DEVELOPMENT AND VALIDATION OF MATHEMATICS ACHIEVEMENT TEST USING THE RASCH MODEL

ALIYU, Rasheed Taiwo
PG/10/11/191557

DEPARTMENT OF GUIDANCE AND COUNSELLING
DELTA STATE UNIVERSITY,
ABRAKA

NOVEMBER, 2015

TITLE PAGE

DEVELOPMENT AND VALIDATION OF MATHEMATICS ACHIEVEMENT TEST USING THE RASCH MODEL

ALIYU, Rasheed Taiwo
PG/10/11/191557
DIP IN COMM (U.I) 1995, ADV. TECH (LONDON) 1999, PGDE (UNAD) 2002, M.Ed. (DELSU) 2008

A Thesis Submitted to the Postgraduate School in Partial Fulfillment of the Requirements for the Award of Doctor of Philosophy (Ph.D) Degree in Measurement and Evaluation of the Delta State University, Abraka.

DEPARTMENT OF GUIDANCE AND COUNSELLING

DELTA STATE UNIVERSITY,

ABRAKA

NOVEMBER, 2015

CERTIFICATION

We the undersigned certify that this research was carried out by ALIYU, Rasheed Taiwo in the Department of Guidance and Counselling (Measurement and Evaluation Unit), Faculty of Education, Delta State University, Abraka.

----------------------------- -----------------------------

Dr J. N. Odili **Date**

(Supervisor)

-------------------------------- --------------------------------

Dr P. U. Osadebe **Date**

(Supervisor)

----------------------------------- --------------------------------

Dr P. U. Osadebe **Date**

(Head of Department)

----------------------------------- --------------------------------

Prof. E. P. Oghuvbu **Date**

(Dean, Faculty of Education)

DECLARATION

I, ALIYU, Rasheed Taiwo declare that this research was carried out by me in the Department of Guidance and Counselling (Measurement and Evaluation Unit), Faculty of Education, Delta State University, Abraka.

------------------------------ ------------------------------

Aliyu, Rasheed Taiwo **Date**

(Student)

DEDICATION

This work is dedicated to my beloved wife- Jummy Olutayo and my dearest son and daughter, Heritage and Hephzibah respectively. Also, it is dedicated to all my teachers and lecturers who made the work possible.

ACKNOWLEDGEMENTS

The researcher is mostly indebted to the Almighty and Sovereign God for the accomplishment of this research; for His protection and guidance are unspeakable. Great is thy faithfulness Oh God!

The researcher is wholeheartedly grateful to his research supervisors Dr. J. N. Odili and Dr. P. U. Osadebe for their understanding, constructive guidance and great pains they took in making this research a huge success. Thank you so much sirs.

Immense gratitude goes to Prof. C. E. Mordi, Prof. P.O. Ikoya, Prof. R. I. Okorodudu, Prof. N.E. Akpotu, Prof. E. P. Oghuvbu, Dr. (Mrs) G.O. Akpochafo, Dr. (Mrs) E. E. Ebenuwa-Okoh, Dr. James Okoro, Dr (mrs) Osakwe, Dr. S. D. Clifford, Dr. (Mrs) Asiyai, late Dr. U. L. Ezenwani, Dr. Anho and other lecturers in the Department of Guidance and Counselling, Delta State University, Abraka, for their constructive criticisms and useful suggestions in making this work a bundle of success.

Two people need my recognition. They are in persons of Prof. Zoon Naqvi and Mrs Ocheli, Eunice. The former is a Professor who assisted in analyzing my data from the United State of America (USA). She is a Rasch fan who has keen interest in any Rasch measurement issue. The latter was always there to assist when it comes to the issue of finance. I thank you both.

Special thanks also go to the researcher's coursemates; Chief Ebisine Sele Sylvester, Ochuko Caroline Alordiah and Onyehi Abaye who were very instrumental towards the successful completion of this study; most especially, Chief Ebisine Sele Sylvester who assisted me financially and materially. Also, I am grateful to sister Faith at the Post Graduate School, Mr Ibeh Ralph, Mr Obaro, Mr Godspower Okenechukwu and his wife Agama Mary Okenechukwu who assisted me during my stay in the school. My good Lord will bless you all (Amen).

I am indebted to my parents, Mr Ismail and Mrs Roseline Aliyu, for their moral, spiritual and financial support throughout the period of this study. Emubor Jeff, Kuyenum Chukumah and Alaba Hammed, thanks a lot for typing and retyping my work when it was out of season and in season, convenient and not convenient for you, God bless you too. Mum Sunshine, you are too special to me. She gave me an ipad and a projector for the smooth running of my Ph.d programme, what a real gem she is! Thanks a lot.

The researcher is indebted to his brothers, leaders and friends, such as Bro Segun Oduyale, Pastor Olajire Adeleke Martins, Pastor Ogundipe and his wife, Pastor Sodeke and his wife, Pastor Tayo and his wife, Ojo Gabriel, Ebun Ojo, Glory Adeniji, Tope Ajani, Tompson (Abraka), Emmanuel Onoriode (Abraka), Taye and Kenny Ojo toward the successful completion of this work and others too numerous to mention. Special thanks also go to the various principals, teachers and respondents who co-operated in administering the constructed instrument. Finally, I want to thank authors whose works were used in the study.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

APPENDICES

LIST OF FIGURES

ABSTRACT

This research centres on the development and validation of Mathematics Achievement Test using the Rasch model. The scales used to measure achievement test in Nigeria rely on classical test theory (CTT) approach. As a result of this, they are faced with some limitations like; poor precision, sample dependency and undue focus on aggregate scores that deny test developers the opportunity of determining how the examinees performed on a test item. These problems were addressed with the application of item response theory (IRT) of the Rasch model. Nine research questions and four hypotheses testable at 0.05 level of significance were used. The study was reviewed under both theoretical and empirical study. An instrumentation research design was adopted. The population of this study consisted of all senior secondary class III students' in Oyo and Delta States. A multistage sampling technique was used to sample out one thousand five hundred (1500) students from the population. The instrument consisted of a self developed 150 items and 25 questionnaires on social economic status (SES). The content and construct validities were examined using the table of specification and factor analysis respectively while face validity was based on some experts' judgment on the development of the items. The selected items by the experts were trimmed to 100 using the principle component analysis (PCA) and rotated component matrix (RCM). The reliability value of the items using KR_{20} was 0.85. Fit analysis of Winsteps 3.75 and t-test were used to investigate how well the Mathematics fit the Rasch model of IRT. The Rasch model addressed several measurement principles that were central to construct validity which were particularly useful for assessing testees' achievements. The result showed that 65 items not only met the Rasch model assumption of measurement construct (fitting and invariant) but also demonstrated good psychometric properties. The result showed that items and person separations indices were 13.17 and 2.93 while item and person reliability were 0.99 and 0.78 respectively. The MNSQ for both infit and outfit were 0.94 and 1.08 respectively while the ZSTD for both infit and outfit are -1.7 and +2.0 respectively which were within the acceptable range of 0.7-1.1 for MNSQ for sample > 1000 while -2.0 to +2.0 for ZSTD. The difficulty level of the items ranges between -1.95logit to 7.45logit. The study contributed to knowledge in that it provided an item bank which can allow test users to pick items that would select his ability of interest. Also, the output results were expressed in both wit and logit units. It was therefore recommended that the calibrated MAT item Bank should be used for formative evaluation before testees write their senior secondary school certificate examinations. Application of the Rasch model principles of test development and validation for Achievement Tests in Nigerian schools and Examination bodies were highly recommended.

CHAPTER ONE

INTRODUCTION

Background to the study

Measurement is crucial to all areas of education, psychology and the social sciences. Therefore, the advancement of education, psychology and science depend on measurement. Experts such as Opasina (2009), Odili (2005), Nenty (2004), Osadebe (2001) and Kerlinger & Lee (2000) have suggested that for measurement to be meaningful, the object to be measured, the numerals that will be assigned and the rules of assignment of numerals must be well defined. Developing good measures, however, can be challenging, particularly in areas where constructs are difficult to define. In a descriptive or survey research, for example, constructs may be ephemeral or ambiguous, and still must be assessed with maximum brevity. This and other measurement issues must be carefully considered as the development of theory is affected when measurement problems overwhelm the data.

For the most part, measurement models found within Item Response Theory (IRT) can provide the information needed to develop and/or assess the qualities of a desirable test (usually an achievement test). A desirable test is one that is simple and easy to use and is characterized by high quality of the information obtained which is usually reported as reliability and validity. Some tests are relatively straightforward, like some of those used in the education or physical sciences.

Development of test items presents a number of challenges as follows: (1) Finding an optimal length for a test item can be difficult. In general, test items should be shorter rather than longer - short instruments that maintain high quality are ideal (Green & Frantom, 2002). When test length is longer, it brings in item redundancy. Therefore, item redundancy adds to testees' burden and may increase item non-response, although decreasing instrument length adversely affects variance and thus impacts reliability and validity. (2) Identifying ways of effectively dealing with missing data requires, at least, a thorough understanding of processes involved in data collection. (3) Inter-item dependencies may present obstacles to developing a test. This has been evidenced in research on item order effects found for some item sequences (Converse & Presser, 1986, Sudman & Bradburn, 1982). (4) Identification of appropriate item and response scale functioning and changes in item and scale functioning across subpopulations over time are critical to the accuracy of conclusions (Green & Frantom, 2002, Golino, Gomes, Commons, & Miller, 2012 and Andrich, 2004). (5) The data collected through administration of the instrument should be capable of meeting criteria for statistical analyses. However, to overcome the stated challenges Odili (2013) suggested a number of guidelines in writing test items that measure a single latent trait or unidimensional in nature: (1) writing items from specific instructional objectives: This refers to the performance which the learner is expected to do at the end of the lesson. (2). Using illustrations and notes in the examination syllabus: This is the document prepared by an examination body that describes the structure of the examination in each subject in terms of content coverage, illustrations and notes required in each topic. (3). Writing short sentences stem: The stem of a multiple choice items is that part that conveys the stimulus task. (4). Avoiding the use of negatives and contrasting terms of quantity in stem of items. (5). Using of familiar synonyms in key position. These guidelines were strictly followed in the development and validation process of this study.

A good test items should yield invariant scores. Invariance describes the 'scope of use' properties of a good test. For example, a ruler provides scores of height in inches. The 'height' scores are invariant: regardless of the ruler used, a person's height remains constant and the ruler can be used with anyone. A ruler's use is not restricted to particular groups of people and is not biased towards men or women. A ruler which is marked wrong will always give the same (wrong) measurements. It is very reliable, but not very valid. Asking random individuals to tell the time without looking at a clock or watch is sometimes used as an example of an assessment which is valid, but not reliable. The answers will vary between individuals, but the average answer is probably close to the actual time. In many fields, such as medical research, educational testing, and psychology, there will often be a trade-off between reliability and validity. But, different item response patterns can provide interesting information about the characteristics of testees. For example, testees whose parents are farmers may have more difficulty getting items that reflect Mathematical Arithmetic such as loss and profit, discount or amount and interest than testees whose parents are business men and women who are into goods and services. This gives us information about the different testees groups. However, the failure of invariance prohibits group comparisons since the variable’s (gender, location, school type or social economic status) definition changes for the different types of testees. This is a validity issue. When test data meet the assumption of unidimensionality, the data also meet the assumption of local independent (Green & Frantom, 2002 and Hambleton, 1991).

