model of IRT framework on Botswana national examinations found that over 97% of the total items included in the Junior Certificate Mathematics National Examination of 2010 were poor items. The research established that item -difficult indices (**b**-parameters)within the range -0.5 to +0.5 were items with fair or medium difficult levels, Items with b-parameters below -1 were labelled easy items (poor) and those with b-parameters greater than +1 were regarded as very hard items(poor) to be answered correctly by examinees. This knowledge impacted the examination boards to verify the items before they were included in the test or national examinations in Botswana. The researchers concluded that items with poor p-values or b-parameters should be improved or modified before they are included in the test since they compromise the quality and function of the test. It was, therefore, recommended that examination bodies should consider improving the quality of their test items by conducting IRT psychometric analysis for validation purposes on test items before use.

2.7.2 Item discrimination and test quality

Quality of a test can be described in terms of item discrimination (**a-**parameter) in IRT theoretical framework. Item discrimination is the ability of an item to separate examinees with high ability from those with less ability in locating correct responses of the items. In other words, we expect that an item that demands more ability should be answered correctly by examinees who have ability equal or more than the required ability and not with examinees with less ability or no ability.

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

On the other hand, CTT can find the item discrimination indices using person product-moment correlation (r) between the items and the total test score (McAlpine, 2002). Correlation (r)ranges from +1 (where there is perfect relationship, examinees that score high marks on the item are the same examinees that also have scored high marks on the same test) to -1 (where there is perfect inverse relationship between those scoring high marks on the item and on the test.

2.7.3 Reliability and test quality

US DLETA (2000) revealed that validity and reliability were the main two technical properties of a test that indicated the quality and usefulness of the test and these must be examined when evaluating the suitability of any test use. Most scholars including Tavakol & Dennick (2011) agree that if the test was not valid and reliable, then must not be administered or its findings (scores) be used to make any judgment or decisions by authority. That meant that for any test, validation or standardization was a "must do" procedure before administration and use of its scores for sound decisions (Stone, 1992).

Additionally, in the USDLETA (2000) research study conducted in United States of America that involved examining the correlation between job performance and performance during selection aptitude (test), the upper limit of the test statistical indices range in reliability was set and interpretation of statistical values were reviewed as follows: reliability coefficient (r) of a good test should range from 0.21 to 3.5 since reliability coefficient r rarely exceeds 4 .Those statistical values meant that the larger the value, the more the reliability of the test and the lower the statistical value, the less the reliability resulting into poor validity and consequently poor quality test.

McAlpine (2002) and US DLETA (2000) agreed across their work that as the statistical values of the reliability improve (increase) for the better, that results into high quality test items or test. Furthermore, both writers through their findings showed that reliability covers a range with end limits.

Tavakol &Dennick (2011), agreed with many researchers like McAlpine (2002) and across literature that reliability could be analyzed and interpreted using Cronbach Alpha statistical indices on a test using Likert scale when examining quality of a test. The researchers further established a lower bound of 0.65 and an upper bound of 1.0 where the values of reliability could be accepted and regarded the test being reliable for use and that large statistical values of Cronbach Alpha were more recommended than smaller ones. Cronbach‘s Alpha is the measure of reliability or internal consistency of test items or set of scale. In short, it measures the strength of consistency. It is a function of the number of items in a test, the variance of the total score and the average of covariance between pairs of items. The results of Cronbach Alpha range from 0 to 1 as overall assessment of a measure’s reliability. If items are completely independent from each other (uncorrelated or share no covariances) then alpha is zero but if items are sharing more covariances and not independent then alpha value approaches 1 as the number of items increases to infinity.

Most researchers, including Tavakol & Dennick (2011), recommend a minimum reliability alpha (a) of 0.65 and a maximum of 0.8 or higher in many cases to rate items to be good quality and fit for test development. Through the same study, Tavakol & Dennick devised six types of tests based on Cronbach Alpha levels as follows:

Table 4: **Showing Cronbach Alpha (**a) **Likert Scale**

Cronbach's Alpha indices	Internal consistency (test quality)
a ≥ 0.9	Excellent test
0.9 >a ≥ 0.8	Good test
0.8 >a ≥ 0.7	Acceptable test
0.7 > a ≥ 0.6	Questionable test
0.6 >a≥ 0.5	Poor test
0.5 >a	Unacceptable test

Source: Mohsen Tavakol & Reg Dennick, 2011

2.8 Views about test as selection tool

Coe et al. (2008) agree with Makori, et al. (2015) that the use of tests as tools for selecting or placing students in schools is not trustworthy and effective because selection tests fail to identify students with more abilities because selection tests were not and will never be fair or adequate to examinees. Furthermore, they argued that learner's ability is multi-dimensional and fluid. Selection limits parental choice such that use of selection test has an adverse effect on learners' performance, which may result into producing worse academic results. A related research conducted in Kenya established that over 70 percent of faith missions and other stakeholders did not trust standard eight nation selection examinations citing reasons ranging from lack of transparency during selection and nation examination failing to pick capable students (Makori et al., 2015; Priscilla, 2011; Burrow,2015; Jacob, Jepkenei, Chepwarwa & Makori , 2015).

Therefore, it is equally important to assess the selection test thoroughly before they are used to assess the learners for selection purposes so that psychometric properties (quality) are checked to establish test reliability, validity and other parameters that count test quality (Souza, Costa, & Guirardello ,2017).

But Test Partnership Limited (2017) differed to Coe et al. (2008) where they recommended the use of selection test when placing students at different levels in education, citing reasons that aptitude (selection) tests were among the most commonly-used assessment tools for predicting academic performance of learners in schools. Kinyua (2014) recommended the selection test as an easy way to obtain learners with high ability from those with low ability to enrolled or award them correctly in schools. He cites that as the reason why tests are used worldwide as selection tools or mechanism for students in schools and employment in workplaces. Another research by Priscilla (2011), in Kenya, did agree with Kinyua (2014) and Test Partnership (2017) that selection test is the best criterion for predicting the future academic performance of students in schools and that use of tests in selection of students into form one is inevitable in the world including Kenya. She gave an example of KCPE which was used to select form one students to various secondary schools in Kenya which operated on premise that their sterling performance of students at KCPE would enable them to perform well at the Kenya Certificate of Secondary Education (KCSE) which would come at the end of secondary education (four years later).The assumption under the use of selection test was that examinees that perform well at selection test would also perform well at year four examinations- that was predicting future performance of learners in Kenya.

A related research by Makori et al. (2015) concurred with Priscilla's (2011) assumption but further advised that, it could be possible for tests to achieve that, only if, the selection tests used were valid and reliable.

Literature revealed that there had been so many selection test boards in the world, both locally and internationally which started long time ago preparing "standardized selection tests", administering, scoring, interpreting and making judgment based on the results (Young & Fraser, 1994). For instance, America (USA) had been making decisions based on a number of selection tests like SAT(Stanford Achievement Test), American College Test(ACT),Graduate Record Examinations (GRE), Iowa Test of Basic Skills (ITBS), terra Nova , Metropolitan achievement tests for selecting students in united states of America schools which were devised sometime back around 1965. In Africa, we could mention Kenya Certificate of Primary Education (KCPE) and Primary Leaving Certificate of Education (PLCE) in Malawi for selecting grade 8 students to join secondary education in form one (Sandikonda, 2013; Priscilla, 2011; Nichols& Berliner, 2007; Wright & Stone, 1979).West & Hind (2016) recommended that admission arrangements of students into secondary schools through selection tests should be always clear, fair, consistent and objective so that the intake for every cohort should have similar characteristic ability with the rest.

2.9 Examining quality of test and test items

There are so many ways of examining (analyzing) the quality of selection test and test items properties (parameters) statistically. Some of the parameters (indices) that describe the quality of a test which can be considered are item-difficult, item-discrimination, differential item functioning (DIF) and test-reliability. These statistics can be checked in IRT psychometrical analysis using test score properties and gauge the level of quality of a test used to generate that data (scores) (Zumbo, 1999).

Carlson & Davier (2013) and McAlpine (2002) define item analysis as a method of gauging the quality of an examination (test) by looking at its constituent parts (items). It seeks to give some ideas of how well the examination (test) has performed relative to its purposes. The primary purpose of item analysis in most high education institutions is that of a measurement tool for assessing the achievements of the examinees (candidates) and thus how future learning will be supported and directed.

2.10 Theories in the Field of Testing, Measurement and Evaluation

There are two main theories in the field of Testing, Measurement and Evaluation, namely Classical Test Theory (CTT) and Item Response Theory (IRT).

2.10.1 Classical Test Theory (CTT)

It is mostly used in Britain. Its underlying assumptions come from psychology and were developed around the turn of 20th century. It is used to handle a range of types of questions including optional ones (Carlson & Davier, 2013).CTT is a theory about test scores which

introduces three main concepts: test scores which are also called observed scores; true scores; and error scores, Kadzitche (2018). This aimed at coming up with error-free test items which, in turn, makes the test to be valid and reliable. Classical measurement models and methods used in CTT still remain viable in many testing programs despite some literature recommending IRT in modern world testing (Hambleton & Jones, 1993). There are three basic assumptions in CTT: there is no correlation between true scores and error scores, the average error score in the group of examinees is zero, and also there is no correlation between error scores on parallel form tests (Hambleton & Jones, 1993). CTT Charles was found by Spearman in 1904 (Traub, 1997).spearman discovered three things which were a major breakthrough in a field of testing and measurement and those were: presence of errors in measurements, conception of error as a random variable and, correction of correlation coefficient for attenuation (Traub, 1997). In the field of CTT item analysis there are two statistics: the difficulty index (p-value) and the discrimination parameter (power) which is the item-total correlation or point biserial correlation, (Hambleton & Jones, 1993). When test development techniques are applied, besides concerns regarding content validity, selection of items is based on item difficulty and item discrimination (Hambleton & Jones, 1993).CTT was used in this study during the analysis of content validity and reliability of the ADUS selection tests. During content validity assessment of the test (domain sampling), primary school subject matter experts were involved to verify each test item the appropriate content core element of the syllabus as well as thinking level.

2.10.1.1 Item-difficult

Item-difficult is essentially a measure of difficulty of an item with a high facility (p-value) indicating an easy item and a low facility indicating a difficulty item. This is given by the formula. The formula below can be used to determine the item- difficult index in a test as was advised by McAlpine (2002).

$$\text{Fac(x)} = \frac{\overline{X}}{\text{Xmax}}$$ McAlpine (2002) Model

(x) = the facility value of question x

X Bar = the mean mark obtained by all candidates attempting question x

Xmax = the maximum mark available on the question

Obinne (2011)) found that it was desirable for the facility value (difficulty index) of a test item to be close to 0.5 to promote maximal differentiation. But importantly a test should have questions which vary in their difficulty, so that all points of the ability stratum may be fully tested. However, it is undesirable to have facility value above 0.85 or below 0.15 (Adedoyin & Mokobi, 2013).

2.10.1.2 Item- discrimination

Item discrimination is the measure of how the candidates perform on a question as opposed to another measure of performance. It can also be calculated by Pearson product-moment correlation between the items and the total test score. Correlation ranges from +1 (where there is perfect relationship between those who score high marks on the item and those who

have high marks on the test) to -1(where there is perfect inverse relationship between those scoring high marks on the item and on the test (Obinne, 2011).

In general, item discrimination should be always positive unless there is good reason to suppose that the assumption of undimensionality has been violated. Negative item discrimination should be regarded as suspect, and has no limit for its statistics. But Massey (1995) suggested that indices below 0.2 are weak and values above 0.4 are desirable in CTT theoretical framework. The formula for item discrimination is shown below

$$r_{xy} = \frac{\sum xy}{NSxSy}$$ McAlpine (2002) Model, p.6

where r_{xy} = the correction between the item x and the test total y

$\sum xy$ = the sum of the product of the deviations of the items and the totals

N= the number of observations

Sx = the standard deviations of the item

Sy= the standard deviation of total marks

But a key problem with CTT indexes (parameters) is that they depend on the group of examinees being tested and, therefore, do not adequately reflect the measurement quality of the test items and a test as a whole. CTT hinges on assumption that every individual or person has a true score, T, and this true score can be obtained if and only if traits are constant and there are no random errors which can affect the result (Yu, 2008; Carlson & Davier, 2013) i.e.

X = T +E

X = the total score/observed score obtained

T = the true score and

E = the error component.

The problems that occur with CTT analysis of the examinees' proficiency and quality of test items are successfully addressed in the framework of item response theory (IRT).

2.10.1.3 Reliability

It's the overall consistency of a measure. A measure (test) is said to have high reliability if it produces similar results under consistent conditions (Kline, **2005)**.

There are three ways of estimating reliability of a test with CTT. These are test-retest where a set of paired scores are analyzed, parallel forms reliability where scores from two tests are correlated and split-half (single test administration). The first two are expensive and time consuming methods according to Massey (1995) and Wilmut, Wood, & Murphy (1996).Using these three ways, reliability of a test in CTT can be estimated by finding the internal consistency (correlation between the items) of the test to be developed. Two tests can also be correlated using same methods. It is possible to find correlation between an already constructed test and all other possible tests which might be constructed from the hypothetical universe of questions measuring that measure the same trait Massey (1995). For multiple choice, an internal consistency measure of over 0.90 is achievable and desirable, for short answer questions, measure in the range of 0.65 to 0.80 are expected while for long essay type examinations and more practical examinations, reliability may be as low as 0.40 without concern being raised (Moss, 1994).