Specific objectivity is another desirable characteristic in a test. Specific objectivity means that a person’s trait is independent of the specific set of items used to measure it. For example, it shouldn't matter which ruler is used to measure a person’s height; any ruler could be used and any one used would be independent of the person’s height. Additionally, a test with specific objectivity would not be affected by missing data. Hence, despite missing data, the test would still be useful and provide credible information of the testees. Test with specific objectivity can be tailored to any given testee, thus permitting individually administered and precluding administration of items that are not appropriate for a particular testee.

A statistic known as ‘fit’ provides an internal mechanism for identifying inappropriate responses to the items, allowing exclusion or re-assessment of persons whose responses make no sense, i.e., do not fit, according to the understanding of the construct (Green & Frantom, 2002). The basic idea that one can capitalize on is that the statistical behavior of "bad" items is fundamentally different from that of "good" items. The items have to be administered to students in order to obtain the needed statistics. This fact underscores a point of view that tests can be improved by maintaining and developing a pool of "good" items from which future tests will be drawn in part or in whole. This is particularly true for instructors who teach the same course more than once.

Once the instructor is satisfied that the test items meet the above criterion and that they are indeed appropriately written, what remains is to evaluate the extent to which they discriminate among students. The degree to which this goal is attained is the basic measure of item quality for almost all multiple-choice tests. For each item the primary indicator of its power to discriminate testees is the correlation coefficient reflecting the tendency of testees selecting the correct answer to have high scores. This coefficient is reported by typical item analysis programs as the item discrimination coefficient or, equivalently, as the point-biserial correlation between item score and total score. This coefficient should be positive, indicating that students answering correctly tend to have higher scores. Similar coefficients may be provided for the wrong choices. These should be negative, which means that students selecting these choices tend to have lower scores (Hambleton, R. K., Swaminathan, H, & Rogers, H.J. 1991).

In the teaching-learning process, testing and evaluation of the testees progress is a common place event. In fact, evaluation of the individuals' or learners' progress is a major part of the instructors' task. The task is how to evaluate the testees learning progress. One of such way of going about this is the use of specially prepared instrument, which measures all aspects of the identified instructional objectives and content domain. However, majority of the prepared instruments are in CTT therefore due to its shortcoming, the IRT method has now been introduced.

Nenty (2004) said 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. Psychometricians generally determine the validity and reliability of such tests. Therefore, in order to measure the performance of learners or testees in an academic setting, measurement instruments called tests are administered. A test can be described as a systematic procedure in which testees are presented with a set of constructed stimuli to which they respond. The responses enable the examiner to assign the testees numerals or sets of numerals from which inferences can be made about the testees' performances or possession of whatever the test is supposed to measure (Yoloye, 2004). Many examining bodies like the West African Senior School Certificate Examination (WASSCE) and National Examination Council (NECO) have some major challenges confronting them. One of such is the development of test items, which will cover the stipulated syllabus and how to objectively assess the performance of the candidates. This is because they have to develop equivalent test items for their numerous candidates. When well developed, valid and reliable tests are utilized, and the performance of the candidates objectively measured, the information generated can be used in selection of candidates for further studies.

When test developers test testees on technically valid and reliable Mathematics items, large amounts of performance data could be generated. These can be used to address a whole range of problems as well as the differences that exist across groups of urban and rural testees; different types of school: public and private; gender types: boys and girls; different social economic status: high and low and co-educational school testees, and also what characteristics are associated with improving performance in Mathematics at the senior secondary school (SSS) level. When an examination warrants a large group of students writing the same paper at the same time for which certificate are to be awarded, then, there is need for uniformity in the examination conditions which the students are exposed to in order to bring about fairness to all the participants. This brings to focus the method of assessment of test item in Mathematics. Opasina (2009) said many researchers such as Nworgu and Harbor-Peters, (1990); Ogochukwu, (1990) and Agwagah (1985), have all based their studies on students' assessment on the Classical Test Theory which is considered not valid enough for ensuring objectivity in measurement. Therefore, there is an urgent need to intervene in the process of setting test items (questions) and method of assessment of Mathematics for the SSS students' final examinations. According to Nenty (2004), when item response theory (IRT) is used in educational assessment, estimate of a given trait level, for example, a proficiency level possessed by a learner can easily be assessed. These call for a change in the theory that underlies the way tests are developed, administered, scored and analyzed.

IRT attempts to model the relationship between an unobserved variable, usually conceptualized as an examinee's proficiency and the probability of the examinee correctly responding to any particular test item. IRT rests on two basic postulates: The performance of an examinee on a test can be predicted (or explained) by a set of factors called traits, latent traits or proficiency. The relationship between examinee's item performance and the set of traits underlying item performance can be described by a Monotone Increasing Function (MIF) called an Item Characteristic Function (ICF) or Item Characteristic Curve (ICC).

Hence, there is the need for another new method of constructing, scoring, analyzing, validating and reporting students' achievement in Mathematics, which will ensure objective measurement of the testees. As a result, the researcher has developed and validated Mathematics Achievement Test using the Rasch model.

Statement of the Problem

The poor performance of students in Mathematics has been of great concern to stakeholders in education, particularly when considering the role of Mathematics in the development of science and technology. Also, admission into the Nigerian Universities is partly based on academic achievement of testees in Mathematics; therefore, good testing instruments in Mathematics (validated tests) should be used in the educational system to assess the performance or ability of students if results awarded to them in Mathematics are to be relied upon by the Nigeria universities. However, there seems to be few or no Mathematics validated tests in use by teachers in Nigerian secondary schools with Item Response Theory. This issue is of great concern to both the educational system and psychometricians in Nigeria at large.

Most developed achievement tests in Nigeria which are used for research, classroom or public examination purposes are based on Classical Test Theory (CTT). As a result of this, they are faced with some challenges like, poor precision, sample dependency and undue focus on aggregate scores that deny test developers the opportunity of determining how the examinees performed on a test item. This problem can be addressed or overcome with the application of item response theory (IRT) of the Rasch model.

Also, most classroom educators who could not undergo the rigors involved in validating and interpreting the Mathematics achievement tests resulted into using this new modern theory; this is because they are not familiar and skilled with the software for test validation. It was also noted that with their deficiency in determining the quality of a constructed test item, many who teach large classes resort into the use of multiple-choice test items of WAEC or NECO examination for the testees. These multiple-choice items constructed by both examining bodies are CTT based.

Beside teachers' inefficiency and unskilfulness in handling the software for analysis and interpretation, the non-availability of tests developed by professional test developers or psychometricians is also a major concern. Hence the method of assessment of the achievement of senior secondary Mathematics students in internal and external examination should be in Item Response theory (IRT) which could enhance objectivity in the cognitive traits of the candidates.

From the above stated problems, the researcher therefore developed and validated Mathematics Achievement Test using the Rasch model of the Item Response Theory. Hence, the statement of problem if put in a question form is: How suitable is the development and validation of a MAT (that would bring about firmness in the measurement of the achievement of examinees) able to determine testees achievement (ability) using the Rasch model?

Research Questions

This study therefore, attempted to answer the following research questions.

(1) What is the validity of the Mathematics Achievement Test (MAT) items using the Rasch model?

(2) What is the reliability of the Mathematics Achievement Test (MAT) items using the Rasch model?

(3) What are the difficulty indices of the Mathematics Achievement Test (MAT) items using the Rasch model?

(4) What items are the poor distracters of the Mathematics Achievement Test (MAT) using the Rasch model?

(5) What is the person-item-map of Mathematics Achievement Test (MAT) items using the Rasch model?

(6) What are the difficulty indices of the Mathematics Achievement Test (MAT) items for male and female testees using the Rasch model?

(7) What are the difficulty indices of the Mathematics Achievement Test (MAT) items for rural and urban testees using the Rasch model?

(8) What are the difficulty indices of the Mathematics Achievement Test (MAT) items for public and private testees using the Rasch model?

(9) What are the difficulty indices of the Mathematics Achievement Test (MAT) items for high social economic status and low social economic status testees using the Rasch model?

Hypotheses

In the light of the above stated research questions the following hypotheses were formulated and tested to check the invariance of the b-parameter value across the subgroups used in the study.

(1) There is no significant difference between the mean difficulty index (b-parameter value) of the male and female testees in Mathematics Achievement Test (MAT).

(2) There is no significant difference between the mean difficulty index (b-parameter value) of the high social economic status and low social economic status of the testees in Mathematics Achievement Test (MAT).

(3) There is no significant difference between the mean difficulty index (b-parameter value) of the public and the private schools' testees in Mathematics Achievement Test (MAT).

(4) There is no significant difference between the mean difficulty index (b-parameter value) of the urban and rural testees in Mathematics Achievement Test (MAT).

Purpose of the Study

The main purpose of the study was to develop and validate Mathematics Achievement Test using the Rasch model. As such, this study was set to achieve the following specific objectives:

i. To determine the validity of the Mathematics Achievement Test (MAT) using the Rasch model.

ii. To determine the reliability of the Mathematics Achievement Test (MAT) using the Rasch model.

iii. To determine the difficulty indices of the Mathematics Achievement Test (MAT) using the Rasch model.

iv. To examine the distracters of the Mathematics Achievement Test (MAT) items using the Rasch model.

v. To examine the person-item-map of Mathematics Achievement Test (MAT) items using the Rasch model.

vi To compare the difficulty index of the Mathematics Achievement test items for male and female testees in order to establish if there is sample invariance.

vii. To compare the difficulty index of the Mathematics Achievement test items for rural and urban testees in order to establish if there is sample invariance.

viii. To compare the difficulty index of the Mathematics Achievement test items for public and private testees in order to establish if there is sample invariance.

ix. To compare the difficulty index of the Mathematics Achievement test items for high social economic and low social economic testees in order to establish if there is sample invariance.

Significance of the Study

The study is significant to test developers, practicing teachers in the development of parallel Mathematics Achievement Test in an objective format and test scores equating due to the invariant properties of the latent trait model item parameters. The use of item fit enables test developer to choose different content areas of a given achievement test and use them to measure traits on a common content. The study provides test developers and practicing teachers with the knowledge and capacity of estimating examinees abilities, even when some items have not been attempted. It also offers researchers the opportunity to clarify measured construct

The result of the study contributes greatly to the need for the use of item banking as an objectively measuring instrument for assessing learners, thereby, freeing the hands of those Mathematics teachers who spend most of their time, duty and efforts designing, developing and administering tests and scoring learners' responses, recording and preserving the scores. The study is significant to teachers who may want to use the instrument for prognostic, diagnostic, formative and summative or end of course achievement test purpose.