Reliability of a test can generally be estimated by two formulas when estimating (calculating internal consistency) coefficients. These are Cronbanch's alpha which is generalized formula of the Kudor –Richardson 20 formula and a Backhouse's P, a specific form of alpha coefficient designed to copy with optional questions (Backhouse, 1972).

Cronbach's Alpha (for a test with compulsory questions) is

Cronbach's Alpha for compulsory questions

$$r_\alpha = \frac{k}{k-1}\left(\frac{\sum S^2}{1-\sum S^2}\right)$$

And for a test with optional questions the appropriate equation used is:

Backhouse's P
(for a test with optional questions)

$$r_\alpha = (\lambda + 1)\left(\frac{\sum_{j=1}^{k} n_j S_j^2}{nS_x^2}\right) - \lambda\left(\frac{\sum_{j,t=1}^{k} n_{j,t} m_j . m_{t.j} . nm^2}{nS_x^2}\right)$$ McAlpine (2002)

K= number of items, n= number of people taking the test, nj = number of people attempting question j, njt = number of people attempting both questions j and t, si = standard deviation of item I, s= standard deviation of the test.

2.10.2 Item Response Theory (IRT)

Item Response Theory (IRT) field provides a framework for modeling and analyzing item response data. It is a modern method of estimating both examinees ability and item

parameters since it gives invariant properties that can be used across different groups of examinees. IRT is a powerful tool used in measurement of examinee ability, selection of test items and for equating tests. The concept of Item Characteristics Curve (ICC) is used in IRT to show the relationship between examinee ability and performance on an item. In IRT ability and item parameters are both estimated based on examinees' response patterns on the test(An & Yung, 2014).Some of the of advantages of using IRT over classical test theory for analyzing mental test data are well explained such as suitable for large data and test equating. It postulates that an examinee's performance on a test depends on a set of unobservable "latent traits" that characterize the examinee. An observed score of an examinee on an item is regressed on the latent traits. The resulting regression model, what is termed as an item response model, specifies the relationship between the item response and the latent traits, with the coefficients of the model corresponding to parameters that characterize the item. This is item level modeling that gives IRT its advantages over classical test theory. Item Response Theory is based on strong mathematical and statistical assumptions, and only when these assumptions are met, at least to a reasonable degree, can item response theory methods be implemented effectively for analyzing educational and psychological test data and for drawing inferences about properties of the tests and the performance of individuals. Checking model assumptions and assessing the model data fit are routine in statistical endeavors (Khalid, 2009).In study IRT was used to select a model (2PLM) to use, dimensionality of items, item dependence, analysis of data and interpretation of findings and choice of which software (BILOG) to use in analysis.

2.10.2.1 Item-discrimination (The a-parameter)

One characteristic of a good test item is that high-ability candidates should answer it correctly more frequently than lower-ability candidates. The a-parameter expresses how well an item can differentiate among examinees with different ability levels. The discrimination indices (a-values) of good items range between +0.5 to +2 and the steeper the slope of an item characteristic curve (ICC), the higher an item's discrimination value. High discrimination level indicates that the item discriminates well between low and high skilled individuals. The *a*-parameter is a measure that can be graphically expressed by the steepness of the ICC. If the values of the item-discrimination are above 1, they are normally desirable values for a good test items and values above 0.75 can also be acceptable sometimes (Obinne, 2011).This parameter is estimated by using the three parameter IRT logistic model (3PLM) that takes the following form:

$$P_i(\Theta) = C_i + (1 - C_i)\frac{1}{1+e^{-Daj(\Theta - bi)}} \quad \cdot \quad 3PLM$$

Adedoyin & Mokobi (2013)

where C_i is the guessing factor, $\mathbf{a_i}$ is the item discrimination parameter commonly known as item slope, bi is the item-difficulty parameter commonly known as the item location parameter, **D** is the arbitrary constant (normally D = 1.7) and θ is the ability level of a particular examinee. The item location parameter is on the same scale of ability,**θ** , and takes the value of θ at the point at which an examinee with the ability-level θ has a 0.50 probability of answering the item correctly. At the point of the location parameter, the item discrimination parameter is the slope of the tangent line of the item characteristics curve (ICC). When the guessing factor is assumed or constrained to be zero (ci =0) the

three-parameter logistic model is reduced to the two- parameter IRT logistic model (2PLM) for which only item location and item slope parameters need to be estimated(An &Yung, 2014).

2.10.2.2 Item difficulty(The b-parameter)

The difficulty of an item, known as the b-parameter, is the point where the S-shaped curve has the steepest slope. The more difficult an item is, the higher an examinee's ability must be in order to answer the item correctly. Items with high b values are hard items, that is, values of b greater than 1 indicate a very difficult item and low-ability examinees are unlikely to answer it correctly. Items with low b-values below -1 indicate easy items, which most examinees, including those with low ability, will have at least a moderate chance of answering correctly. When the values of *b* are between -0.5 to +0.5, then the test items with such difficulty indexes have medium difficulty levels.

The appropriate IRT model for this parameter to be estimated is:

$$P_i(\Theta) = \frac{1}{1+e^{-Daj(\Theta - bi)}} \qquad 2PLM$$

Adedoyin & Mokobi (2013)

If another restriction is imposed that stipulates that all items have equal and fixed discrimination, then aj becomes a constant rather than a variable. As such, this parameter does not require estimation, and the IRT model is further reduced to one PL model (1PLM) which is used to determine the guessing parameter called pseudo-guessing parameter (c-parameter).

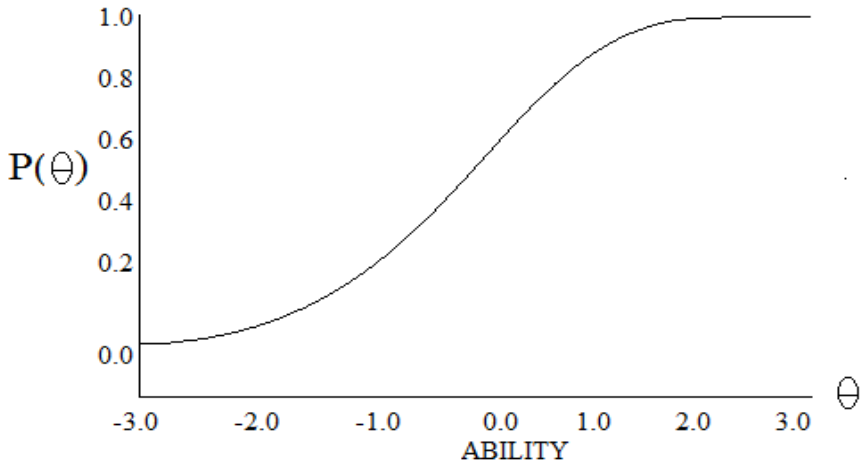
2.10.2.3 Pseudo-guessing (The c-parameter)

The **c**-parameter in IRT expresses the likelihood of an examinee with very low ability to be able to guess the correct response to an item and, therefore, has a greater-than-zero probability of answering correctly. The item guessing parameter c, is the lowest value that an ICC curve attains. For example, an examinee who randomly selects responses to items that have four response (multiple) choices can answer these items correctly about 1 out of 4 times, meaning that the probability of guessing correctly is about 0.25(Carlson & Davier, 2013). This parameter is well estimated using 1 parameter model which has the following logistic equation:

$$P_i(\Theta) = \frac{1}{1+e^{-D(\Theta - bi)}} \qquad \text{1PLM}$$

Adedoyin & Mokobi (2013)

These logistic equations when graphed produce plots that are called item characteristic curves (ICC). When ICCs are plotted the ability of the examinee is denoted by theta (θ) on the x-axis while the probability of an examinee correctly answering the question is denoted by P(θ) on the y-axis. ICCs typically take the S – shaped curve called ogive (∫)(An &Yung, 2014).



Adedoyin & Mokobi (2013)

Figure 1: Example of Item Characteristics Curve (ICC)

The probability of the correct response is closer to zero at the lowest levels of the trait and it increases to the highest levels of the traits where the probability of correct response approaches 1 (Hambleton, Swaminathan, & Rogers, 1991). To describe the ICC, two technical properties are used, the values of item difficulty and item discrimination. The value of item difficulty denoted by (b) is a location parameter, indicating the position of the item characteristics curve in relation to the ability that is required for an examinee to have a 50% chance of getting the item right. The item discrimination provides information on how well an item separates people with high and low ability levels (An & Yung, 2014).

Furthermore, IRT provides test item information that contributes the estimation of ability at any given point along the ability continuum. The information of an item is gathered from the formula:

$$I_i(\theta) = \frac{2.89a_i^2\,(1-C_i)}{\left[C_i + e^{1.7ai(\theta - bi)}\right]\left[1 + e^{-1.7ai\,(\theta - bi)}\right]^2}$$

(Adapted from McAlpine, 2002)

Where $I_i(Q)$ =item information provided by the item I at point Q on the ability scale, a_i = the discrimination of the item, b_i = the difficulty of the item, C_i = the pseudo-guessing (chance) level of the item and e = Euler's constant (2.71)

Test information which is mostly given graphically called test characteristic curve (TCC) at a given ability is the sum of the related information of items of a test. As the test information increases, the standard error of estimation (atθ) decreases. The standard error of estimation can be calculated by:

$$SE(\theta) = \frac{1}{\sqrt{\sum_{i=1}^{n} Ii(\theta)}}$$ (adapted from McAlpine, 2002, p. 22)

Where SE (θ) =the standard error of estimation (SE_{est}) at ability level θ

$\sum_{i=1}^{n} Ii(\theta)$ = the sum of the item information ability level θ for all items in the test.

IRT parameters and their models are the modern frameworks for estimating the quality of any test items and whole test as well if chosen (Embretson & Reise, 2000). Therefore, the

researcher would like to examine the quality of selection test items and tests used by faith mission secondary schools when selecting students into their respective secondary schools. ADUS selection test was used as a case study.

A number of researches on validity and reliability of tests have been conducted worldwide using IRT and generally their research findings portrayed similar results that most of the tests used in selecting students lack validity and reliability. Research by Adedoyin & Mokobi (2013) found that 2010 Botswana Mathematics JC paper 1 had only one best item out of 40 items which had accepted and perfect item parameters. Through the same research study, they recommended that examination bodies especially in Kenya and beyond should improve their quality of test items by conducting item analysis using IRT psychometric analysis for validation purposes and further said that there was a need to shift from CTT to IRT when constructing and analyzing items for public examinations in Africa.

A related research study was carried out in Malaysia (Razak et al., 2012) which analyzed the Malaysian secondary school from two mathematics item bank and found that 19 out of 160 test items (12%) did not fulfill quality requirements for Rasch Measurement Model (1PL model) and those were excluded from item bank of Malaysia from two mathematics item bank. The researchers recommended use of IRT framework analysis and advised all examination boards and item developers to use and analyze their item bank through that measurement to ensure the test fairness, appropriateness, reliability and validity to the examinees.

IRT methods have been shown, in many studies, to be more superior to other methods of data and item analysis (Ironson, 1977; Ironson & Subkoviak, 1979; Runder, Getson & Knight, 1980; Camilli & Shepard, 1994; Subkoviak, Mack, Ironson, &Craig, 1984). IRT methods worked well with large samples as estimated by Embretson & Riese (2000) that sample should not be less than 500 examinees for stable IRT item parameter estimates. Therefore, this research used ITR framework to examine and analyze the quality of test and test items of faith (church) selection test in terms of three parameters: item difficulty, item discrimination and test reliability due to large sample of examinees drawn.

CHAPTER 3

METHODOLOGY

3.1 Chapter overview

This chapter discusses study approach, research design, population, sampling procedures, data generation, instrumentation, data analysis. The chapter ends with how validity and reliability were enhanced and ethical considerations that guided the research study.

3.2 Study approach

The study followed a quantitative research tradition. Quantitative research is an approach for testing objective theories by examining the relationship among variables (Creswell, 2014, p.3). Quantitative approach was ideal for this study because data were in form of numbers (scores) from the ADUS selection test that could objectively measured and analyzed using statistical procedures. Quantitative approach is a method that makes use of positivism that beliefs that social environment is real and constant regardless of time and setting (Creswell, 1994).

3.3 Design of study

The study used a descriptive research design to examine the quality of ADUS selection test in terms of item-difficulty, item-discrimination and test reliability parameters: The descriptive research describes and interprets the current status and is concerned with conditions that exist, practices that prevail or trends that are developing (Mehraj, Taufique,

Ona, Nusrat, & Uddin, 2014).; Yildiz et al., 2017). Descriptive research design simply describe data on variables of interest and the attractions of a survey lie in its appeal to generalizability within given parameters (Cohen, Manion, & Morrison, 2007).The descriptive research design with cohort longitudinal survey type was chosen for this study because ADUS selection test was to be described in terms of its quality by collecting and analyzing data (scores) statistically of one particular group of 2018 standard 8 candidates (Kothari, 2004).

3.4 Population and Sampling procedures

3.4.1 Study population

The population of interest, in this study was the 2018 standard 8 candidates (8569) who were registered by MANEB in Boma and Chimbende education zones in Mangochi district in the southern region of the country. These were all students who completed standard 8 work in 2018 and were waiting to sit for 2018 MANEB in less than two weeks.

The research targeted 2018 standard 8 examinees from full primary schools which offered Bible Knowledge subject at standard8 because the ADUS selection test had Bible Knowledge items section and no Religion and Moral Education studies (RME) section.

3.4.2 Sampling technique and Sample size.

Sampling of schools (clusters) and examinees involved “multistage” systematic random sampling technique with the use of “inclusion and exclusion criteria”. Since some schools were offering Bible Knowledge and others Religious and Moral Education (RME)

First stage: Random sampling technique was used to draw 10out of 33primary schools from two education zones (Boma and Chimbende).