The result of this study also serves as a model test to Mathematics teachers. More so, other researchers will find the result of this study useful as a guide for future research work especially those who are interested in the development and validation of achievement tests; and those who will want to use the instrument for studies on academic achievement of secondary school testees.

Scope and Delimitation of the Study

The scope of this study was to develop and validate Achievement Test multiple choice (objective) items in Mathematics which was adequate enough to measure senior secondary school (SS III) performance covering the stipulated syllabus in Oyo and Delta states. The testees were drawn from rural and urban school area (geographical location), male and female (gender), public and private schools (school type) and high and low socio-economic status (SES). These formed part of the variables for the study. The Rasch model of item response theory was used to measure and validate the testees' achievement. This means it focused on the development and validation of Mathematics Achievement Test (MAT) using the Rasch model.

This study was restricted or delimited to some secondary schools in Oyo and Delta States and one thousand five hundred (1500) testees were used for the study.

Operational Definition of Terms

For clarity and precision, the terms used in the study are explained as follows:

Items Response Theory (IRT): This is the relationship between an unobserved variable, usually conceptualized as an examinee's ability, and the probability of the examinee correctly responding to any particular test-item.

Item Parameter: This refers to the four item parameters in Item Response Theory. These include; the difficulty index (b-parameter), discrimination index (a-parameter), the guessing index (c-parameter) and the carelessness index (d-parameter).

Latent Trait/Proficiency: This refers to the innate, inherited, mental distinguishing characteristics that cause consistent performance in an individual examinee. These mental characteristics cannot be observed physically. They are referred to as latent trait that is hidden.

Scale Score: This refers to the estimate of a respondent's ability or other attributes not based on the number of correct score only but also on extrapolated parameters.

Assumption of Local Independence: This is the assumption that response to an item is independent of responses to other items in a scale after controlling for the latent trait (construct) measured by the scale.

Unidimensionality: This is a situation where several constructs exist in a set of data but only one which clearly dominates (essential unidimensionality) is investigated.

Assumption of Unidimensionality: This is the assumption that the set of items are measuring a single continuous latent variable (construct).

Theta (θ): This is the unobservable (or latent) construct being measured by the test. These constructs or traits are measured along a continuous scale.

Item Difficulty: The item difficulty (threshold) parameter-b for the 1 and 2- parameter logistic model is the point on the latent scale θ where a person has 50% or 0.5 chance of responding correctly to the scale item. Item difficulty threshold parameter b for the 3-parameter logistics model is the point on the latent scale θ where a person has a little bit above 50% (0.5) chance of responding correcting to the scale item.

Item Discrimination: The item discrimination (Slope) parameter-a_l describes the strength of an item's discrimination between examinees with trait levels θ below and above the threshold, b. The parameter, a, may always be interpreted as describing how an item may be related to the trait measured by the scale.

Guessing: This is the probability that an examinee with lowest ability would get an item correct.

Abbreviation

MAT:	Mathematics Achievement Test
CTT:	Classical Test Theory
ETS:	Educational Testing Service
ICC:	Item Characteristic Curve
ICF:	Item Characteristics Function
IRT:	Item Response Theory
TIF:	Test Information Function
IIF:	Item Information Function
MIF:	Monotone Increasing Function
SES:	Social Economy Status
FA:	Factor Analysis
PCA:	Principle Component Analysis (SPSS)
PCAR	Principle Component Analysis of the Residual (Rasch model)

CHAPTER TWO

REVIEW OF RELATED LITERATURE

This chapter focuses on the review of related literature and it is organized under the theoretical, empirical studies and summary of review of literature. They are discussed under the following sections:

i. Theoretical Framework

ii. Theories of Measurement

(a) Classical Test Theory (CTT)

(b) Item Response Theory (IRT)

i. A brief history and its Conceptual Background

ii. Item Response Theory models and its Assumptions

(c) The Rasch Model for Measurement Theory

iii. Rasch Model Analysis and Its Validity

iv. The Rasch Model Critics (Criticism of the Rasch model)

v. The Construct Validity of Test Items

vi. An Overview of the Rasch model indices

vii. (a) Students' Achievement in Mathematics

(b) Location and Students' Achievement in Mathematics

(c) Gender and Students' Achievement in Mathematics

(d) SES and Students' Achievement in Mathematics

viii. Empirical Studies Relating to the Rasch Analysis Model

ix. Appraisal of Literature Review

Theoretical Framework

The theoretical framework of this study is based on the Rasch Model and the six aspects of Messick's validity process. Baghaei & Amrahi, (2011); Beglar, (2010); Wolfe & Smith (2007), Bond (2003) and Smith (2001), have all attempted to point out how the analyses carried out within Rasch framework can be linked to the six Messick's validity states. Rasch model, named after the Danish mathematician and statistician Georg Rasch, is a prescriptive probabilistic mathematical ideal. It is highly distinguished for its two remarkable properties of invariance and interval scaling which are obtained in case the basic assumption of unidimensionality underlying the model is met, i.e. when the data fit the model.

The model is referred to as a prescriptive model because it prescribes specific conditions for the data to meet. This means that the whole research process, from the very beginning, must be in line with the model's specifications. One of the basic assumptions of the Rasch model is the unidimensionality principle: the measurement instrument must measure only one trait at a time. Though theoretically sound, practically it is almost impossible to construct a test which measures only one attribute or to prevent the interference of extraneous factors. One may unintentionally measure language proficiency in a Mathematics test which is primarily intended to measure the test takers' mathematical ability. This is usually the case with Mathematics tests including worded problems, especially when the test is administered to non-native speakers of the test language. Moreover, in almost all testing situations, a number of extraneous factors are involved which contaminate the measurement. Henning, Hudson & Turner (1985) clarifies the point: Examinee performance is confounded with many cognitive and affective test factors such as test wiseness, cognitive style, test-taking strategy, fatigue, motivation and anxiety. Thus, no test can strictly be said to measure one and only one trait. As achieving this strong version of unidimensionality is impossible, a more relaxed formulation has also been advanced (Bejar, 1983).

The unidimentionality with which the Rasch and IRT models are concerned is psychometric unidimensionality and not psychological. Thus unidimentionality within Rasch model means "a single underlying measurement dimension; loosely, a single pattern of scores in the data matrix" rather than "a single underlying (psychological) construct or trait" (MacNamara, 1996). In order for the data to meet unidimensionality condition, the response patterns should follow Guttman pattern. If items are rank ordered from easy to difficult, a person who has responded correctly to a difficult item should respond correctly to all the easier items as well. In other words, it is not expected that a person respond correctly to difficult items, but miss the easier ones or vice versa. The more the data is Guttman-like, the more it is likely to fit the Rasch model.

Having calculated the probabilities of providing correct responses to items of specific estimated difficulties by persons of particular estimated abilities, one should check whether the model's expectations realized in the form of probabilities are consistent enough with the observed data. This is done by checking the probabilities against the real observed data which can be carried out statistically as well as graphically. It should be noted that there always exists some difference between the model's predictions and the real data since the model is a perfect mathematical ideal, a condition impossible to meet in the real world. If deviation of data from the ideal set by the model is tolerable, it is said that the data fit the model, thus enabling one to benefit from the attractive properties provided by the model. If not, the remarkable properties of the model which are in fact the properties of fundamental measurement are lost. Although over forty fit indices have been developed by Psychometricians to check the accord between data and the model mainly two of them are implemented in Rasch software written in North America and Australia: infit and outfit statistics. While the former is sensitive to the unexpected patterns of response in the zones where the items are quite targeted to the person's abilities, the latter is highly sensitive to lucky guesses and careless mistakes. Both types of fit statistics are expressed in the form of mean square values as well as standardized values. The ideal value is 1 for mean square values and 0 for standardized ones. The acceptable range for mean square values is from 0.70 to 1.1 if the sample is above 1000 and for standardized ones from -2 to +2. In case the data fit the model, one can be confident that the item measures are independent of the person measures and vice versa.

Invariance of the measures can also be tested by splitting the items or persons into two halves and running independent analyses to check whether the item and person estimates remain invariant across the analyses. To be more specific, either the same test is given to two groups of people or the sample to which the test is given is divided and considered as two groups. Then, the difficulty estimates of each item, derived from two separate analyses, are correlated against each other. Also, t-test analysis can be used to test the invariance of the measures with the above principle or procedure (Andrich, 2011). The procedure here is the same for persons, but in this case of persons there are two groups of persons and one group of item. That is, two ability estimates for each item is estimated based on the two sets of persons and then the ability estimates are correlated against each other and tested using the t- test analysis.

In this part, summarizing briefly the works of Wolfe and Smith (2007), Bond (2003) and Smith (2001), the contribution that Rasch analysis can make to demonstrate different aspects of the Messick six faceth validity is pointed out. A number of analyses are performed to provide evidence for the content aspect of validity within Rasch framework. Fit indices are used to check the relevance of the test content to the intended construct. Misfitting items may be measuring a totally different and irrelevant construct. Moreover, person-item map and item strata are two important criteria for checking the representativeness of the items. Noticeable gaps in the item difficulty hierarchy point to the fact that some area of the construct domain has not been covered by the test (Baghaei, 2008). Item strata, i.e. "the number of statistically distinct regions of item difficulty that the persons have distinguished" (Smith, 2001), is another clue which is drawn upon to check representativeness. There should be at least two item difficulty levels distinguished so as to judge the items as being appropriate representatives of the intended content. Furthermore, technical quality of the test items can be assessed via fit indices as well as item-measure correlations since the former is a good indicator of multidimensionality, poor item quality or miskeying and the latter is an indicator of "the degree to which the scores on a particular item are consistent with the average score across the remaining items." (Wolfe & Smith, 2007). With regard to the expected values of the item-measure correlations, Wolfe and Smith (2007) summarize the issue as: Item-measure correlations should be positive, indicating that the scores on the item are positively correlated with the average score on the remaining items. Negative item-measure correlations typically indicate negatively polarized items that were not reverse- scored. Near zero item-measure correlations typically indicate that the item is either extremely easy or difficult to answer correctly or to endorse or that the item may not measure the construct in the same manner as the remaining items (Ahmad et al, 2012 and Wolfe & Smith, 2007).

Person fit statistics and, in the case of multiple-choice tests, multiple choice distracter analysis is considered to be important indicators of substantive aspect of validity. Person fit statistics provide empirical clues for "the extent to which a person's pattern of responses to the items correspond to that predicted by the model" (Smith, 2001). Person misfit may be due to factors like carelessness, guessing, etc. Distracter analysis within Rasch framework involves distracter p-values, choice means and distracter-measure correlations. P-values indicate "the proportion of respondents choosing each distracter" (Wolfe & Smith, 2007). Ideally, it is expected that the distracters be equally attractive; however, this seems to be almost impossible in practice. Thus, p-values are used to detect malfunctioning as well as non-functioning distracters. Choice means represent "the average measure of respondents who choose each distracter" (Wolfe & Smith, 2007, p. 209). They indicate the discrimination power of the distracters. It is expected that distracters be chosen by less able test takers, thus discriminating between test takers of high and low ability levels. As Wolfe & Smith (2007) put it," If a distracter does not attract less able respondents, then its validity as a measure of the underlying construct is questionable. Finally, distracter-measure correlations are correlations between distracters and test takers' ability measures and indicate "the degree to which each distracter is a plausible answer to the prompt" (Wolfe & Smith, 2007). Since, again, it is expected that test takers of low ability choose the distracters (rather than a correct option), thus negative values for correlations are desired. However, since the number of test takers choosing a particular distracter may be small, it is likely that the distracter measure correlations be attenuated and consequently result in correlation values which are not considerably negative. In such cases, choice means are drawn upon to compensate for the attenuation effect.

Fit statistics are used to assure whether the test is unidimensional and guide one to decide upon the way the test should be scored. That is, in case the test is shown to be unidimensional, reporting a single score for the whole test would suffice. However, in case of multidimensionality, separate scores should be reported for each dimension, and one should be cautious not to add up the scores on different dimensions. Thus, fit statistics provide helpful evidence with regard to the structural aspect of construct validity. Checking the invariance of item measures across different populations or over time, as well as checking the invariance of person measures across different sets of items can be employed to check the generalizability aspect of construct validity.

In the case of external aspect of construct validity, the extent to which the meanings of the scores of a test hold relations with some other related test results or non-test behaviors is usually checked via building Multitrait Multimethod matrices. The external aspect of validity is usually checked via monotrait and heterotrait correlations which have traditionally been referred to as convergent and discriminant evidence respectively. It is expected that monotrait correlations be higher than the heterotrait ones in order to serve as evidence for the external aspect of validity. Moreover, the capacity of a test to detect within-individual changes (over time, e.g. as a result of treatment) and between-group differences, is another indicator of the external validity. This capacity can be checked via visual inspection of person-item map as well as checking the person strata index. If a test is given to a group before a treatment and the map manifests "a floor effect, and a wide dispersion of item calibrations beyond the highest person measure" (Wolfe & Smith, 2007), the test is said to be responsive to treatment and thus capable of detecting within individual changes. The same applies to situations where the test is used to compare different groups which undergo different experimental treatments. Person strata index which represents the number of statistically separate ability strata that the test can distinguish is another evidence for external aspect of construct validity. High values for person strata (at least 2) are needed to confirm the external aspect of validity of a test.

Rasch has not explicitly put forward a way to check the consequential aspect of validity. However, issues like item bias and examination of Differential Item Functioning (DIF) or a close examination of the person-item map- which reveals the amount of information on the basis of which decisions for action are taken- can provide helpful evidence to decide about the consequential aspect of construct validity of a test (Bagheai & Amrahi, 2011).

For this current analysis, the above are applied to a multiple choice MAT to demonstrate the six Messick's construct validity. The PCAR, item misfit order or item measure(difficulty), item strata, item separation, person item maps, item measure correlations (PT measure correlation or Point Biseral), item distracters and fit statistics were measured and examined. Rasch analyses corresponding to various Messick's validity aspects are conducted to show how Rasch model is applied in practice for content, substantive, structural, generalizability, external and consequential validation.

Theories of Measurement

Classical Test Theory

Item analysis: When employing the standard test development techniques of classical test theory, item analysis consists of (a) determining sample-specific item parameters by employing simple mathematical techniques and moderate sample sizes, and (b) deleting items based on statistical criteria. Standard item analysis techniques involve an assessment of item difficulty and discrimination indices and item distracters (Nenty, 2004).

(1) Preparation of test specifications.

(2) Preparation of the test item pool.

(3) Field testing the items.

(4) Revision of the test items.

(5) Test development.

(6) Pilot testing.

(7) Final test development.

(8) Test administration (for norming and technical testing)

(9) Technical analyses (e.g., compiling norms, setting standards, equating scores, reliability and validity studies).

(10) Preparation of administrative instructions and

(11) Technical manual.

(12) Printing and distribution of tests and manuals.

Important differences in test development using classical test theory and item response theory occur at Steps 3, 5, and 9 (Golino et al, 2012 and Hambleton et al 1991).

Steps in test development

Because item statistics depend to a great extent on the characteristics of the examinee sample used in the analyses, an important concern of test developers applying classical test theory is that the examinee sample should be representative of the overall population for whom the test is intended. Heterogonous samples will, generally, result in higher estimates of item discrimination indices as measured by point-biserial or biserial correlation coefficients, whereas item difficulty estimates rise and fall with high- and low-ability groups, respectively (Beck & Gable, 2001). Despite the inherent difficulty of obtaining a representative sample, an advantage of this approach to item analysis is that item statistics can be accurately calibrated on examinee samples of modest size. For each item the primary indicator of its power to discriminate students is the correlation coefficient reflecting the tendency of students selecting the correct answer to have high scores. This coefficient is reported by typical item analysis programs as the item discrimination coefficient or, equivalently, as the point-biserial correlation between item score and total score. This coefficient should be positive, indicating that students answering correctly tend to have higher scores. Similar coefficients may be provided for the wrong choices (Hambleton et at,1991 and Hambleton, 2000). These should be negative, which means that students selecting these choices tend to have lower scores.

Detection of poor items (at least for norm-referenced tests) is quite straightforward and is basically accomplished through careful study of item statistics. A poor item is identified by an item-total score correlation (Osadebe, 2010). It is appropriate to point out that classical item analysis procedures, together with an analysis of distracters, have the potential to provide the test item developer with invaluable information concerning test item quality regardless of which measurement model is applied in the stages of test development (Osadebe, 2001 and Odili, 2005, Osadebe, 2014).

Item selection: When applying test development techniques, in addition to concerns for content validity, items are selected on the basis of two characteristics: item difficulty and item discrimination. An attempt is always made to choose items with the highest discrimination parameters. The choice level of difficulty is usually governed by the purpose of the test and the anticipated ability distribution of the group for whom the test is intended. For example, it may be the case that the purpose of a test is to select a small group of high-ability examinees for the award of a scholarship. In this situation, items are generally selected that are quite difficult for the population at large. Most norm-referenced achievement tests are commonly designed to differentiate examinees with regard to their competence in the measured areas. That is, the test is designed to yield a brand range of scores maximizing discriminations among all examinees taking the test. When a test is designed for this purpose, item are generally chosen to have a medium level and narrow range difficulty.

Item Response Theory: A Brief History and its Conceptual Background

While many think of item response theory as modern psychometric theory, the concepts and methodology of IRT have been developed for over three-quarters of the century. Thurstone (1925) laid down the conceptual foundation for IRT in his paper, entitled Method of Scaling Psychological and Education Tests." In it, he provides a technique placing the items of the Binet and Simon (1905) test of children's mental development on an age-graded scale (Opasina, 2009).

Thurstone dropped his work in measurement to pursue the development of multiple factor analysis, but his colleagues and students continued to refine the theoretical bases IRT (Steinberg & Thissen, 1995). Opasina (2009) said Ferguson (1943) and Richardson (1936) introduced normal ogive model as a means to display the proportions correct for individual item as a function of normalized scores.

Three persons deserve special recognition as regards their various contributions to development of IRT. They are Lawley (1943), Lazarsfield (1950) and Lord (1952). Lawley (1943) extended the statistical analysis of the properties of the normal ogive and described maximum likelihood estimation procedures for the item parameter and linear approximations to those estimates. Lord (1952) of the Educational Testing Services (ETS) may be regarded as the father of IRT. In his Ph.D thesis (1952), he presented IRT as a model in its own right. He introduced the idea of a latent trait or ability and differentiated this construct from observed test score. Lazarsfield (1950) described the unobserved variable as accounting for the observed interrelationships among the item responses. Graphs of the proportions of children in successive age cross-sections succeeding on successive Binet tasks and the effective location of each item on chronological age reflect marks of the features suggestive of IRT (Opasina, 2009).

In 1960, George Rasch published his one parameter sample free IRT model. This research stirred up much interest and 'considerable studies were undertaken. The leading proponent was Benjamin Wright of the University of Chicago: The Rasch model assumes a single underlining ability usually a continuous, unbiased variable designated as the ability of examinees but varying in the characteristics they ascribe to items (Harris, 1989) In 1968, Lord and Novick in their book Statistical Theories of Mental Test scores looked into the assumption underlying the application and usage of IRT. A substantial part of Lord and Novick's chapters on IRT were by Allan Birnhaum (1968) who proposed the two parameter model for IRT. The effort of Lord and Novick was further complemented by Urry (1974) in his Ph.D dissertation, where he compares the various models of IRT. Other publications related to IRT include that of Samejima (1969, 1972), Wright & Stone(1979), and Wright & Panchapakesan (1969).

Considered a milestone in psychometrics, the textbook written by Embretson & Reise (2000), entitled, *"Statistical Theories of Mental Test Scores"* provides a rigorous and unified statistical treatment of classical test theory. The remaining half of the book, written by Allen Bimbaum, provides an equally solid description of the IRT models. Bock and several students collaborators at the University of Chicago, including David Thissen, Eiji Muraki, Richard Gibbons and Robert Mislevy developed effective estimation methods and computer programs such as Bilog, Multilog, Parscale and Testfact. Along with Aitken, Bock developed the algorithm of marginal maximum likelihood method to estimate the item parameters that are used in many of these IRT programs (Bock & Aitken, 1981).

In a separate line of development of IRT models, George Rasch (1960) discussed the need for creating statistical models that maintain the property of specific objectivity, the idea that people and item parameters be estimated separately, but comparable on a similar metric. Rasch inspired Gerhard 'Fischer (1968) to extend the applicability of the Rasch models into psychological measurement and Ben Wright to teach these methods and help to inspire other students in the development of the Rasch models. These students, including David Andrich, Geoffrey Masters, Graham Douglas, and Mark Wilson, helped to push the methodology into education and behavioral medicine (Wright, 1997). Since there will always be improvement in every sphere of life, psychometricians will continue to bring about new innovations into both education and behavioural medicine.

Its Conceptual

Item response theory (IRT) offers a different way to construct tests. The characteristics of individual test items are at the heart of the theory (Nenty, 2005). IRT has a peculiar way in which the test item are viewed and used. The proportion of individuals getting a valid item correct is correlated with ability (θ). Tests developed using IRT raise additional validation issues, such as the fact that items are unidimensional and locally independent. These results frequently come from factor-analytic studies. When individual test items are negatively correlated or uncorrelated with the ability assessed by the test, which is usually estimated by the total score on the test, the items ought to be deleted. Deletion of such items ensures the development of a homogenous test.