Second stage: Primary schools that offer **Bible knowledge** and not **Religious and moral education** subject at standard (grade) 8 were further selected from 10 sampled schools using the "inclusion and exclusion criteria". The sampled primary schools then dropped to nine. Lastly, the systematic random sampling was used to get final six sampled primary schools. Two schools were from Boma zone and four schools from Chimbende zone.

3.4.3Selection of examinees

The total population of study area was8569 examinees. These were candidates who had just completed grade 8 and were waiting to sit for Primary School Leaving Certificate Examinations (PSLCE) administered by Malawi National Examinations Board (MANEB) in the two zones of Boma and Chimbende in Mangochi district. Then systematic random sampling was used to draw a sample of 1003 (447 boys and 556 girls) from the six selected schools. The sample size was determined using IRT theoretical framework which recommended the minimum sample of 500 if to generate accurate, valid, reliable and invariant parameters during data analysis (Lord, 1980) and Chafutwa (2017) recommended even the use of a sample as small as 200 examinees in study still give viable parameters. Chafutwa in his study used a sample of 200 examinees in 'estimating magnitude of measurement error through the application of generalizability theory: case of remarked MSCE Mathematics paper 1'.

During systematic random sampling of examinees, selection interval (r) was calculated using the formula, r = N/n (where N=population size, n = sample size, r =selection interval). This meant that every K^{th} examinee was selected to form a representative sample of the population i.e. r, r+K, r+2k …r+(n-1)k. Every participant had an equal chance (probability) of being included in the sample of 1003 from the six schools (Cohen et al, 2007).Then selection interval (r) was calculated and found to approximately 2. That meant that every 2^{nd} examinee was picked and included in the final representative sample of examinees. In short, all examinees represented by numbers which were multiple of 2 between 1 and 2007 were picked e.g. 2, 4, 6, 8, 10...

3.5 Data generation and instrumentation

Data were test scores which examinees obtained after ADUS 2015 selection test was re-administered to 2018 standard 8 candidates prior to writing Primary School Leaving Certificate Examinations (PSLCE) to obtain test scores. The test was re-administered in their respective schools with matched conditions of time, duration and environment.

The data generating instrument was ADUS selection test which was used in 2015 to select students into their ADUS mission secondary schools.

The tool (2015 selection test) length was 64 items: thus 18 constructed response items and 44 multiple choice items to be answered in 2.5 hours.

Three experienced MANEB item developers and raters (makers) from primary school practitioners were employed to score the examinees scripts (papers) using set standards of marking key by officials from department of education in the ADUS. The marking key used was the same key which ADUS used in 2015 selection of candidates into form one in

ADUS mission secondary schools in SEED. The researcher generated preliminary data of the sample to give an insight of the examinees that were used in the research study.

3.6 Data analysis

Carlson & Davier (2013) and McAlpine (2002) define item analysis as a method of gauging the quality of an examination (test) by looking at its constituent parts (items). It seeks to give some ideas of how well the examination (test) has performed relative to its purposes. The primary purpose of item analysis in most high education institutions is that of a measurement tool for assessing the achievements of the examinees (candidates) and thus how future learning will be supported and directed.

The quality of test and test items in any public examinations like selection test is always examined through item analysis of examinees' responses either by CTT or IRT methods (theoretical frameworks) (McCowan & McCowan, 1999; McAlpine, 2002; Zimowski, Muraki, Mislevy & Bock, 1996). The study analyzed data (scores) using item response theory(IRT)to establish the quality of test in terms of item difficult, item discrimination and test reliability.BILOG-3.0 and SPSS software were used to generate IRT item parameter estimates like item difficult (**b**-value), item discrimination(**a**-value), item graphics and test reliability (**r**)from scores which examinees got after the ADUS selection test was re-administered to 2018 examinees to establish the quality of the selection test. Bilog-3.0 was chosen by the researcher since it can be used by both dichotomous and polytomous scored data (Thissen, 1991). The two softwares produced parameters at 0.95 confidence level of measurement. The b-Parameters and a-parameters were generated

using two Parametric Logistic Model (2PL model) formula by Adedoyn and Makobi (2013) as given:

$$P_i(\Theta) = \frac{1}{1+e^{-Daj(\Theta - bi)}} \qquad 2PLM$$

(Adedoyin & Makobi, 2013 p. 5)

where $\mathbf{a_i}$ is the item discrimination parameter commonly known as item slope, bi is the item-difficulty parameter commonly known as the item location parameter, **D** is the arbitrary constant (normally D = 1.7) and θ is the ability level of a particular examinee.

The data were interpreted using Adedoyn and Makobi (2013) IRT theoretical framework. Reliability was measured and analyzed using total test information curve (TCC) using the formula given below:

$$SE(\theta) = \frac{1}{\sqrt{\sum_{i=1}^{n} Ii(\theta)}}$$ (adapted from McAlpine, 2002, p. 22)

Where SE (ɵ) =the standard error of estimation (SE_{est}) at ability level ɵ

$\sum_{i=1}^{n} Ii(\theta)$ = the sum of the item information ability level ɵ for all items in the test.

The data were interpreted against McAlpine (2002) theoretical framework as total test information curve.

3.7 Validity and reliability

The researcher used same original 2015 selection test, mark scheme and PCAR guidelines of assessment and interpretation of results tool for primary school which were used by ADUS officials in 2015. Furthermore, three long serving and experienced practitioner experts who were also primary school item developers and rater were used to do the testing processes. The research targeted and administered to 2018 standard 8 candidates only who completed standard 8 syllabuses.

The research study drew a large sample deliberately to increase accuracy (reliability and Validity) in IRT theoretical framework (Lord, 1980).

3.8Ethical considerations

All ethical issues and standards were critically taken care off in the study. According to Strenbert & Carpenter (1999), these help to protect and keep dignity to subjects involved in the research (research participants) and avoid harm that may arise within or after the research report or findings are out. Therefore, the research took all the appropriate steps and measures to safeguard all stakeholders who were involved in data generation procedure like item developers, schools and their examinees, invigilators/ teachers, head teachers of concerned primary schools, government officials, ADUS officials from education department, ADUS secretariat and SEED offices.

The participation of most concerned members in the research was voluntary. Furthermore, the researcher solicited consent from parents of each child (candidate/examinee) and serial numbers rather than real names were used during research study.

All information collected during research was kept under key and lock. The participants were kept anonymous throughout the research or used "pseudo- names*"*. Participants were provided with opportunity to withdraw or not to take part in the research if they wished to do so during generation of these primary data.

CHAPTER 4

RESULTS AND DISCUSSIONS OF THE FINDINGS

4.1 Chapter overview

This chapter presents the findings of the study according to research questions and their discussion. Given first was brief description of how test quality parameters could be calculated and analyzed. Then furthermore, gives the general overview of characteristics of sample examinees and the preliminary data (scores) analysis with discussions on the preliminary findings of the sample to provide an insight of the research findings to the reader. Lastly this chapter concludes by presenting detailed results on item-difficulty, item discrimination, and test information (reliability) parameters and each parameter is immediately accompanied by a discussion of that parameter.

4.2 Quality of test, calculation and analysis of parameters

The quality of test and test items in any public examinations like selection test is always examined through item analysis of examinees' responses either by CTT or IRT methods (theoretical frameworks) (McCowan & McCowan, 1999; McAlpine, 2002; Zimowski et al. 1996). There are so many ways of examining (analyzing) the quality of selection test and test items properties (parameters) statistically. Some of the parameters (indices) that can be statistically are item-difficult, item-discrimination, differential item functioning (DIF), test-reliability and validity.

These statistics can be checked in IRT psychometrical analysis using test score properties and gauge the level of quality of a test used to generate that data (scores) (Zumbo, 1999).

In this research three key parameters of test quality namely item difficulty (b -parameter), item Discrimination (a- parameter) and test reliability (r) in form of test information were analyzed using IRT framework.

The data (scores) obtained using ADUS selection test were analyzed using **BILOG 3.0** at 95% confidence level of measurement to generate the three parameters b, a and r for all 1003 grade 8 examinees and recorded in separate tables.

Then the parameters were interpreted using Adedoyin & Mokobi (2013) work and Test Partnership Limited, 2017 theoretical frame standards in IRT.

4.3 Characteristics of Sample examinees and preliminary data(scores) analysis

The sample consisted of 1003 grade 8 primary school examinees (447 boys and 556 girls) drawn from 6 schools that were randomly selected, using "inclusion and exclusion technique", from 41 Schools in Boma and Chimbende zones.

Examinees minimum and maximum ages were 10 and 20 respectively. All 1003 standard examinees sample sat for ADUS selection test. Their scripts were rated by three professional primary school practitioners and the scores were preliminary analyzed using SPSS software to give the insight of the data.

The descriptive statistics of scores of 1003 examinees obtained using ADUS selection test were shown on the table5.

Table 5: Descriptive statistics of scores of 1003 sample examinees who wrote ADUS Test

Statistic	Statistic Value
Mean	26.54
Median	25
Mode	22
Std. Deviation	11.1
Minimum score	3
Maximum score	75
Skewness	0.706
Range	72

Table 5 above shows that the skewness of data is 0.706, which meant that the distribution of the examinees' scores was positively skewed. This means that most examinees got low scores (marks). That was also confirmed by the most frequent score (mode) being 22 with the mean of 25 which is just slightly above mode but very far from the passing mark of Primary Curriculum and Assessment Reform (PCAR) performance scale of 40.

Furthermore, the scores were widely spread considering the standard deviation value of 11.10 which was just too large for such data which expected a range of scores from 0 to 100. The range of 72 showed again that the difference between maximum and minimum marks (scores) was large.

In summary, all the three central tendency statistics mean, mode and median presented on the table 5were of very low values according to pass mark of 40% (PCAR performance scale and interpretation**.**

It was likely for the study topredict from the preliminary descriptive statistical analysis of scores of 1003 examinees that the grade 8 examinees performed poorly during the 2015

ADUS selection test before actual test quality analyses were carried out by the researcher. Statistic values of mean (26.54), mode (22.0), median (25.0) are of a very lower side and positive skewness (0.706) clearly indicated that most of the examinees preformed very poor on the test.

4.4 Difficult level of the test items

The quality of a test is also determined by the item difficult parameters of items in the test. Item difficult is known as the *b* parameter in IRT. The more difficult an item is, the higher an examinee's ability must be in order to get the item correctly.

The research findings were then analyzed and interpreted using Adedoyin & Mokobi (2013) classification and analysis of item difficult. The item difficult parameters which were found from ADUS selection test were tabulated on **Table6** below.

Table 6: Showing item-difficulty (b-parameters)

Item	b-par	Item	b-par	Item	b-par	Item	b-par
1	0.727	**17**	-1.127	**33**	1.312	**49**	2.554
2	1.262	**18**	-1.049	**34**	4.695	**50**	2.984
3	0.891	**19**	0.517	**35**	0.486	**51**	3.289
4	0.819	**20**	0.250	**36**	2.740	**52**	3.535
5	0.422	**21**	2.117	**37**	1.437	**53**	2.549
6	1.539	**22**	6.535	**38**	3.026	**54**	2.742
7	0.991	**23**	0.554	**39**	0.125	**55**	2.728
8	0.233	**24**	9.853	**40**	2.763	**56**	3.737
9	1.205	**25**	10.507	**41**	0.926	**57**	2.722
10	1.234	**26**	22.315	**42**	1.076	**58**	3.522
11	0.373	**27**	6.150	**43**	0.824	**59**	60.991
12	3.908	**28**	0.117	**44**	1.236	**60**	60.991
13	1.800	**29**	2.071	**45**	0.316	**61**	60.991
14	0.102	**30**	6.755	**46**	2.378	**62**	60.991
15	3.702	**31**	1.525	**47**	1.825	**63**	60.991
16	-0.145	**32**	7.402	**48**	2.425	**64**	60.991

According to Adedoyin & Mokobi (2013) classification of items, items with b-value below -1 were easy items (lower order abilities) for examinees to answer correctly, items with b-value ranging from -0.5 to +0.5 were regarded as items with fair or medium b- values. Those items were said to be not very easy or very difficult for examinees to get correct responses and those items which had b-values greater than +1 were very hard items to be answered correctly.

Using Adedoyin & Mokobi (2013) classification of item-difficulty, **Table 6** clearly shows thatitem17 (-1.127) and item18 (-1.049) had b-values below -1which meant that they were

"very easy items" on the ADUS selection test which was administered in 2015. Those items represented 3.0% of the total test items.

Furthermore, table 7 indicates that items which had b-parameters within the recommended and acceptable range of -0.5 to +0.5

Table 7: Showing items with b-parameters ranging from -0.5 to +0.5

Item	b- par	Item	b- par
5	0.422	**8**	0.117
11	0.373	**35**	0.486
14	0.102	**39**	0.125
16	-0.145	**45**	0.316
20	0.25		

Table 7 indicated that nine items were regarded as items with medium (fair) difficulty parameters levels. Therefore, 9 out of 64 items represented 14.0 % of the total test items, which were good items for grade 8 examinees.

The third category of items had b–parameter values greater than +1. Those items were labeled as difficult items (very 'high order abilities') for the grade 8 examinees whose abilities were very low. In other words, they were beyond the scope of the grade 8 examinees ability in terms of item difficulty level.