Fundamentally, the most important product of education is the learner's score, which is generated to represent learning character or trait and other abilities which underlie individual's capability to perform in life, in work situations and hence the enablement to contribute to development in the society (Nenty, 2005). According to Hashway (1998), the score obtained from an assessment is an estimate of the degree of success an individual would experience if confronted with a representative event from the content domain reflected by the events composing that assessment. He further states that any educational process is effective to the extent that it has been able to, during such process, improve on and hence maximize such score for each individual learner. The assumption is that every human behaviour be it cognitive, affective or psychomotor can be validly assessed, quantified, qualified or graded. The score, in its several forms, operationalises cognitive and non-cognitive behaviour including performance, achievement and ability. All that those involved in education are doing is to find out and implement all possible means of maximizing the score for every learner. A score has meaning only to the extent that it is valid, that is to the extent that it defines the behaviour and only the behaviour it represents. Education is of quality to the extent that it is able to maximize such score or grade for every learner in the society.

Improving on and hence maximizing this all important score is not obvious unless the educational measurement process, like the physical measurement process, could put the initial and all subsequent scores by a leather on a particular behaviour or ability on a single continuum based on which changes could be determined, compared and evaluated (Nenty,2005). Education enhances or desirably changes a learner's ability or behaviour, and this is reflected through enhanced performance and achievement. Learning is said to be operationalised by gains or desirable changes in the behaviour of the learner, hence education is concerned with maximizing desirable changes in human behaviour. Changes in behaviour can only be determined through comparison between what was before and what is now, and this call for an objective scale with invariant properties. In order words, education as defined is amenable more to self-reference measurement than to the commonly used norm and criterion-referenced measurement.

The score or grade if validly generated represents the extent to which learning has taken place. That is, the level to which education, whose task it is to ensure learning, has done its job. Therefore, the size and quality of the score represent the level of success or failure in the task of equating. It is this score in its one form of quantitative or qualitative transformation or the other that provides the most important input into every form of decision making from the classroom to the national level. These are specifically with regards to feedback to learners to ensure improvement in subsequent performance and enhance learning; learners promotion or advancement, certification or graduation; selection for and promotion in employment; as well as input into educational policy decisions (Nenty, 2005). The score in several of its analyzed or transformed form is the basis of every educational decision making.

IRT is a model for expressing the association between an individual's response to an item and the underlying latent variable (often called "ability" or "trait" or proficiency) being measured by the instrument. The latent variable, expressed as theta (θ), is a continuous unidimensional construct. That explains the covariance among item responses (Steinberg & Thissen, 1995). People at higher levels of theta (θ) have a higher probability of responding correctly or endorsing an item.

IRT models use item responses to obtain scaled estimates of θ, as well as to calibrate items and examine their properties (Mellenbergh, 1994). Each item is characterized by one or more model parameters. The item difficulty or threshold, (parameter b is the point on the latent scale (θ), where a person has a 50% chance of responding positively to the scale item (question). Items with high threshold are less often endorsed (Steinberg & Thissen, 1995). The slope or discrimination, (parameter-a) describes the strength of an item's discrimination between people with trait levels theta (θ) below and above the threshold b. The a parameter may also be interpreted as describing how an item may be related to the trait measured by the scale and is directly related, under the assumption of a normal θ distribution, to the biserial item test correlation (Linden & Hambleton, 1997).

The slope parameter is linearly related (under some conditions) to the variable loading in a factor analysis. Some IRT models, in education research, include a lower-asymptote parameter or guessing parameter c which possibly explains why people of low ability level of the trait are responding positively to an item.

Another important feature of IRT models is the information function, in index indicating the range of trait level θ over which an item or test is most useful for distinguishing among individuals. In other words, the information function characterizes the precision of measurement for persons at different levels of the underlying latent construct, with higher information denoting more precision. Graphs of the information function place persons' trait level on the horizontal x-axis, and amount of information on the vertical y-axis.

The shape of the item information function is dependent on the item parameters. The higher the item's discrimination, the more peaked the information function will be; thus, higher discrimination parameters provide more information about individuals whose trait levels θ lie near the item's threshold value. The item's difficulty parameter(s) determines where the item information function is located (Opasina 2009, Flannery, Reise, & Widaman, 1995). With assumption of local independence reviewed, the item information values can be summed across all of the items in the scale to form the test information curve (Lord, 1980).

At each level of the underlying trait (θ), the information function is approximately equal to the expected value of the inverse of the square-root of the standard errors of the θ estimates (Lord, 1980). The smaller the standard error of measurement (SEM) the more information or precision the scale provides about (θ). For example, if a measure has a test information value of 16 at θ=2.0, then examinee scores at this trait level have a standard error of measurement of $1/\sqrt{16} = 0.25$, indicating good precision (reliability is approximately 0.94) at the level of theta (θ) (Flannery, Reise & Widaman, 1995).

Scale scoring in item response theory (IRT) has a major advantage over classical test theory (CTT). In classical test theory (CTT), the summed scale score is dependent on the difficulty of the items used in the selected scale, and therefore, not an accurate measure of a person's trait level. The procedure assumes that equal ratings on each item of the scale represent equal levels of the underlying trait (Cooke & Michie, 1997). Item response theory, on the other hand, estimates individual latent trait level scores based on all the information in a participant's response pattern. That is, IRT takes into consideration, which items were answered correctly (positively and which ones were answered incorrectly, and utilizes the difficulty and discrimination parameters of the items when estimating trait levels (Weiss, 1995). Persons with the same summed score but different response patterns may have different IRT estimated latent scores. One person may answer more of the highly discriminating and difficult items and receive a higher latent score than one who answers the same number of items with low discrimination or difficulty. IRT trait level estimation uses the item response curves associated with the individual's response pattern. A statistical procedure, such as maximum likelihood estimation, finds the maximum of a likelihood function created from the product of the population distribution with the individual's trace curves associated with each item's right or wrong response. The IRT models always focus on the measurement of change in trait level which indicates the level of the positive behavioural change that has taken place and is therefore very useful in an academic setting.

Item Response Theory (IRT)

Item analysis: When employing item response theory, item analysis consists of: (a) determining sample-invariant item parameters using relatively complex mathematical techniques and large sample sizes, and (b) utilizing goodness-of-fit criteria to detect items that do not fit the specified response model (Hambleton et al 1991). The property of sample invariance inherent within IRT means that test developers do not need a representative sample of the examinee population to calibrate test items. They do, however, need a heterogeneous and large examinee sample to insure proper item parameter estimation. As can be seen, even when examinee samples differ, the test developer is able to use the principles of IRT to estimate the same ICC regardless of the examinee samples used in the item calibration process. However, the test developer using IRT is faced with a different problem. Because IRT requires larger sample sizes to obtain good item parameter estimates, the test developer must ensure that the examinee sample is of sufficient size to guarantee accurate item calibration. *A poorly fitting IRT model will not yield invariant item and ability parameter (Hambleton, etal, 1991). Also,when data meet the assumption of unidimensionality, the data also meet the assumption of local independent (Hambleton, 2000).*

The detection of poor items using response theory is not as straightforward as when classical test theory is used (Aliyu & Ocheli, 2012). Items are generally evaluated in terms of their goodness-of-fit to a model using a statistical test or an analysis of residuals. It is important to emphasize that an adequate fit of model-to-data is essential for successful item analysis; otherwise, items may appear poor as an artifact of poor model fit. Poor items are usually identified through a consideration of their discrimination indices (the value of a_i will be a low positive or even negative) and difficulty indices (items should be neither too easy nor too difficult for the group of examinees to be assessed) (Hambleton et al,1991 and Osadebe, 2001).

Item Selection: As is the case with classical test theory, item response theory also bases item selection on the intended purpose of the test. However, the final selection of items will depend on the information they contribute to the overall information supplied by the test. A particularly useful feature of the item information functions used in IRT test development is that they permit the test developer to determine the contribution of each test item to the test information function independently of other items in the pool. Lord (1977) outlined a procedure, originally conceptualized by Birnbaum (1968), for the use of item information functions in the test building process. Basically, this procedure entails that a test developer take the following four steps:

(1) Describe the shape of the desired test information function over the desired range of abilities. Lord (1977) calls this the target information function.

(2) Select items with item information functions that will fill up the hard-to-fill areas under the target information function.

(3) After each item is added to the test, calculate the test information function for the selected test items.

(4) Continue selecting test items until the test information function approximates the target information function to a satisfactory degree.

Item Response Theory Models and its Assumptions

Broadly speaking, IRT models can be divided into two families: unidimensional and multidimensional. Unidimensional models require a single trait (ability) dimension θ. Multidimensional IRT models model response data hypothesized to arise from multiple traits. However, because of the greatly increased complexity, the majority of IRT research and applications utilize a unidimensional model.

IRT models can also be categorized based on the number of scored responses. The typical multiple choice item is *dichotomous*; even though there may be four or five options, it is still scored only as correct/incorrect (right/wrong). Another class of models apply to *polytomous* outcomes, where each response has a different score value (Andrich, 2004). A common example of this Likert-type items, e.g., "Rate on a scale of 1 to 5."

Thissen & Orlando (2001) discusses two approaches to model building in item response theory. One approach is to develop a well-fitting model to reflect the item response data by parameterizing the ability or trait of interest as well as the properties of the items. The goal of this approach is item analysis. The model should reflect the properties of the item response data sufficiently and accurately, so that the behavior of the item is summarized by the item parameters. The philosophy is that the items are assumed to measure as they do, not as they should (Thissen & Orlando, 2001).

Another approach of IRT model building is to obtain specific measurement properties defined by the model to which the item response data must fit. If the item or a person does not fit within the measurement properties of the IRT model, assessed by analysis of residuals (i.e., item and person fit statistics), the item or person is discarded. This approach follows that of the Rasch (1960) model, and in the cases where the data fits the model, offers a simple interpretation for item analysis and scale scoring. This approach to model building believes optimal measurement is defined mathematically, and then the class of item response models that yield such measurement is derived.

The two approaches described above yield a division in psychometrics. Those who believe that educational research measurement should be about describing the behaviors behind the response patterns in a survey will use the most appropriate IRT model (e.g Rasch or One-Parameter Logistic Model, Two-Parameter Logistic Model, Graded Model and Three-Parameter Logistics Model) to fit the data. The choice of the IRT model is data dependent. Researchers from the Rasch tradition believe that the only appropriate models to use are the Rasch family of models, which retain strong mathematical properties such as specific objectivity (person parameters and item parameters estimated separately) and summed score simple sufficiency (no information from the response pattern is needed). Several advantages of the Rasch model include the ability of the model to produce more stable estimates of person and item properties when there are a small number of respondents, when extremely non-representative samples are used, and when the population distribution over the underlying trait is heavily skewed.

Embretson & Reise (2000) suggest one should use the Rasch family of models when each item carries equal weight (i.e., each item is equally important) in defining the underlying variable, and when strong measurement model properties (i.e., specific objective simple sufficiency) are desired. If one desires fitting an IRT model to existing data or desires highly accurate parameter estimates, then a more complex model such as the Two-Parameter Logistic Model or Graded Model should be used.

IRT is generally regarded as an improvement over classical test theory (CTT). For tasks that can be accomplished using CTT, IRT generally brings greater flexibility and provides more sophisticated information. Some applications, such as computerized adaptive testing are enabled by IRT and cannot reasonably be performed using only classical test theory. Another advantage of IRT over CTT is that the more sophisticated information IRT provides allows a researcher to improve the reliability of an assessment.

The IRT model is based on the assumption that the items are measuring a single continuous latent variable θ ranging from -∞ to +∞. IRT entails three assumptions:

1. A unidimensional trait denoted by θ;
2. Local independence of items;
3. The response of a person to an item can be modeled by a mathematical *item response function* (IRF).

The trait is further assumed to be measurable on a scale (the mere existence of a test assumes this), typically set to a standard scale with a mean of 0.0 and a standard deviation of 1.0. 'Local independence' means that items are not related except for the fact that they measure the same trait, which equivalent to the assumption of unidimensionality, but presented separately because multidimensionality can be caused by other issues. The topic of dimensionality is often investigated with factor analysis, while the IRF is the basic building block of IRT.

The unidimensionality of a scale can be evaluated by performing an item-level factor analysis, designed to evaluate the factor structure underlying the observed co-variation among item responses. The assumption can be examined by comparing the ratio of the first to the second eigenvalue for each scaled matrix of tetrachoric correlations. This ratio is an index of the strength of the first dimension of the data. Similarly, another indication of unidimensionality is that the first factor accounts for a substantial proportion of the matrix variance (Lord, 1980; Reise & Wailer, 1990).

In the IRT model, the item responses are assumed to be independent of one another: the assumption of local independence. The only relationship among the items is explained by the conditional relationship with the latent variable (θ). In other words, local independence means that if the trait level is held constant, there should be no association among the item responses (Thissen & Steinberg, 1988). Violation of this assumption may result in parameter estimates that are different from what they would be if the data were locally independent; thus, selecting items for scale construction based on these estimates may lead to erroneous decisions (Chen & Thissen, 1997). The assumptions of unidimensionality and local independence are related in that items found to be locally dependent will appear as a separate dimension in a factor analysis. For some IRT models, the latent variable (not the data response distribution) is assumed to be normally distributed within the population. Without this assumption, estimates of θ for some response patterns (e.g., respondents who do not endorse any of the scale items) have no finite values resulting in unstable parameter estimates (Chen & Thissen, 1997).

The item response function

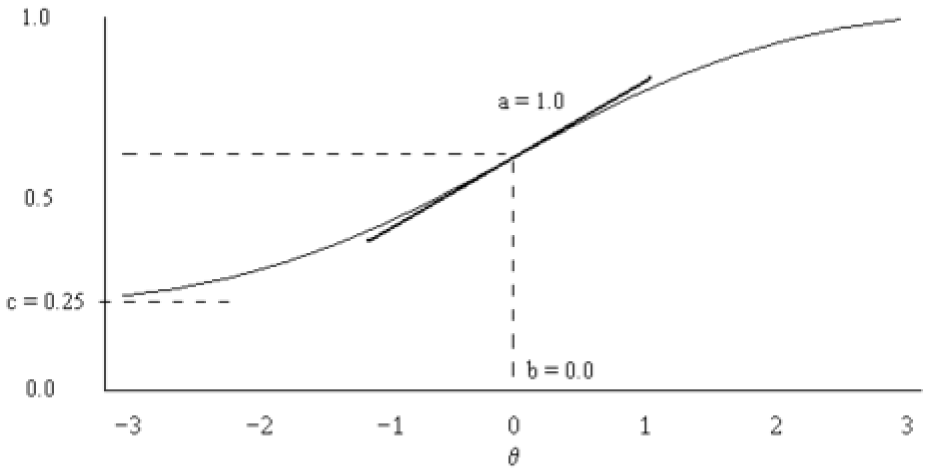


Fig.1: IRF

The IRF gives the probability that a person with a given ability level will answer correctly. Persons with lower ability have less of a chance, while persons with high ability are very likely to answer correctly; for example, students with higher math ability are more likely to get a math item correct. The exact value of the probability depends, in addition to ability, on a set of *item parameters* for the IRF. For example, in the *three parameter logistic* (3PL) model, the probability of a correct response to an item *i* is:

$$p_i(\theta) = c_i + \frac{1 - c_i}{1 + e^{-a_i(\theta - b_i)}}$$

Where θ is the person (ability) parameter and a_i, b_i, and c_i are the item parameters.

The item parameters simply determine the shape of the IRF and in some cases have a direct interpretation. The figure to the right depicts an example of the 3PL model of the ICC with an overlaid conceptual explanation of the parameters. The parameter b_i represents the item location which, in the case of attainment testing, is referred to as the item difficulty. It is point on θ where the IRF has its maximum slope. The example item is of medium difficulty, since b_i=0.0, which is near the center of the distribution. Note that this model scales the item's difficulty and the person's trait onto the same continuum. Thus, it is valid to talk about an item being about as hard as Person A's trait level or of a person's trait level being about the same as Item Y's difficulty, in the sense that successful performance of the task involved with an item reflects a specific level of ability.

The item parameter a_i represents the discrimination of the item: that is, the degree to which the item discriminates between persons in different regions on the latent continuum. This parameter characterizes the slope of the IRF where the slope is at its maximum. The example item has a_i=1.0, which discriminates fairly well; persons with low ability do indeed have a much smaller chance of correctly responding than persons of higher ability.

For items such as multiple choice items, the parameter c_i is used in attempt to account for the effects of guessing on the probability of a correct response. It indicates the probability that very low ability individuals will get this item correct by chance, mathematically represented as a lower asymptote. A four-option multiple choice item might have an IRF like the example item; there is a 1/4 chance of an extremely low ability candidate guessing the correct answer, so the c_i would be approximately 0.25. This approach assumes that all options are equally plausible, because if one option made no sense, even the lowest ability person would be able to discard it. So IRT parameter estimation methods take this into account and estimate a c_i based on the observed data (Bock & Aitkin, 1981).

IRT item parameters are not dependent on the sample used to generate the parameters, and are assumed to be invariant (within a linear transformation) across divergent groups within a research population and across populations. IRT models measure scale precision across the underlying latent variable being measured by the instrument (Cooke & Michie, 1997; Hays, Morales & Reise, 2000). An IRT-estimated person's trait level is independent of the questions being used. This is because the expected scale score is computed from their responses to each item (that is characterized by a set of properties), the IRT estimated score is sensitive to differences among individual response patterns and is a better estimate of the individual's true level on the trait continuum (Santor & Ramsay 1998).

According to Lord & Novick (1968) in Opasina (2009), latent traits are psychological dimensions necessary for the description of individuals. This means that latent trait is a theoretical attribute or characteristic(s) that independently and completely defines test performance. This indicates that latent trait represents ability measured by a test. Cressic (1983) states that the development of latent trait models begins with the assumptions that the probability that a person will answer an item correctly is the product of an ability parameter pertaining to the person and the difficulty parameter, pertaining only to the item. However, the ability of the examinee is not directly observable. It is referred to as latent trait or latent variables, which are not static. It refers simply to characteristics of examinees, which are not directly measurable or observable but are assumed to determine test performance and are referred generically to as ability (θ), which the test item is attempting to measure (Guiton & Ironson, 1983).

IRT offers the privilege for one to be able to answer any question on an item, a test or even an examinee. The scaling procedure assumes that the properties of the natural variable can be arranged along a continuum based on the magnitude of the trait possessed by a person in each property. If a test is administered to the persons in one of the sets, this model assumes that all persons have the same probability correctly on endorsing the item. This theory assumes that an individual's behaviour can be accounted for, to a substantial degree by defining certain human characteristics called traits, quantitatively estimating the individual's standing and the use of the numerical valued obtained to predict or explain performance in relevant situation (Lord & Novick, 1968).

IRT is an attempt to model the relationship between an unobservable variable referred to as the examinee's ability or trait, and the probability of the examinee correctly responding to any particular test item (Lord, 1980) since the traits are not directly measurable and unobservable, they are called latent traits. IRT makes strong assumptions about a person's behaviour when responding to test. One further note is that item response theory is the only one step towards the goal of the creation of reliable and valid measures. Hambleton (2000) says that IRT is the solution to all our instrument and measurement problems. It is a mathematical model which when used can demonstrate that one model fits the data of interest and model parameter are properly estimated. Item response theory was developed within the frame work of educational testing and so most of the literature and terminology are oriented towards discipline (Hambleton & Swaminathan, 1985). Some researchers (Zickar & Drasgow, 1996) also argued that while ability testing such as verbal or mathematical skill in educational measurement can be conceptualized, it is clear that a latent trait (proficiency) of individuals can be modelled in the same manner. The increasing need for psychometrically-sound measures calls for better analytical tools. In order to infer what the students know or can do or have accomplished more generally, item response theory is an appropriate tool to obtain valuable information.

The Rasch Model for Measurement Theory

In the Rasch model, the probability of a specified response (e.g. right/wrong answer) is modeled as a function of person and item parameters. Specifically, in the simple Rasch model, the probability of a correct response is modeled as a logistic function of the difference between the person and item parameter. In most contexts, the parameters of the model pertain to the level of a quantitative trait possessed by a person or item. For example, in educational tests, item parameters pertain to the difficulty of items while person parameters pertain to the ability or attainment level of people who are assessed. The higher a person's ability relative to the difficulty of an item, the higher the probability of a correct response on that item. When a person's location on the latent trait is equal to the difficulty of the item, there is by definition a 0.5 probability of a correct response in the Rasch model. As Hambleton (1995) makes clear, the Rasch model can be regarded as a special case of the three-parameter model when the discrimination parameters are held equal, and when the ‘guessing’ parameter is fixed at zero.

The purpose of applying the model is to obtain measurements from categorical response data. Estimation methods are used to obtain estimates from matrices of response data based on the model (Linacre & Wright, 1999). A Rasch model is a *model* in one sense in that it represents the structure which data should exhibit in order to obtain measurements from the data; i.e. it provides a criterion for successful measurement. Beyond data, Rasch's equations model relationships we expect to obtain in the real world. For instance, education is intended to prepare children for the entire range of challenges they will face in life, and not just those that appear in textbooks or on tests. By requiring measures to remain the same (invariant) across different tests measuring the same thing, Rasch models make it possible to test the hypothesis that the particular challenges posed in a curriculum and on a test coherently represent the infinite population of all possible challenges in that domain. A Rasch model is therefore a model in the sense of an *ideal* or standard that provides a heuristic fiction serving as a useful organizing principle even when it is never actually observed in practice.

The perspective or paradigm underpinning the Rasch model is distinctly different from the perspective underpinning statistical modelling. Models are most often used with the intention of describing a set of data. Parameters are modified and accepted or rejected based on how well they fit the data. In contrast, when the Rasch model is employed, the objective is to obtain data which fit the model (Andrich, 2004; Linacre & Wright, 1999). The rationale for this perspective is that the Rasch model embodies requirements which must be met in order to obtain measurement, in the sense that measurement is generally understood in the physical sciences.