Table 8: Showing items with b-parameters greater than +1

Item	b-par	Item	b-par	Item	b-par	Item	b-par
2	1.262	25	10.507	40	2.763	55	2.728
6	1.539	26	22.315	42	1.076	56	3.737
8	2.2331	27	6.15	44	1.236	57	2.722
9	1.205	29	2.071	46	2.3781	58	3.522
10	1.234	30	6.755	47	1.825	59	60.991
12	3.908	31	1.525	48	2.425	60	60.991
13	1.8	32	7.402	49	2.554	61	60.991
15	3.702	33	1.312	50	2.984	62	60.991
21	2.117	34,	4.695	51,	13.289	63	60.991
22	6.535	36	2.74	52	3.535	64	60.991
23	4.554	37	1.437	53	2.549		
24	9.853	38	3.026,	54	2.742		

Table 8 showed items with unacceptable b-parameters (item-difficulty) since they had statistic values above+1and were regarded as difficult for grade 8 examinees to give correct responses. That gave a total number of 46 items out of 64 items which represented approximately 72.0% of items in the test being difficult and did demand more or high abilities from the grade 8 examinees in order to get correct responses. They were supposed not to be included in the test for a good quality test parameters.

In addition, items from number 59 to 64 (six items) had same levels of item–difficulty and were identified as too difficult items (very high order abilities) since they had very large b-parameter (60.991) and very far from a recommended range of b-parameters.

In short, the test comprised a total of 53items which were difficult for the grade 8 examinees which represented a total of 83.0% of the items in the ADUS selection.

To sum up, the test had 2 very easy items, 9 good items and 53 very difficult items for the examinees who had just completed grade 8 work.

In terms of percentage, the test had very difficult (83.0%), very good (14.0%) and very easy (3.0%) items as shown on **Table8.**

Therefore, the researcher concluded that the2015 ADUS selection test was a very hard test using Adedoyin & Mokobi (2013) theoretical framework scale. When test is hard, both high ability and low ability examinees fail to locate correct responses hence both get low scores. This means that it is difficult to identify individual ability levels of each examinee and likely to affect the decision making resulting into making erroneous judgement if the results are used by authority. Similarly, when a selection test is easy both high ability and low ability examinees get high scores hence the test fail again to identify examinees individual abilities resulting into all examinees seem to have same abilities that brings errors in decision making (Razak et al., 2012).

That meant the ADUS selection test could fail to identify standard 8 examinees with high and low abilities and hence selecting examinees who did not deserve.

Figure 2 summarized the data on item-difficulty parameters of the whole test in percentage by levels of difficulty

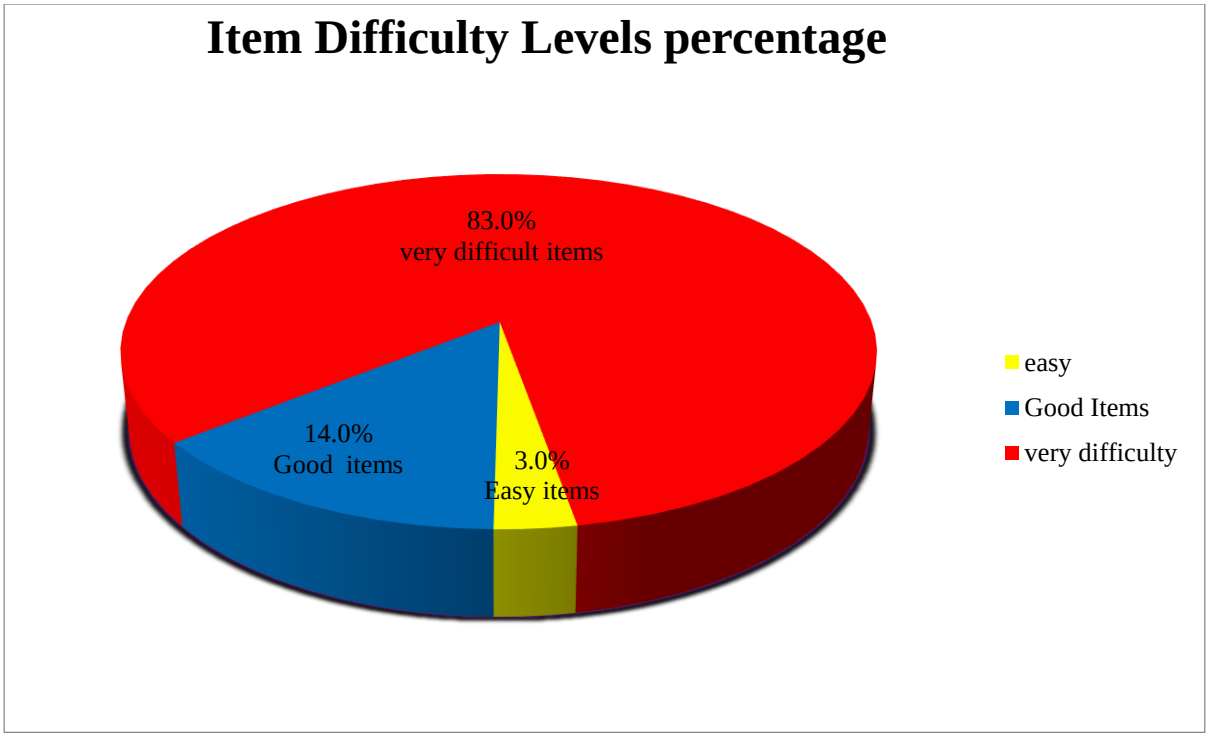


Figure 2: Classification of items according to item difficulty of ADUS selection test

Source: Researcher

4.5 Discrimination indices of selection test items

Quality of a test can as well be described in terms of item-discrimination (**a-**parameter) in IRT analysis. Item discrimination is the ability of an item to separate examinees with high ability from those with low ability. In other words, we expect that an item that demands more ability should be allocated by examinees who have ability equal to or more that the required ability and not examinees with less ability or no ability.

The **a-**parameter is an expression of how well an item can differentiate among examinees with different ability levels. The **a-**parameter is a measure that can also be graphically expressed by the steepness of the ICC.

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

Table 9: Shows discrimination parameters (a-parameters) for the ADUS selection test items

item	a-par	Item	a-par	Item	a-par	Item	a-par
1	8.308	**17**	0.410	**33**	0.740	**49**	0.431
2	1.804	**18**	0.325	**34**	0.599	**50**	0.571
3	2.348	**19**	0.372	**35**	0.735	**51**	0.853
4	2.170	**20**	0.392	**36**	0.250	**52**	0.646
5	1.654	**21**	0.211	**37**	0.266	**53**	0.374
6	0.686	**22**	0.172	**38**	0.198	**54**	0.543
7	0.742	**23**	0.175	**39**	0.393	**55**	0.732
8	0.521	**24**	0.106	**40**	0.154	**56**	0.717
9	0.804	**25**	0.113	**41**	0.585	**57**	0.334
10	0.805	**26**	0.053	**42**	0.810	**58**	0.615
11	0.368	**27**	0.097	**43**	0.480	**59**	0.036
12	0.182	**28**	0.707	**44**	0.868	**60**	0.036
13	0.313	**29**	0.154	**45**	0.418	**61**	0.036
14	0.190	**30**	0.103	**46**	0.711	**62**	0.036
15	0.219	**31**	0.559	**47**	0.529	**63**	0.036
16	0.493	**32**	0.428	**48**	0.447	**64**	0.036

Table 9shows that discrimination parameters (**a**-parameters) of item 2 (1.804) and item5 (1.654) were excellent desired **a**-parameters. Those items were excellent items (excellent discriminators) which constituted 3.0% of the total items of the test which formed the good quality part of ADUS selection test. Those items were the only quality items that could produce very high quality test for selection out of 64 items in terms of discrimination parameter (powers).

Next to those items were Item 9(0.804), item10 (0.805), item42 (0.810), item44 (0.868) and item51 (0.853) which had "acceptable **a**-parameters" in the rangeof+0.75 to +1. Those items could be included in a test since they had item-discrimination parameters that could not compromise test quality very much. They contributed to 8% of the total items of the ADUS selection test.

The following items had "questionable" a- parameters that could be very difficult to draw conclusion on their functions in the test on item-discrimination: item6 (0.686), item7(0.742), item8 (0.521) , item28(0.707), item31 (0.559, item33(0.740), item34 (0.599), item35(0.735), item41(0.585), item46(0.711), item47(0.529),item50 (0.571), item52(0.646), item54(0.543), item55 (0.732), item56(0.717) and item58 (0.615). There were 17 items out of 64 represented 27% of items whose functions could not be easily defined in the test pertaining to item-discrimination. In short, those items were poor item discriminators since their value fall between +0.5 and +0.75 according to Adedoyin & Mokobi (2013) theoretical framework scale.

On the other hand, undesired items fall into two categories; items with a-parameters above+2 and items with parameters below +0.5, since they are out of the "desired and recommended" a-parameters range.

Only three items had **a**-values above the range; item1 (8.308), item3 (2.348) and item4 (2.170). That meant the items failed to discriminate due to some reasons or errors in the item construction process, such that every examinee could get them correct or wrong despite having different abilities. Those 3 items were regarded as poor items to be included in the selection test which represented about 5.0% of the total items in the ADUS selection test.

Contrary to the above, were items which had their **a-**parameters below the required range. Such items were also regarded as poor items since they were easy to every examinee and failed to identify different examinee abilities. Those were item11(0.03), item12(0.03),item13(0.03),item14(0.03),item15(0.03),item 16(0.03),item 17(0.03), item 18(0.03),item 19(0.03),item20(0.03), item 21(0.03),item 22(0.03),item 23(0.03), item 24(0.03), item 25(0.03),item 26(0.03), item 27(0.03), item 29 (0.03),item30 (0.03), item 32(0.03), item 36(0.03), item 37(0.03),item 38(0.03), item 39 (0.03)item 40(2.33), item 43 (2.03), item 45(5.5), item48(2.03) , item 49(3.83), item53( ), item 57(3.13), item 59(17.31), item 60 (4.41), item 61(2.41), item 62(2.03),item 63(3.73), and item 64(6.01), their discrimination **a**–parameters were below the criteria.

Surprisingly, 6 items were constructed with same discrimination values (**a-**parameter) of 0.036 which was not supposed to be in the selection test for their discrimination values

were outside the recommended and acceptable range of **a**-parameters. Those 37 items constituted 58% of the total items of the test.

In summary, the findings have shown that only 2 items were regarded as excellent discriminating items, 5 items were acceptable items, 17 items were questionable items and 37 items were not desired items. Therefore, poor items contributed to 89% of the total test items while the remaining 7items out of 64 items (about 11%) were fit to be included in the test.

Therefore, 89% of the items in the test, in general, failed to discriminate grade 8 examinees. In other words, the test had 11% good quality items and 89% poor or low quality items in terms of item discrimination parameters as shown in **Figure 3**

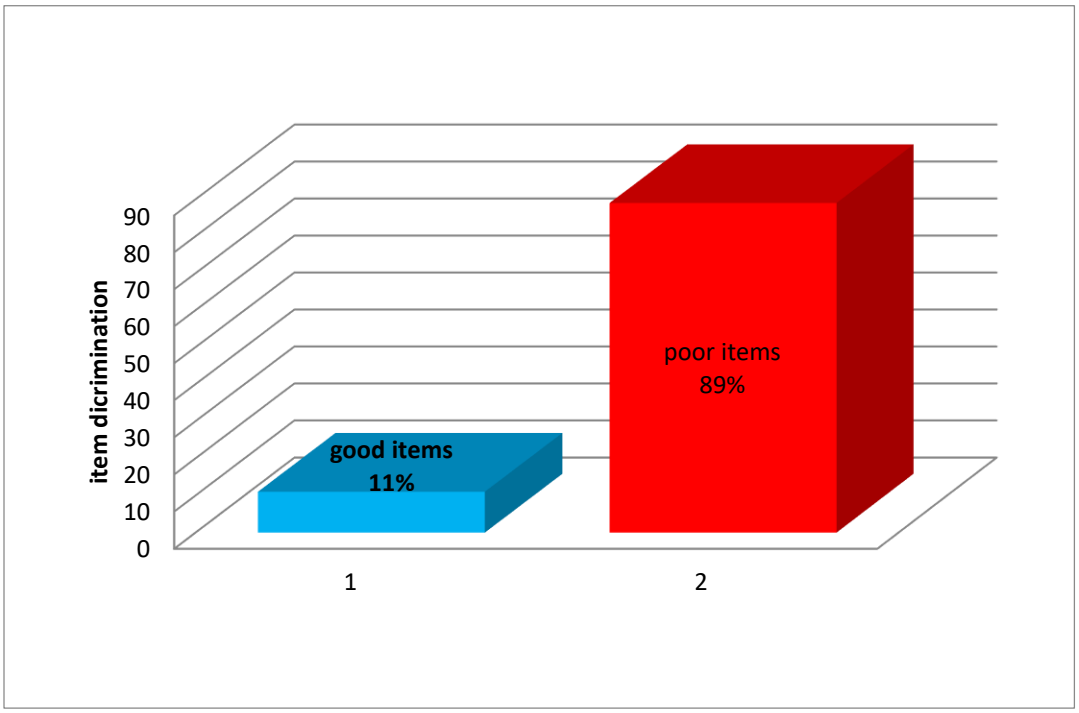


Figure 3: Item-discrimination level
Source: Researcher

Further, critical analysis and observation on parameters, item-difficulty and item-discrimination, out of 64 items, there was only one item (item5)which had excellent a-parameter and fairly acceptable b-parameter but no any item which qualified as an excellent item on both item-difficulty and item-discrimination representing 2% fair items and 0% excellent items of test.