A useful analogy for understanding this rationale is to consider objects measured on a weighing scale. Suppose the weight of an object A is measured as being substantially greater than the weight of an object B on one occasion, then immediately afterward the weight of object B is measured as being substantially greater than the weight of object A. A property we require of measurements is that the resulting comparison between objects should be the same, or invariant, irrespective of other factors. This key requirement is embodied within the formal structure of the Rasch model. Consequently, the Rasch model is not altered to suit data. Instead, the method of assessment should be changed so that this requirement is met, in the same way that a weighing scale should be rectified if it gives different comparisons between objects upon separate measurements of the objects.

Data analysed using the model are usually responses to conventional items on tests, such as educational tests with right/wrong answers. However, the model is a general one, and can be applied wherever discrete data are obtained with the intention of measuring a quantitative attribute or trait.

Scaling

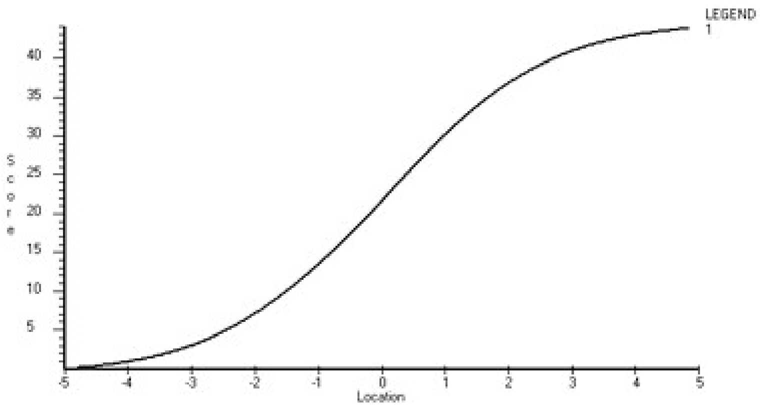


Figure 2: Test characteristic curve (TCC) showing the relationship between total score on a test and person location estimate

When all test-takers have an opportunity to attempt all items on a single test, each total score on the test maps to a unique estimate of ability and the greater the total, the greater the ability estimate. Total scores do not have a linear relationship with ability estimates. Rather, the relationship is non-linear as shown in Figure 2. The total score is shown on the vertical axis, while the corresponding person location estimate is shown on the horizontal axis. For the particular test on which the test characteristic curve (TCC) shown in Figure 2 is based, the relationship is approximately linear throughout the range of total scores from about 10 to 33. The shape of the TCC is generally somewhat sigmoid as in this example. However, the precise relationship between total scores and person location estimates depends on the distribution of items on the test. The TCC is steeper in ranges on the continuum in which there are a number of items, such as in the range on either side of 0 in Figures 2 and 3.

In applying the Rasch model, item locations are often scaled first, based on methods such as those described below. This part of the process of scaling is often referred to as item *calibration*. In educational tests, the smaller the proportion of correct responses, the higher the difficulty of an item and hence the higher the item's scale location. Once item locations are scaled, the person locations are measured on the scale. As a result, person and item locations are estimated on a single scale as shown in Figure 3.

Interpreting scale locations

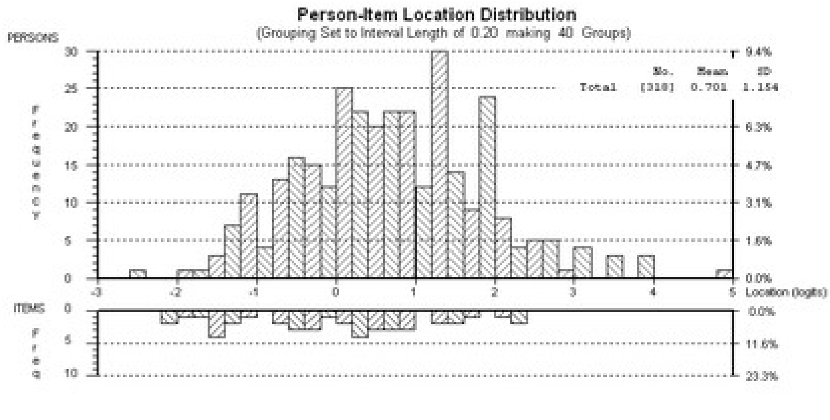


Figure 3: Graph showing histograms of person distribution (top) and item distribution (bottom) on a scale

For dichotomous data such as right/wrong answers, by definition, the location of an item on a scale corresponds with the person location at which there is a 0.5 probability of a correct response to the question. In general, the probability of a person responding correctly to a question with difficulty lower than that person's location is greater than 0.5, while the probability of responding correctly to a question with difficulty greater than the person's location is less than 0.5. The Item Characteristic Curve (ICC) or Item Response Function (IRF) shows the probability of a correct response as a function of the ability of persons.

When responses of a person are listed according to item difficulty, from lowest to highest, the most likely pattern is a Guttman pattern or vector; i.e. {1,1,...,1,0,0,0,...,0}. However, while this pattern is the most probable given the structure of the Rasch model, the model requires only probabilistic Guttman response patterns; that is, patterns which tend toward the Guttman pattern. It is unusual for responses to conform strictly to the pattern because there are many possible patterns. It is unnecessary for responses to conform strictly to the pattern in order for data to fit the Rasch model.

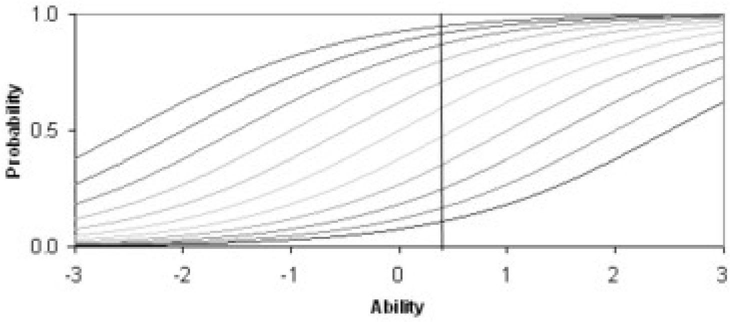


Figure 4: ICCs for a number of items. ICCs are coloured to highlight the change in the probability of a successful response for a person with ability location at the vertical line. The person is likely to respond correctly to the easiest items (with locations to the left and higher curves) and unlikely to respond correctly to difficult items (locations to the right and lower curves).

Each ability estimate has an associated standard error of measurement, which quantifies the degree of uncertainty associated with the ability estimate. Item estimates also have standard errors. Generally, the standard errors of item estimates are considerably smaller than the standard errors of person estimates because there are usually more response data for an item than for a person. That is, the number of people attempting a given item is usually greater than the number of items attempted by a given person. Standard errors of person estimates are smaller where the slope of the ICC is steeper, which is generally through the middle range of scores on a test. Thus, there is greater precision in this range since the steeper the slope, the greater the distinction between any two points on the line.

Statistical and graphical tests are used to evaluate the correspondence of data with the model. Certain tests are global, while others focus on specific items or people. Certain tests of fit provide information about which items can be used to increase the reliability of a test by omitting or correcting problems with poor items. In Rasch Measurement the person separation index is used instead of reliability indices. However, the person separation index is analogous to a reliability index. The separation index is a summary of the genuine separation as a ratio to separation including measurement error. As mentioned earlier, the level of measurement error is not uniform across the range of a test, but is generally larger for more extreme scores (low and high).

Rasch Model Analysis and Its Validity of Test Items

The theoretical framework of this study is based on the Rasch Model and the six aspects of Messick's validity process. These six aspects of Messick's validity have all been linked to a variety of outputs produced by Rasch-based analyses, in providing validity evidence for tests (Baghaei & Amrahi, 2011; Beglar, 2010; Wolfe & Smith, 2007; Boone & Scantlebury, 2006; Bond, 2003; Smith, 2001).

For the content aspect, the fit statistics are used to check the relevance of the intended test construct. These also highlight any misfitting items that are possibly representing a different construct. Person-item maps (graphical representations of the difficulty of all items and the ability of all test-takers) allow for verification of the representativeness of test content since gaps indicate that some domain of the construct has not been assessed (Baghaei, 2008). This can also be examined with item strata, which identify statistically distinct levels of difficulty and ability. Examining item-strata is generally used for the purposes of ensuring that a range of item-difficulties have been included (Smith, 2001). The technical quality aspect of content validity can be examined with item-measure correlations, which indicate how strongly the item is measuring the direction of the construct. Specifically, this measure identifies any items causing high ability students to respond incorrectly when low ability students are responding correctly (Boone & Scantlebury, 2006).

Fit statistics can also be used to provide evidence of the substantive aspect since they examine how a test-taker's response patterns match those predicted by the model (thus representing to some extent, the degree to which test-takers are engaged with the item). The fit statistics can also guide scoring of the test – for instance, they may indicate that reporting a single score is insufficient and that separate scores should be reported. This kind of information has implications towards the structural aspect of construct validity. Multiple choice Item (MCI) distractor analyses, which ensure that the distractors are indeed distracting test-takers in a meaningful way, provide additional arguments towards the structural aspect (Wolfe & Smith, 2007).

In terms of generalizability, the item measures for different populations, or person measures over time can be checked. For the consequential aspect of validity, Rasch does not provide any explicit calculation as evidence (Baghaei & Amrahi, 2011), however, acceptable thresholds for fit statistics can either be made more stringent or more relaxed depending on the stakes of the test. This addresses consequential validity by ensuring more rigorous conditions for the acceptance of the results (Linacre, 2007). For the external aspect of construct validity, other test results can be correlated (as the simplest check). Messick (1989) notes that analysis of group differences or the effects of experimental treatment are two additional ways. If a test produces different results prior to and following experimental treatment then it may be said that the test has external validity.

Furthermore, an externally valid test can differentiate between test-takers who possess the construct and those who do not, or at least, those who have achieved varying levels of the construct (Baghaei & Amrahi, 2011). Other checks such as differential item functioning (DIF) or differential test functioning (DTF) (where different items, or the test overall, do not function consistently for all test takers) have also been used as checks for external validity (Bagheai & Amrahi, 2011).

Analyzing data according to the Rasch model, that is, conducting a Rasch analysis, gives a range of details for checking whether or not adding the scores is justified in the data. This is called the test of fit between the data and the model. If the invariance of responses across different groups of people does not hold, then taking the total score to characterize a person is not justified. Of course, data never fit the model perfectly, and it is important to consider the fit of data to the model with respect to the uses to be made of the total scores. If the data do fit the model adequately for the purpose, then the Rasch analysis also linearises the total score, which is bounded by θ and the maximum score on the items, into measurements. The linearised value is the location of the person on the unidimensional continuum - the value is called a parameter in the model and there can be only one number in a unidimensional framework. This parameter can then be used in analysis of variance and regression more readily than the raw total score which has floor and ceiling effects.