In general, for a test item to qualify as excellent item to be included in a selection test for excellent quality selection test, it must produce excellent properties in all parameters under consideration when analyzed. Therefore, the test showed that it had no any item which qualified on both parameters as excellent or good item for item-difficulty and item-discrimination.

The current findings are contrary to Sahin & Anul (2017) study who found that a test with more than 10 items administered to more 750 examinees provided responses with high accurate **a** and **b** parameters using 2PLM. Therefore, the test length and sample size had no effect on the accuracy of the parameters in this study. That meant, there were other test properties (parameters and conditions) that caused a test to produce unacceptable **a** and **b** parameters (Baker, 1998; Stone, 1992; Yen, 1987). This observation agreed with Nanty (2004) who hinted that, in educational practice, one of the principal tasks is the development of tests that measure the facets of learning with the greatest precision and accuracy, and that is associated with the quality of test items in addition to test length.

4.6 Reliability of the selection test

Reliability of items that form a test is a very important property that describes the quality of the test. It ranges from 0 to 1 but a test to be accepted or used must have reliability above 0.70. The reliability increases by increase in number examinees or test length or both (Test Partnership Limited, 2017).

In this study, reliability is measured in terms of total test information function (TCC). The test information function for 64 items were summed up as shown in **Figure 4.5** from individual item characteristic curves (ICC) on appendix A. the BILOG 3.0 was used to produce the item curves from the scores which standard 8 examinees obtained from ADUS selection test.

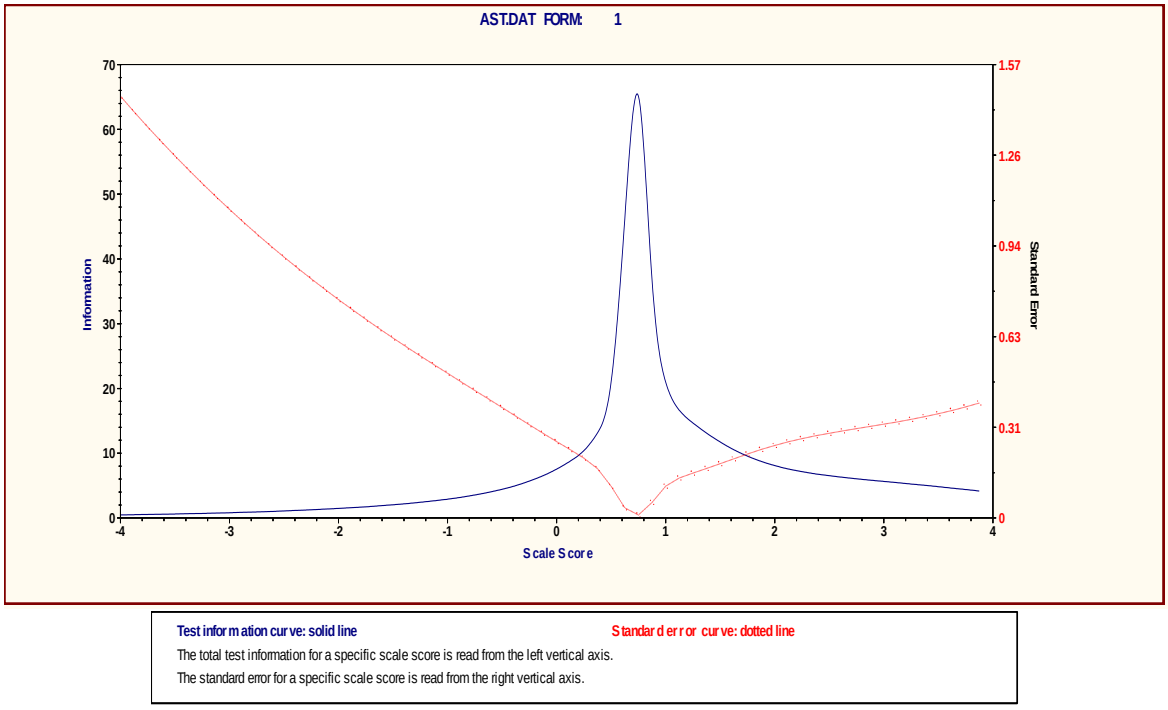


Figure 4: Showing test information of ADUS selection test

Source: Searcher

From **Figure 4**, it is clear that the test provided a lot of information regarding the ability of the examinees in the ability levels between +0.5 and +1.5 with a peak at +0.7. This amount of information is 65, which is adequate but not enough for quality test reliability. The test provided very little information between the ability levels -4 and 0.5, and 1.5 and +4, of the examinees. The test failed to produce minimum test information (reliability) of above 0.70 to account for validity in the IRT framework which meant that it left a lot of information that would help in making decision in selection process by ADUS. This finding implied that the test was best for examinees with ability levels between 0.5 and 1.5 (Test Partnership Limited, 2017). This implies that the test had a narrow range of examinees abilities to assess that might result into leaving a lot of examinees information un assessed hence difficult to make correct discussion on the results by authorities of ADUS.

Partnership Limited (2017) study found that a test with narrow range has implication on amount of test information produced about the examinees. Such tests produce very little information about the examinees which could not be based on when giving decisions.

That meant the selection test was not of the level of the standard 8 examinees since the test demanded more abilities beyond the scope of examinees level since failed to assess wide range of abilities on examinees. The information characteristic curve (ICC) on Fig 4. 5 was generated from the individual item characteristics curves of the test as shown in the appendix A.

The current findings are inconsistent with the findings of Harwell & Jonosky (1991) who found that reliability estimates of a test increase accuracy from a 25 item test with a sample of not less than 250 using 2PLM. Furthermore, Sahin & Anil (2017) suggested a short test

with 10 items with a sample of 750 also give viable parameters (accurate reliability) using same 2 PLM.But this selection test with 64 items and 1003 examinees had failed to achieve minimum requirement of reliability of over 0.7. That meant the test was not very reliable and valid though was lengthy sat by many examinees (Test Partnership Limited, 2017).

The data on appendix A, shows clearly that item 31, item 58, item 59, item 60, item 61, item 62, item 63 and item 64 contribute very little or no information pertaining to test information that determine the reliability of the selection test. These eight items contribute 13% of the total test items. Such items are not supposed to be included in the selection test since they contribute nothing that one can used in making important decision on selection of examinees (Robertson & Mike, 1986).

That meant by including those 8 items which provided less or no information lowered the total test information (reliability and validity) of the test hence the selection test failed to explore abilities of examinees. The test information curve (reliability) provided was not enough to make viable judgement regarding the selection of examinees that deserved, hence was source of error. The selection test is regarded as of good quality (reliable and valid) if it provides test information of above 70% (Razak et al.,2012; Test Partnership Limited, 2017).

CHAPTER 5

CONCLUSIONS, IMPLICATIONS AND RECOMMENDATIONS

5.1 Chapter overview

This chapter is making conclusions with some recommendations based on the research findings by focusing on quality of selection tests of faith mission secondary schools. The conclusion is based on three parameters of the test; item difficulty, item- discrimination and test reliability of ADUS selection test. These conclusions and recommendations are immediately followed by Implications, study's contribution to knowledge and proposed areas for further research areas and for practice is offered last.

5.2 Conclusions

As the study tried to examine quality of ADUS selection tests that were used to select form one students into their schools in relation to disparities in academic performance, the following have been identified:

The general findings of the study, test items of ADUS selection test were constructed from the current syllabus (curriculum) of Malawi. But the quality of the items and their content representation was a problem (poor). Test items had the following challenges that made the whole ADUS selection tests being described as poor quality test for standard 8 examinees:

The preliminary descriptive statistics obtained from ADUS selection test indicated clearly that examinees obtained very low scores and failed the test that could be very hard to use in making decision on selection of examinees. Over 80% of the test items predominantly assessed high order ability skills from examinees that compromised the item-difficulty parameters and reduced test quality. Most of the items had b-parameters below -1 and above +1 which represented easy and very hard items. Those items reduced quality of test in general according to Adedoyn and Makobi (2013) who found that a test with item-difficulty parameters outside the range of -1 and +1 is a poor test.

Most of the test items (about 90%) failed to discriminate examinees with high ability levels from those with low ability examinees. The items were failing to identify and separate examinees with different ability (cognitive) levels and due to that some items were found to have same discrimination parameters (powers) that diluted the quality of test items in terms of item-discrimination parameters. This showed that item developers concentrated on few elements of the curriculum and cognitive levels (abilities) of examinees which symbolized the teacher made test characteristics found by Newel (2002) during his study where he discovered that such tests assess limited part of the curriculum and few cognitive levels.

The study found that the selection test failed to achieve minimum recommended value of reliability (over 0.7). That meant the test lacked both reliability and validity since reliability is ‘precondition’ for validity. The study concluded that the scores obtained through such tests were not reliable to use and the tests were measuring different abilities altogether from

the examinees rather than the intended ones (Test Partnership Limited, 2017). Therefore, as far as the standard 8 examinees (schools) used in the sample, the ADUS selection test had low test quality properties in terms of item-difficulty, item-discrimination and test reliability. The ADUSS selection test could have contributed to disparity in performance between FMSS and GSS in faith mission secondary schools.

In short, the ADUS selection tests were making significant errors in selecting standard 8 examinees which contributed to low academic achievements of FMSS compared to GSS in faith mission secondary schools. The study recommended the use of experts in test development and testing process for better quality tests that may function properly and pick examinees that deserve and perform better at their schools. The ADUS should devise a deliberate mechanism that could help to enroll both faithful and unfaithful students into their schools that could perform similar to GSS at national level (MSCE). Stakeholders should be educated and make the 'Church commitment Criteria' transparent as much as possible to gain public confidence in faithful.

ADUS authority should consider developing or using table of test specifications and follow it during test construction as a guide. The study recommended that any test which ADUS would like to use as a selection test (high stakes) should be pre-tested and analyzed to check and establish the quality of test parameters in order to estimate the degree of error it could make before final testing the examinees.

5.3 Implications

The use of ADUS selection test may have the implications on selection of standard 8 examinees as follows:

There was possibility that the deserving students might have been left out during selection due the poor quality properties of tests which could become the root cause for disparities in performance between government selected students and faith mission selected students at faith mission secondary schools.

The study might have influence on policy formulation that will govern the administration processes of selection and enrolment of examinees into their faith mission secondary schools.

The knowledge from the study would help to perfect and increase trust worth, reliability and validity of data so that decision would be made on valid and reliable data.

5.3 Study's contribution to knowledge

The knowledge will assist ADUS authority to consider developing table of test specifications as a guide for reliable and valid examination.

The issue of cut scores will be considered as variable and not fixed as they used to do with ADUS selection tests.

Test or item analysis of selection test will be considered vital for the ADUS to reduce errors in selection processes.

5.4 Proposed areas for further research studies

However, further research studies are required to compliment the research or related ones:

There is a need for research quality of ADUS selection tests using qualitative approach to appreciate and compare the findings since the study used purely quantitative paradigm.

There is a need to compare MANEB selection tests and ADUS selection tests and set equivalent cut scores rather than just using 50 without proof which Malawians use and so that the scores could be interchangeable used.

Further study is required on other faith selection tests that are equally used to select students for enrolment into faith secondary schools since the study concentrated on ADUS tests from so many faith mission schools.

5.5 For practice

Furthermore, knowledge from this research will assist authority at ADUS in understanding the issues that have to be considered when developing quality selection tests and possibly realize area(s) in test development where improvement could be needed.

It will act as guide in any operations regarding selection test processes.

The knowledge of the study will assist test item developers and experts in ADUS and mission secondary schools to shift fully from mix of CTT and IRT to ITR conceptual frame work in data or item analysis since they deal with very large data.

REFERENCES

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*.

An, X. & Yung, Y. (2014).*Item response theory: What It is and how you can use the IRT procedure to apply it*. USA, Chapel Hill: Professional Testing Inc.

Association of Christian Educator in Malawi and Malawi Government (ACEM) (2007).*Memorandum of understanding*. Lilongwe: ACEM.

Australian Medical Assessment Collaboration (AMAC) (2014). *Determining the quality of assessment items in collaboration: aspects to discuss to reach agreement*. Retrieved from http://www.acer.org>file>quality.

Backhouse, J. (1972).Reliability of GCE examinations: A theoretical and empirical approach. In D.L. Nuttall & A.S. Willmott (Eds.), *British Examinations: Techniques of Analysis*(pp. 56-68). Slough: NFER Publishing Co.

Baker, F. B. (1998). An investigation of the item parameter recovery of a Gibbs sampling procedure. *Applied Psychological Measurement*, *22*(2), 153–169.

Banda, K. (1996). *History of education in Malawi*.Blantyre: Dzuka Publishing Company.

Barrett, P. Kline, P., Paltiel, L., & Eysenck, J. H. (1996). An evaluation of the psychometric properties of the concept 5.2 occupational personality questionnaires. *Journal of occupation and organizational psychology*, (1996), 69, 1-69.

Bisika, J. (1996). *Malawi Policy Implementation Framework*. Zomba: University of Malawi.

Blinkhorn, S. &Johnson, C. (1990). The insignificance of personality testing. *Nature*, *348(6303),* 671-672.

Boupathiraj, C., & Chellamani, K. (2013). Analysis of Test Items on Difficulty Level and Discrimination Index in the Test for Research in Education. *International Journal of Social Science and Interdisciplinary Research*, *2* (2), 189-193

Burrows, O. (2015). Form one selection was fair, says Kaimenyi. *Capital News*. Retrieved from www.capitalism.co.ke/news/2015/02/form-one-selection-was-fairsays-kaimenyi.

Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items*. Newbury Park, CA: Sage.

Carlson, J. E. & Davier, M. V.(2013). Item response theory Educational Testing Service. Princeton, New Jersey *ETS Research Report No. RR-13-28*.

Chafutwa, S.E.A. (2017). *Estimating the magnitude of measurement error through the application of generalizability theory: a Case of Remarked MSCE Mathematics Paper I* Zomba: University of Malawi.

Chakwera, E. W.J. (2002). *Introduction to testing, measurement and evaluation.* Education Module 8. Zomba: Domasi College of Education.

Chimombo, J., Meke, E., Zeitlyn, B.,& Lewin, K. M. (2014).Increasing access to secondary school education in Malawi: Does private schooling deliver on its promises? *Privatisaion in Education Research Initiative: ESP Working Paper Series 2014 No. 61*

Churches Together in Britain (CTB), & Ireland, (2015). *Admission to church schools in Malawi. Retrieved from https://www.ctbi.og.uk*

Coe, R., Jones, K., Searle, J., Kokotsaki, D., Kosnin, A., & Skinner, P. (2008*). Evidence on the effects of selective educational systems (Report).* UK: Durham University.

Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education* (6th ed.). Milton Park: Routledge

Creswell, J. W. (2004). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research.* Upper Saddle River, NJ: Merrill Prentice Hall.

Creswell, J. W. (2014).Research design: Qualitative, quantitative, and mixed methods approach (4th ed.). Thousand Oaks, California: SAGE Publications.

Ebel, R. L. (1979). *Essentials of educational measurement* (3rd ed.). Eaglewood Cliffs, NJ: Prentice-Hall.

Embretson, S.E., & Reise, S.P. (2000). *Item Response Theory for Psychologists*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Fletcher, C. (1991). Personality tests: The great debate. *Personnel Management, September (1991(, 38-42.* Hambleton, R. K. & Jones, R. W. (1993). Comparison of Classical Test Theory and Item Response Theory and their Application to Test Development.
Educational Measurement: Issues and Practice, 12(3), 38-47.

Hambleton, R. K., Swaminathan H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.

Harwell, M. R., & Janosky, J. E. (1991). An empirical study of the effects of small datasets and varying prior variances on item parameter estimation in BILOG. *Applied Psychological Measurement, 15*(3), 279–291.

Hopkins, K. D. (1998). *Educational and psychological Measurement and evaluation* (8th ed.). London: Prentice Hall.

Huang, C. (2003).*Psychometric analyses based on evidence-centered design and cognitive science of learning to explore students' problem-solving in physics*. USA: University of Maryland, College Park publication.

Hughes, A. (2005).*An Exploration into the effectiveness of Personality testing within the workplace for the purpose of selection and recruitment.* Dublin:H. R.M. National College of Ireland.

Ironson, G. H. & Subkoviak, M. (1979). A comparison of several methods of assessing item bias. *Journal of Educational Measurement, 16(4), 209-225.*

Ironson, G. H. (1977). *A comparative study of several methods of assessing item bias* (doctoral dissertation). University of Wisconsin, USA.

Jacob, P., Jepkenei, E.,& Chepwarwa, J., Makori, R. A.,(2015,September 3). *Admission into Public secondary school in Kenya: Understanding parental preferential limitations*. Retrieved from https://www.semanticscholar.org>paper……..

Jailos, H.O. (2017, August17th to 19th). *Education Report. Presented at the Sixth Anglican Diocesan of Upper shire Holy Synod*,. Church of Ascension Likwenu Parish, Malosa, Zomba.

Kadzitche, P. (2018).*Validity and reliability of teacher-made tests in Zomba Primary Schools: the case of mathematics, science and technology and English*. Zomba: University of Malawi, Chancellor College:

Khalid, M.N. (2009). *IRT Model Fit From Different Perspectives* (Doctoral Thesis). University of Twente, Netherlands.

Kaira, L. (2002). *Malawi teachers' knowledge of and attitudes towards standardized tests Master's Capstone Projects*. Retrieved from .http://scholarworks.umass.edu/cie_capstones/96.

Kinyua, S.G. (2014). *Determinants of students' performance in Kenya Certificates of Secondary Education using Logistic Regression* (Master's thesis). Nairobi University, Kenya.

Kline, T.J.B. (2005). *Classical test theory: Assumptions, equations, limitations, and item.* Thousand oaks, CA: SAGE Publication.

Kline, T.J.B.(2015). *Psychological testing: A practical approach to design and evaluation.* Thousand oaks, CA: SAGE Publication.

Kothari, C. R. (2004). *Research methods: Methods and techniques* (2nd ed.). New Delhi: New Age International Limited Publisher.

Linn, R. L.,& Gronlund, N. E. (2000*). Measurement and assessment in teaching* (8th ed.). New Jersey: Prentice-Hall Inc.

Litereko, E. (2017). *Rooted In Jesus*: *Report to Church of England*. Malosa, Zomba, Anglican Diocese of Upper Shire.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale,NJ: Lawrence Erlbaum.

Malambo, B. (2012). *Factors affecting pupil performance in grant aided and non-grant aided secondary schools: A case of selected secondary schools in the Western Province of Zambia*. The University of Zambia: Lusaka Zambia.

Makori, A., Onyura, G., Cheboiwo, F., Yegon, J.& Kandie, J. (2015). Form one selection process, an encouragement or a discouragement: Examining parents' perceptions in Baringo County, Kenya. *Merit Research Journal of Education and Review*, 3(7), 228-234

Manjombe, E. (2018). *Equating standard and modified test forms for examinees with total visual impairment: a study of MSCE Geography examinees in South East Education Division*. Zomba: Chancellor College.

Massey, O. (1995). Evaluation and analysis of examination data: some guidelines for reporting and interpretation, UCLES internal report, Cambridge.

Mazzeo, J., Schmitt, A. P., & Bleistein, C. A. (1993). *Sex-related performance differences on constructed-response and multiple-choice sections of advanced placement examinations (College Board Report No. 92-7).* New York: College Entrance Examination Board.

Mbunge, J. C. F. (1986). *Content validity analysis of the Junior Certificate of Education (JCE) examinations in Geography and History in relation to school syllabus and their relevance to everyday life in Malawi: 1975-1985* (Master's thesis). University of Malawi, Zomba.

McAlpine, M. (2002). *A summary of methods of item analysis.* England: University of Glasgow.

McCowan, R. J., & McCowan, S. C., (1999). *Item analysis for criterion references*. Test Research Foundation of SUNY State University College Buffalo. New York: Elmwood.

Mehraj, H., Taufique, T., Ona, A. F., Nusrat, A., & Uddin, A.F.M. J.,(2014). Performance of gerbera cultivars under different wavelengths of solar spectrum. *Journal of Bangladesh Academy of Sciences, 38 (1), 27-37*.

Messick, S. (1989). Validity. In R. L. Linn (Ed.),*Educational measurement* (3rd ed., pp. 23-38). New York: American Council on Education and Macmillan.

Ministry of Education, Science and Technology [MOEST] (2016). *Education Management Information System.* Lilongwe: Author

Moss, P.A. (1994). Can there be validity without reliability? *Education Researcher, 23(2), 5-12.*

Nenty, H.J.,(2004). *The application of item response theory in strengthening assessment's role on the implementation of national education policy*. Gaborone: University of Botswana (UB).

Nichols, S. L., & Berliner, D.C., (2007). C*ollateral damage: How High-stakes testing corrupts America's Schools.* Cambridge, MA: Harvard Education Press.

Nitko, A, J.(1983). *Educational tests and measurement: An introduction.* New York: Harcourt Brace Joovanovich, Inc.

Nitko, A.J., (1996). *Education assessment of students* (2nd ed.). Jersey City: Prentice-Hall.

Norman, G.R., Smith, E.K.M., Powles, A.C., Rooney, P.J., Henry, N.L., & Dodd, P.E. (1987). Factors underlying performance on written tests of knowledge. *Medical Education*, 21(4), 297–304.

Obinne, A.D.E., (2011).A Psychometric analysis of two major examinations in Nigeria: Standard error of measurement. *International Journal of Education Science, 3*(2),137-144

Priscilla, A.A. (2011). Teaching in a changing Africa: Differential academic performance of students from academies and public primary schools at KCSE Examination in Kenya. *International Journal of Innovative Interdisciplinary Research, 1(1),8-18.*

Razak, N.A, Khairanib, A.Z, Thien, L.M. (2012). Examining quality of mathematics test items using rasch model: Preliminarily analysis.*Procedia Journal - Social and Behavioral Sciences, 69* (2012), *2205 – 2214.*

Robertson, I. & Mike, S.J. (1986). *The theory of practice of systematic personnel selection*. Retrieved on https://www.palgrave.com/research>book

Runder, L. M., Getson, P. R., & Knight, D. L. (1980). A Monte Carlo comparison of seven biased item detection techniques. *Journal of Educational Measurement, 17(1), 1-10.*

Safuli, S. (1996). *Education Development in Malawi: History of Education in Perspective*. Lilongwe: Maneno Press.

Sahin, A., & Anıl, D. (2017). The effects of test length and sample size on item parameters in item response theory. *Educational Sciences: Theory & Practice, 17(1),* 321–335.

Sandikonda, V.C. (2013). *Admission policy of students into Malawi secondary schools*. A dissertation.Retrieved fromhttps://wwww.uir.unisa.ac.za>edu.

Saville, P. & Sik, G. (1992). *Selection tests*. Retrieved from https://www.researchgate.net>publications

Schmitt, N., Gooding, R. Z., Noe, R. A. & Kirsch, M. (1984). Meta-analysis of validity studies published between 1964 and 1982 and the investigation of study characteristics. *Personnel Psychology,37(3),407-422.*

Schuwirth, L.W.T., Bosman, G., Henning, R.H., Rinkel, R., &Wenink, A.C. (2010). Collaboration on progress testing in medical schools in the Netherlands. *Medical Teacher*, 32(6), 476–479.

Shepard, L. A., Camilli, G., & Williams, D. M. (1985). Validity of approximation techniques for detecting item bias. *Journal of Educational Measurement*, 22(2), 77-105.

Shepard, L. A., Camilli, G& Averill, M. (1981). Comparison of procedures for detecting test-item bias with both internal and external ability criteria. *Journal of Educational Statistics, 6(4), 317-375.*

Smith, B.D. (1988). *Measurement of intelligence and personality within the cattellian psychometric model*. Retrieved from https://www.researchgate.net>publication

Smith, E.V. Jr.& Smith, R.M. (Eds.) (2004).*Introduction to Rasch measurement: Theory, models, and applications*. Maple Grove, MN: JAM Press.

Souza, A.C., Costa N.M., & Guirardello A.E. (*2017). Psychometric properties in instruments evaluation of reliability and validity*. Universidade Estadual de Campinas, Brasil.

Stone, C. A. (1992). Recovery of marginal maximum likelihood estimates in the two-parameter logistic response model: An evaluation of Multilog. *Applied Psychological Measurement, 16*(1), 1–16.

Strenbert, H.J.,& Carpenter, D.R. (Eds.)(1999). *Qualitative research in nursing: Advancing the humanistic imperative*(2nd ed.). Philadelphia: Lippincott.

Subkoviak, M. J., Mack, J. S., Ironson, G. H., & Craig, R. D. (1984). Empirical comparison of selected item bias detection procedures with bias manipulation. *Journal of Educational Measurement*, 21(1), 49-58.

Tavakol, M. & Dennick, R. (2011). Making sense of Cronbach Alpha. *Internal Journal of Medical Education*, 2(1), 53-55.

Tengatenga, J. (2006*). Church, state and society in Malawi, the Anglican Case*. Blantyre: Dzuka Publications.

Tengatenga, J. (2010). *The UMCA in Malawi, History of Anglican Church*. Zomba: Kachere Series.

Test Partnership Limited (2017, March). *Insights Series Technical Manual* Retrieved from https://www.testpartnership.com/psychometric-test.htm

Thissen, D. (1991). *Multilog: Multiple category item analysis and test scoring using item response theory* [computer software]. Chicago: Scientific Software International.

Traub, R. (1997). Classical Test Theory in Historical Perspective. *Educational Measurement: Issues and Practice*, *16* (4), 8-14. Retrieved from doi:doi:10.1111/j.17453992.1997.tb00603.x

US-Department of Labor, Employment and Training Administration (US-DLETA) (2000, November). National guideline standards of apprenticeship for international pipe trades joint training.Retrieved from https://www.onetcenter.org>dl-files.

Ward, W.C. (1982). A comparison of free responses and multiple-choice forms of verbal aptitude tests. *Applied Psychological Measurement*, 6(1), 1-11.

Weaver, K. (2011). *Standardized testing measurement of academic achievement.* Virginia: Liberty University Lynchburg.

West, A., & Hind, A. (2016).*Secondary school admissions in London 2001 to 2015: Compliance, complexity and control.* London: LSE Academic Publishing.

Wilmut, J., Wood, R., & Murphy, R. (1996). *A review of Research into the Reliability of Examinations.* Retrieved from https://cerp.aqa.org.uk>pdf-upload.

Woltjer, L. (2006). *Analysis of vocational education and training in Malawi. Retrieved from https://www.teveta.mw*

World Bank (WB) (2004). *Cost, Finance and School Effectiveness of Education in Malawi a Future of Limited Choices and Endless Opportunities. Retrieved from siteresources.worldbank.org.mw>educ*

Wright, W.D., & Stone, M.H. (1979). *Best test design*: *Rasch measurement*. Chicago: MESA Press.

Yen, W. M. (1987). A comparison of the efficiency and accuracy of Bilog and Logist. *Psychometrika*,*52*(2), 275–291.

Yildiz, O., Altundag, E., Cetin, B., Guner, S.T., Saginci, M., & Toprak, B. (2017). Afforestation restoration of saline-sodic soil in the Central Anatolian Region

of Turkey using gypsum and sulfur. *Silva Fen nice, 51 (1B),41-45.* Retrieved from http://doi.org/10.14214/sf.1579.

Young, D. J., & Fraser, B. J. (1994). Gender differences in science achievement: Do school effects make a difference? *Journal of Research in Science Teaching*. *Retrieved from https://books.google.mw>book.s*

Yu, C., (2008). *True score model and item response theory*. Retrieved from https://www.rg.nl>files>thesis.

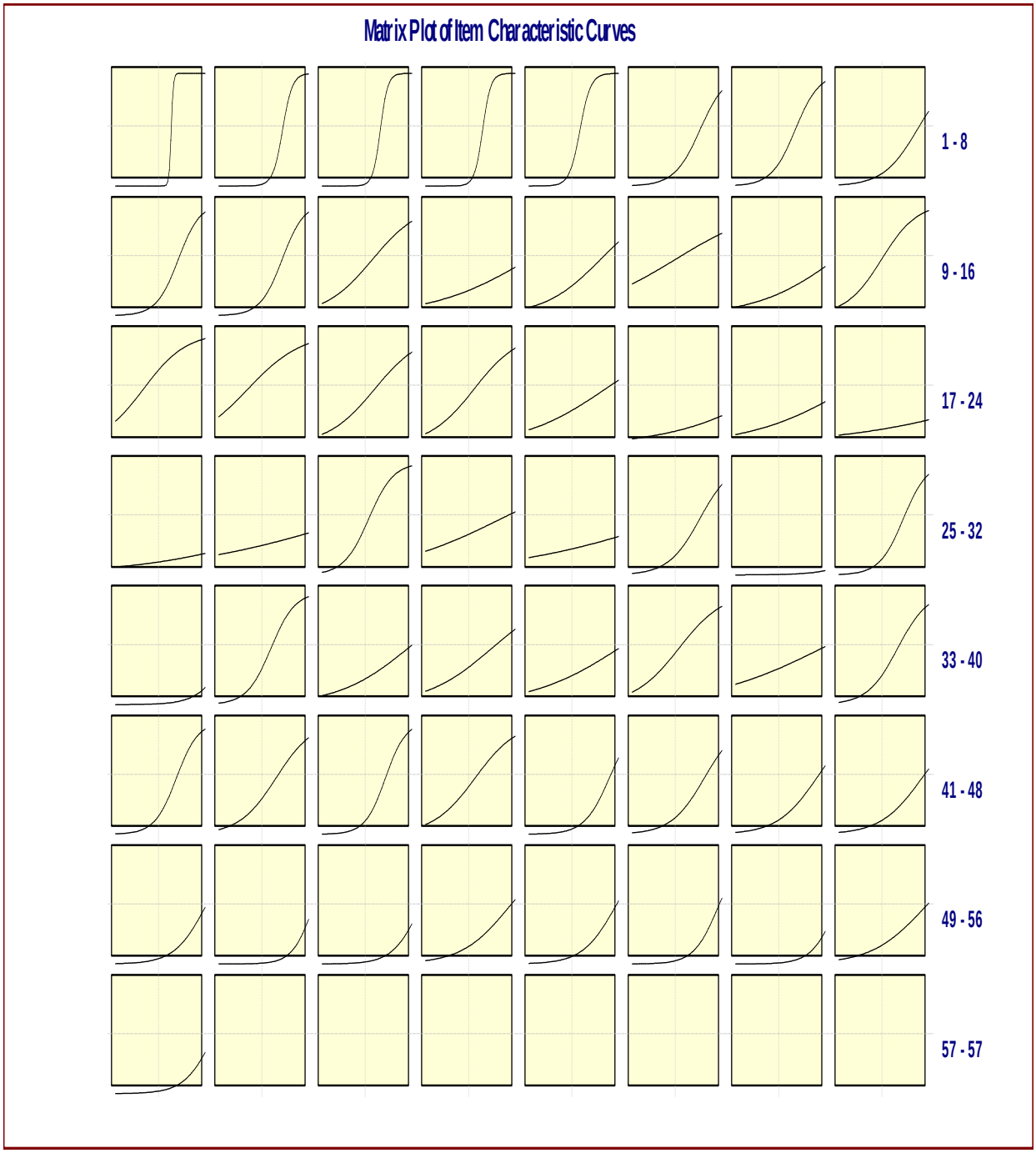
Zenisky, A.L., Hambleton, R.K., & Robin, F. (2004). DIF detection and interpretation in large-scale science assessments: Informing item writing practices. *Educational Assessment,* 9 (1-2), 61-78.

Zimowski, M.F., Muraki, E., Mislevy, R.J., & Bock, R.D., (1996). *Bilog-mg: Multiple-group IRT analysis and test maintenance for binary items [computer software*]. Chicago: Scientific Software International.

Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and likert-type (Ordinal) item scores.* Ottawa, ON: Department of National Defense.

APPENDICES

Appendix A: Matrix Plot of Item Characteristic Curves



Appendix B: ADUS selection test for 2015

ANGLICAN DIOCESE OF UPPER SHIRE

2015 FORM ONE ENTRANCE EXAMINATIONS FOR MALOSA AND ST. MICHAEL'S SECONDARY SCHOOLS

100 Marks

Date: Friday, 05th June 2015 **Time allowed: 2½**

Instructions

1. This paper contains 14 pages. Please check.
2. There are four sections in this paper. Section A: English, Section B: Mathematics, Section C: Primary Science and Section D: Bible Knowledge.
3. Follow instructions for each section very carefully.
4. Make sure you write your Name, Primary School Leaving Certificate Examination Number, Parish, Diocese, clearly on top of each page of the question paper.
5. Cheating will lead to disqualification.
6. Hand in your answer sheet to the invigilator when time is called to stop writing.

SECTION A: ENGLISH

COMPOSITION (15 MARKS)

Answer **ONE** question only from this Section, using the lined answer sheet provided on this question paper. Write between 100 and 150 words.

EITHER

1. (**A**) Write a composition of three paragraphs with the title "The competition" include the following information:

   Paragraph 1

   - What was the competition about?
   - When it took place?

Paragraph 2

- Explain where the competition took place
- What competition prizes were given?

Paragraph 3

- What happened after you had won the competition?
- What you did with the prizes which you received?

OR

(**B**) Write a letter to your Uncle, telling him that you have lost a book, include the following information:

Paragraph 1

- Tell him that you have lost a book
- When it was lost

Paragraph 2

- What the teacher told you to do?
- How much you should pay and when you should pay it?

Paragraph 3

- State your feelings
- How sorry you are that you trouble him with money

_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____

SECTION B
COMPREHENSION
(5 MARKS)

Read the following passage and answer questions that follow:

Florence Nightingale was a very famous nurse. She was born and lived in England.

As a young girl, Florence liked to take care of sick and very old people in the area where she lived. This encouraged her to become a nurse when she grew up.

At the age of seventeen, she wrote in her dairy that god had called her to become a nurse. She studied nursing. In 1854, there was war in Europe. Florence volunteered to go and help the English Soldiers who were wounded in war. She became the leader of women nurses in the war. The first thing which Florence did was to improve the living conditions of the wounded Soldiers. She worked very hard every day. She cleaned all rooms where wounded soldiers lived and she washed their clothes and beddings. She also gave the Soldiers good and clean food.

QUESTIONS

2. Who was Florence Nightingale?
   _____
3. Where was she born?
   _____
4. What was her nationality?
   _____
5. What is the meaning of <u>volunteered</u> according to the passage?

_____

6. Which was the first thing Florence did for the wounded Soldiers?

_____ _____

ACTIVE VOICE (5 MARKS)

In questions 7-11 change the sentences form Active voice to Passive voice

7. He teaches English

_____

8. George will meet us

_____

9. Mr. Brown gave us a lesson

_____

10. Mary is cleaning the room

_____

11. The dog killed the rat

_____

SENTENCE CONSTRUCTION (5 MARKS)

In questions 12 – 16 make sentences using the words given to show that you understand their meaning

12. Observe

_____

13. Watch

_____

14. Expect

_____

15. Consider

_____

16. Suggest _____

STRUCTURE (5 MARKS)

In questions 17 – 21 choose the most suitable word that best completes the following sentences

17. The road was so crowded _____ people

a. in b. of c. with d. at

18. Have you found the solution _____ the problem?

a. with b. to c. about d. from

19. The teacher was angry _____ the student who came late
a. to b. for c. of d. with

20. The clerk was dismissed _____ his post
a. off b. from c. out d. up

21. The teacher pointed _____ the student's tense mistakes
a. into b. over c. up d. out

CLAUSES (5 MARKS)

In questions 22 – 26 choose the answer that gives the correct type of clause and its function

22. Sandile, <u>whose eye was injured,</u> is in hospital
a. Noun clause, object of 'Sandile"
b. Adjective clause qualifying noun 'Sandile'
c. Adverb clause, modifying 'Sandile'
d. Noun clause, subject of 'is'

23. We arrived <u>while we were having breakfast</u>
a. Adverb clause, modifying 'arrived'
b. Adjective clause, qualifying 'we'
c. Noun clause, object of 'arrived'
d. Noun clause, subject of 'were having'

24. You will fail <u>unless you work much hard</u>
a. Adjective clause, qualifying 'you'
b. Noun clause, complement of 'unless'
c. Adverb clause, modifying 'will fail'
d. Noun clause, object of 'will fail'

25. She said <u>that she was feeling ill</u>
a. Adverb clause, modifying 'said'
b. Noun clause, object of 'said'
c. Adjective clause, qualifying 'she'
d. Noun clause, subject of was feeling

26. It was Bonga <u>who three the stone at me</u>
    a. Noun clause object of 'was'
    b. Adverb clause, modifying 'was'
    c. Noun clause subject of 'was'
    d. Adjective clause, qualifying 'Bonga'

SECTION B
MATHEMATICS
(25 MARKS)

Answer questions 27 - 36 by encircling the letter A, B, C or D representing the right answer

27. Which of the following is the largest fraction?
    A. $\frac{1}{10}$ B. $\frac{1}{5}$ C. $\frac{2}{5}$ D. $\frac{1}{2}$

28. In a Khola $\frac{1}{5}$of goats are he-goats. How many she-goats are there if the total number of goats is 75?
    A. 15 B. 90 C. 80 D. 60

29. Solve the inequality $6y+4 \leq 16$
    A $y \leq 4$ B. $y \leq 2$ C. $y \leq 20$ D. $y \leq 16$

30. Simplify $(0.112 \div 5.6) \times 4.2$
    A. 0.084 B. 8.4 C. 0.84 D. 84

31. By selling a radio at K4200, a trader makes a loss of 30%, calculate the cost price of the radio
    A. K1260 B. K2940 C. K5460 D. K6000

32. Subtract 502kg 72g from 1201kg 20g
    A. 698kg 48g B. 698kg 948g C. 699kg 20g D. 699kg 948g

33. By how much is 0.75 greater than $\frac{3}{5}$. Give your answer as a decimal fraction?
    A. 0.45 B. 750 C. 0.15 D. 1.35

Study the following table and answer questions that follow:

Learner	Number of cups
Isaac	☐ ☐ ☐ ☐ ☐
Mable	☐ ☐ ☐
Steven	☐

34. If ☐ represents 4 cups, how many cups has Isaac?
    A. 3 B. 20 C. 5 D. 15

35. A lorry can carry goods worth 7 tonnes (1 tonne = 1000kg). How many bags of 70kg each would the lorry carry?
    A. 300 bags B. 226 bags C. 100 bags D. 319 bags

36. Find the simple interest on K400 for 6 months at 10% per annum
    A. K80.00B. K20.00 C. K200.00 D. K380.00

37. A cook is paid K2500 in 10 days of work. What is the pay for 24 days? **(3 marks)**

    _____
    _____
    _____
    _____
    _____
    _____
    _____

38. Ms. Moyo spent of her monthly salary on food and on bus fare. If she left with K4560, how much was her salary? **(3 marks)**

    _____
    _____
    _____
    _____
    .

39. A school started 5086 learner. By the end of the school year 827 joined and 591 had dropped out. What was the enrolment at the end of the year? **(3 marks)**

_____
_____
_____
_____
_____
_____

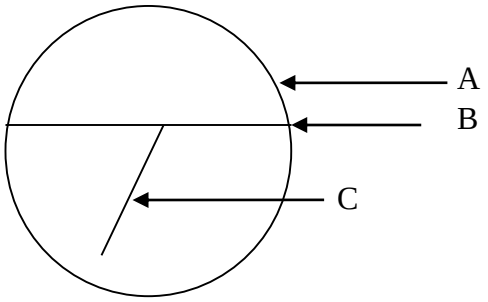
Table 1 shows premium on insured property

Type of Property	Premium per month (in Kwacha)
Bus	K6000
Toyota car	K2500

40. If a company insured two buses and a Toyota car, calculate the total insurance paid for one year **(3 marks)**

_____
_____
_____
_____
_____
_____

41. Name the parts of the circle marked A, B and C below **(3 marks)**



A _____
B_____
_

C_____
_

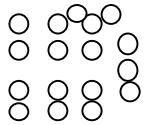
SECTION C

PRIMARY SCIENCE (20 MARKS)

Question 1 to 5 encircle the letter A, B, C or D representing the right answer

42. What name is given to the production of ornamental plants?
A. Arboriculture B. Agronomy C. Floriculture D. Horticulture

Figure 1 represents samples showing texture classes of soil. Use it to answer question 2



X

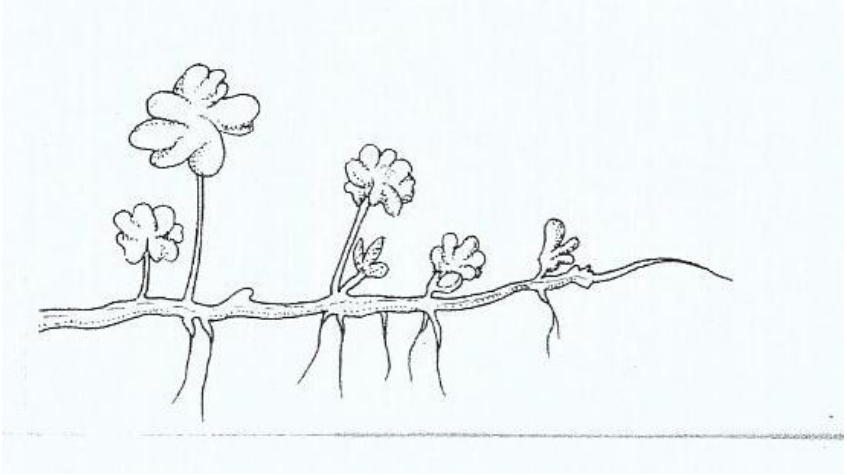
Z

43. The soil sample Z holds more water than X because it has:-
A. many air spaces C. small soil particles
B. large air spaces D. large soil particles

44. Which of the following plants produce their own food?
1. mosses 2. liverworts 3. moulds 4. ferns
A. 1,2 and 3 B. 1,2 and 4 C. 2,3 and 4 D. 1,3 and 4

45. Name the last stage in scientific investigation
A. Planning C. Conclusion stage
B. Implementation D. pre-requisite

Fig 2. Shows one type of stem. Study it and answer question 47



46. Name the type of stem drawn above

A. creeping B. climbing C. erect D. underground

In question 6 to 10 write your answers in the spaces provided after each question

47. List down any two ways of improving soil fertility **(2 marks)**

a. _____

b. _____

48. State any two factors to consider when selecting a site for a fish pond **(2 marks)**

a. _____

b. _____

49. Mention any two areas where pineapples are grown in Malawi **(2 marks)**

a. _____

b. _____

50. Which malnutrition diseases can be prevented by **(2 marks)**

a. eating a variety of food rich in Vitamin A

_____

b. eating a variety of food rich in iodine

_____

51. State any two causes of damage to the brain **(2 marks)**

a. _____

b. _____

52. Which method can be used to separate the mixture of tea leaves from tea **(1 marks)**

_____

53. Mention any two problems of the digestive system **(2 marks)**
    a. _____
    b. _____

54. Name two examples of ball and socket joints **(2 marks)**
    a. _____
    b. _____

SECTION D

BIBLE KNOWLEDGE
15 MARKS

55. Name the animal which was used as a sacrifice during the Passover **(1 mark)**

_____

56. Why did Moses run away from Egypt to Midian? **(2 marks)**

_____

57. What does the word trinity in Christianity mean? **(1 mark)**

_____

58. Where did Jesus start his ministry? **(1 mark)**

_____

59. Mention any two Judges of Israel **(2 marks)**

_____

60. What does the Bible teach on how to avoid getting HIV/Aids? **(2 marks)**

_____

_____

61. Mention two things which people laid on the road when Jesus entered Jerusalem **(2 marks)**

_____

_____

62. What is the journey of the Israelites out of Egypt called? **(1 mark)**

_____

_____

63. How many children did Jacob have? **(1 mark)**

_____

_____

64. Give one reason why Jesus performed Miracles on sick people? **(2 marks)**

_____

_____

END OF QUESTION PAPER

Appendix C: ADUS selection test for 2015 marking guide

MARKING GUIDE

SECTION A:

ENGLISH

Question One (composition)

Heading	**(1 mark)**
Introduction	**(1 mark)**
Paragraph	**(3 marks)**
Grammar	**(5 marks)**
Body layout	**(2 marks)**
Sequences flow of words	**(2 marks)**
Conclusion	**(1 mark)**

OR

LETTER WRITING

Address	**(2 marks)**
Date	**(1 mark)**
Salutation	**(1 mark)**
Grammar	**(5 marks)**
Body (i.) Layout	**(2 marks)**
(ii) Sequences flow of words	**(2 marks)**
Conclusion	**(1 mark)**
Ending	**(1 mark)**

COMPREHENSION (1 MARK FOR EACH QUESTION)

2 was a very famous nurse
3. she was born in England
4. English
5. Working without receiving salary or money
6. Improve the living conditions of the wounded soldiers

QUESTION 7 – 11 PASSIVE VOICE

7. English is taught by him
8. We shall be met by George
9. A lesson was given us by Mr. Brown
10.The room is being cleaned by Mary
11. The rat was killed by the dog.

QUESTION 12 – 16 SENTENCE CONSTRUCTION

SAMPLE ANSWERS

12. Observe: I have observed the lessons
13. Watch: They are watching football matches
    She is wearing a wrist watch
14. Expect: He is expecting good results
15. Consider: I won't consider you

16. Suggest: You should bring your suggestion boxes

QUESTIONS 17 – 21 STRUCTURE

17. C with
18. B to
19. D with
20. B from
21. D out

QUESTIONS 22 – 26 CLAUSES

22. B Adjective clause, qualifying 'Sandile"
23. A Adverb clause, modifying 'arrived'
24. C Adverb clause, modifying 'will fail'
25. B Noun clause, object of 'Said'
26. D Adjective clause, qualifying 'Bonga'

SECTION B
MATHEMATICS
QUESTION 27- 36

27. D	29. B	31. D	33. C	35. C	
28 D	30. A	32. B	34. B	36. B	

Structured 37-41

37. 10 days = K2500
    24 days = more

$$\frac{24 \times K2500}{10dys}$$

24xK250
= **K6000** **(3 marks)**

38. $\frac{2}{5} + \frac{1}{8} = \frac{16+5}{40} = \frac{21}{40}$

Fraction used = $\frac{21}{40}$

Fraction left = $\frac{40}{40} - \frac{21}{40} = \frac{19}{40}$

$\frac{19}{40}$ = K4560

$\frac{40}{40}$ = more

= $\frac{40}{40}$ x $\frac{40}{19}$ x 4560

= **K9,600** **(3 marks)**

39. 5086 + 827 – 591
= 5913 – 591
= **5322** **(3 marks)**

40. Total premium for 2 buses + premium for Toyota Car
(2xK6000x12) + (1x12xK2500)
K144000 + 30000
=K174000 (**3 marks)**

41. A. Circumference
B. Diameter
C. Radius **(3 marks)**

SECTION C
PRIMARY SCIENCE
Questions 42-46 multiple choice
42 A 43. C 44. B 45. C 46. A

Questions 47-54 structured
47. - Mulching
- Applying manure
- Practicing mixed cropping **(2 marks)**

48. Nearer to water source
Type of soil
Nearness to market or school

Type of fish to stock **(2 marks)**

49 . Mulanje, Nkhata Bay
Thyolo, Ntchisi

50 . a. Night blindness
b. Goitre **(2 marks)**

51. blow to the head
disease
neck and back injuries

52. Sieving **(1 mark)**

53. Constipation **any 2**
Haemorrhoids
Intestinal ulcer
Intestinal parasites
Diarrhea **(2 marks)**

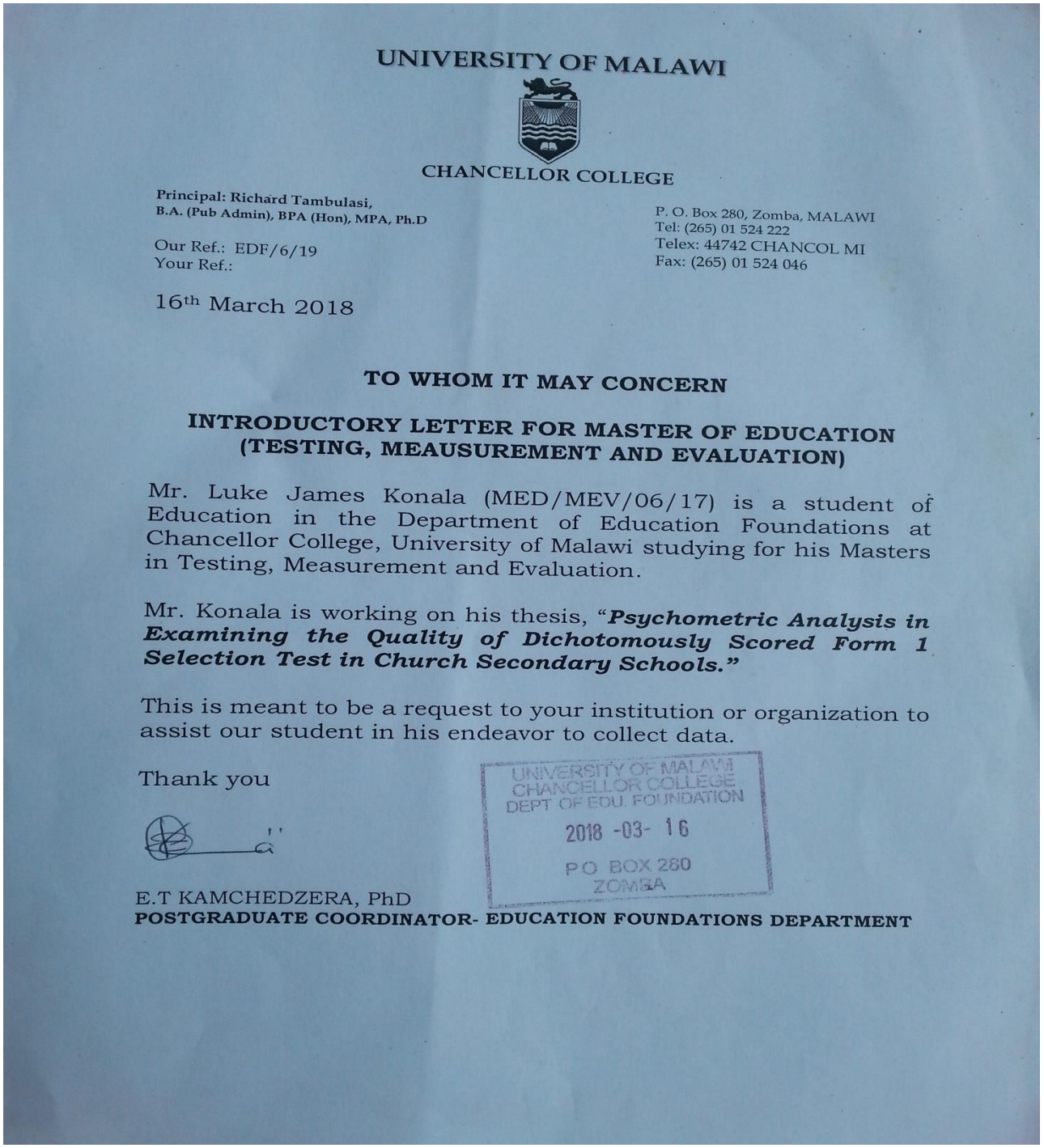
54 a. shoulder joint
b. hip joint **(2 marks)**

SECTION D
BIBLE KNOWLEDGE
Questions 55- 64 structured response

55 Sheep **(1 mark)**
56. He killed an Egyptian **(2 marks)**
57Father, son and holy spirit **(1 mark)**
58. In Galilee **(1 mark)**
59. Any two of the following:
Deborah, Gideon, Samson, Jephthah **(2 marks)**
60. By not having more than one sexual partners in marriage **(2 marks)**
61. Palm leaves and clothes **(2 marks)**
62. Exodus **(1 mark)**
63. 12 Children **(1 mark)**
64. Any one of the following:
a. He was requested to do so
b Jesus felt sorry for the sick **(2 marks)**

END OF MARKING GUIDE

Appendix D: Introductory letter for master of education research



UNIVERSITY OF MALAWI

CHANCELLOR COLLEGE

Principal: Richard Tambulasi,
B.A. (Pub Admin), BPA (Hon), MPA, Ph.D

P. O. Box 280, Zomba, MALAWI
Tel: (265) 01 524 222
Telex: 44742 CHANCOL MI
Fax: (265) 01 524 046

Our Ref.: EDF/6/19
Your Ref.:

16th March 2018

TO WHOM IT MAY CONCERN

INTRODUCTORY LETTER FOR MASTER OF EDUCATION (TESTING, MEAUSUREMENT AND EVALUATION)

Mr. Luke James Konala (MED/MEV/06/17) is a student of Education in the Department of Education Foundations at Chancellor College, University of Malawi studying for his Masters in Testing, Measurement and Evaluation.

Mr. Konala is working on his thesis, "***Psychometric Analysis in Examining the Quality of Dichotomously Scored Form 1 Selection Test in Church Secondary Schools.***"

This is meant to be a request to your institution or organization to assist our student in his endeavor to collect data.

Thank you

UNIVERSITY OF MALAWI
CHANCELLOR COLLEGE
DEPT OF EDU. FOUNDATION
2018 -03- 16
PO BOX 280
ZOMBA

E.T KAMCHEDZERA, PhD
POSTGRADUATE COORDINATOR- EDUCATION FOUNDATIONS DEPARTMENT