Many assessments in these disciplines involve a well defined group of people responding to a set of items for assessment. Generally, the responses to the items are scored 0, 1 (for two ordered categories); or 0, 1, 2 (for three ordered categories); or 0, 1, 2, 3 (for four ordered categories) and so on, to indicate increasing levels of a response on some variable such as health status or academic achievement. These responses are then added across items to give each person a total score. This total score summarise the responses to all the items, and a person with a higher total score than another one is deemed to show more of the variable assessed. Summing the scores of the items to give a single score for a person implies that the items are intended to measure a single variable, often referred to as a unidimensional variable.

A researcher who is developing items of a test or questionnaire intending to sum the scores on the items can use a Rasch model analysis to check the degree to which this scoring and summing is defensible in the data collected. For example, if two groups are to be compared on the variable of interest (e.g. males and females), it is important to demonstrate that the workings of the items is the same in the two groups. Working in the same way permits interpreting the total score as meaning the same in the two groups.

In checking how well the data fit the model, it is important to be able to diagnose very quickly where the misfit is the worst, and then proceed to try to understand this misfit in terms of the construction of the items and the understanding of the variable in terms of its theoretical development. A very important part of the Rasch analysis from this perspective is to be in dynamic and interactive control of an analysis and to be able to follow the evidence to see where the responses may be invalid.

A Rasch analysis is consistent with the Rasch paradigm when the researcher is in control when accumulating evidence of the validity of the responses. No one single statistic is generally enough to decide whether a set of data fit the model for the purpose. Each analysis is a case study in determining the diagnostic evidence for the internal consistency and validity of the data. Often, there is no simple "yes" or "no" answer. It is important to use both statistical and graphical evidence simultaneously and interactively, and not mechanistically and sequentially, in making different decisions, such as whether to discard or modify an item. The researcher must use professional judgment by considering all the evidence, statistical, graphical and conceptual, in making decisions on evidence produced by a Rasch analysis.

Advantages of the Rasch Measurement Model In Test Development and Validation Analysis

The purpose of this page is to summarize the advantages and disadvantages of the Classical Test Theory (CTT) and Rasch measurement models for developing and maintaining testing standards. The intent is to provide an overview of the selected theoretical issues that endorse or fail to endorse either model; for further discussions see Crocker and Algina (1986) and Wright and Stone (1979).

Classical test theory provides a way to estimate the average error in test scores for a given test and population. As the basis of conventional measurement methods, a typical analysis involves estimating item difficulty and discrimination, as well as the reliability of test scores based on a random sample of examinees from the population.

Modern measurement methods, commonly referred to as “Item Response Theory” (IRT), also provide a way to estimate the average error and reliability of test scores. One of these models, the Rasch measurement model (Rasch, 1960), estimates the ‘goodness of fit’ between item difficulty and person ability. Unlike other IRT models, the Rasch model is unique in its capability to condition, or remove, the effects of person ability from test scores, which results in *sample-free* estimates of examinee performance; this advantage has far-reaching implications.

Several limitations of CTT have been brought forth in the measurement literature; its primary weakness is that it does not promote sample-free estimates of population values (Kundiger-Schildkmp, 1976; Hambleton & Swamiinathan, 1985; Wright, Mead, and Draba, 1976), which means that item difficulty and discrimination, as well as reliability estimates are dependent upon test scores from beta samples. This is a significant cause for concern, since beta samples typically fail to represent the population of interest. As a result, item difficulty values tend to be positively bias for high ability samples and negatively bias for low ability samples. Moreover, indices of item discrimination will tend to be higher when estimated from a more heterogeneous group, and reliability estimates will tend to over- or underestimate examinees’ true ability depending upon the variance in test scores (Hambleton et al., 1985).

A second limitation of CTT rests in its lack of control for developing strictly parallel test forms. Since sources of variance like, practice effects, influence test scores over repeated trials, examinees never test exactly the same on the second administration of an exam. In light of this, researchers must be satisfied with accepting reliability estimates for parallel forms that either over- or underestimate true ability. The Rasch model solves this dilemma by removing person variance, regardless of their true ability.

A third limitation of CTT, according to Tucker (1946), suggests that the rules for item selection fail to represent the optimal selection of items. If item variance is greatest at p=.50, then the optimal choice of item difficulties for any exam would harbor around this value. If this is the case, then a test of any number of items would be no better than a test of only one item at p = .50.

The Rasch model seeks to answer three questions: (1) which items are biased and for whom, (2) which items define the trait to be measured, and (3) which persons are properly measured by the items that define the trait (Wright et al., 1976). The model assumes that a more able person always has a better chance of success on an item than a less able one, and that any person has a better chance of success on an easy item compared to a difficult one. In other words, a persons' measure on any trait is a simple function of their *ability* and the item'*s difficulty*. All of the information needed about a person's ability is contained in their simple and unweighted response count of the number of items answered correctly on any measure.

The basic mathematical foundation of the Rasch model specifies how to convert observed counts into linear (ratio) measures. Ability (β) and difficulty (б) are combined by forming their difference (i.e., β_v - $б_i$). The difference is applied to the exponent of a 'logit' (the basic unit of measurement in the Rasch model), which is used to compute the probability of a correct answer. Ultimately, the fit of the data is evaluated by calculating the residual after the data have been used to estimate ability and difficulty. Wright (1977) describes it this way:

"When person v has more of the latent ability than item I requires, then β_v [ability] is more than $б_i$ [difficulty], their difference is positive and person v's probability of success on the item is greater than 0.5. The more person v's ability surpasses the item's difficulty, the greater this positive difference and the higher the probability of success" (p. 98).

Consequently, the Rasch model allows for observations that conform to a ratio scale. An examinee with twice the ability of another examinee has twice the odds of success on that item. And, when one item is twice as easy as another item, an examinee has twice the odds of successfully answering the easier one. (Hambleton et al., 1985).

Summarily, assaults on the Rasch model typically employ two strategies (Licarce, 1996), an attack on the models mathematical foundation, and attacks on its use as a *general* and *viable* solution for measurement problems. The following section refutes these attacks by providing evidence that supports the contrary.

Counter arguments

Those who attack the Rasch model's mathematical foundation of unweighted counts fail to realize that raw scores are sufficient measures of ability, and "since the Rasch model is a mathematical derivation from the requirement that linear measures be constructed from ordered qualities, it has the same standing as Pythagoras' theorem. It cannot be mathematically disproved" (Licacre, 1996, p. 512).

Those who attack the Rasch model as an inappropriate and impracticable solution to measurement problems posit that, although Rasch produces the same results as more complex (IRT) and less complex (CTT) models, it fails to factor guessing and item discrimination into the analysis of fit. For Office Specialist exams, where guessing is difficult, this assumption is nullified since there are far too many variants within an item to compute the possible range of response options. Moreover, Licarcre (1996) defends the attack this way, "if there is little guessing and item discriminations are similar, then IRT and Rasch produce similar results. If all scores are central, then raw score analysis [CTT] and Rasch produce similar results. But, neither IRT nor raw score analysis implement quality control of the construct and the data effectively" (p. 512). In other words, CTT and other IRT models fail in their lack of effort to produce sample-free estimates, whereas Rasch assumes that sample-free estimates can be obtained by removing person ability.

Finally, to counter the argument that Rasch assumes equal discrimination between items, Licarce (1996), adds that unequal discrimination of items is a diagnostic of item malfunction. "The Rasch model assumes that item parameters are the same across all samples. Constant item parameters imply a constant construct. Different item parameters across samples of the relevant population imply that the construct has changed" (Licarce, 1996, p. 512). Accordingly, to include items that discriminate provides misinformation about examinee ability, and if the construct is in a state of constant variation, we can't measure it reliably.

Advantages of the Rasch model in test development far outweigh its criticisms. The following examples provide several illustrations (Wright ,1977):

(1) *Item Fit:* Item fit can be evaluated and a decision can be made about the extent to which an item is bias. Not only does the analysis of fit determine which item is poor functioning, an analysis of residuals permits an investigation of bias in test scores with respect to language, cultural background, gender, or any other factor that can be isolated.

(2) *Item calibration is sample free:* Traditional item difficulty values differ by sample. The Rasch model adjusts item fit for the influence of sample ability, which produces sample-free estimates. This factor has noteworthy implications for the Office Specialist program since psychometric beta testing can never guarantee representative samples.

(3) *Item precision and sample size:* For tests containing more than 20 or 30 items positioned within a logit or two of the average ability level, samples as small as 100 examinees (generating standard errors about .12 logits) are generally enough.

(4) *Item banks:* Item difficulties that have been estimated from a variety of samples (or seeded into existing exams) can be transformed onto a common scale, thereby providing a subset of items that can be used to replace poor functioning ones without altering the reliability of the exam.

(5) *Linking tests and test networks:* Provided there are equivalent and existing items within an item bank, traditional test equating becomes obsolete. As the number and difficulty range of items increases, an efficient linking system can be developed which calibrates all items onto a common scale which can then be translated into equivalent forms of an exam.

(6) *Transforming Job Task Analysis into Content Valid Exams:* Rasch analysis is also useful for understanding job analysis data. Lunz, Stahl, and James (1989), for example, used the Rasch model for transforming job analysis data into test specifications from which test questions were written and content valid examinations constructed.

(7) *Setting and Evaluating Certification Standards:* Grosse & Wright (1986), described how the Rasch model could be used to facilitate test content judgments in standard setting, and also described how test standards could be held constant across time.

The Partial-Credit Model: A Special Case of the Rasch Model

Traditional Rasch analysis is used when items are scored dichotomously (e.g., correct/incorrect, true/false, yes/no), and would not be applicable to the multi-step architecture of Office Specialist items. An extension of this model developed by Masters (1982), however, specifically incorporates the possibility of having differing numbers of steps for different items on the same test. Under the *partial credit model*, the number of steps into which an item is divided and their relative difficulties can vary between items. Like the traditional Rasch model, the partial credit model conditions person ability from the estimation equations, allows for the analysis of fit, and makes possible item estimates, which are sample-free (Bond, & Fox, 2001).

Compared to CTT, the Rasch measurement model bears several advantages. The primary advantage is that it permits object-free instrument calibration and instrument-free object measurement, which makes it possible to generalize beyond the calibrated sample (Wright, 1967). Moreover, not only does the Rasch model permit traditional item analysis (e.g., item difficulty and reliability), it has the added advantages of test linking and equating. For a large-scale testing program, where the overexposure of items seems certain, the ability to equate items onto a common scale and replace poor functioning ones becomes a necessary component to help ensure exam security and integrity.

Limitations of the Rasch Model in Practice

The Rasch model is termed a "strong" model since its assumptions are more difficult to meet than those of classical test theory. The benefits derived from its use come at the cost of, on occasion, failure to define a measure at all. When data do not adequately fit the model, the instrument construction process must begin anew. An overall failure could occur if items are poorly constructed or are not comprehensible to the population, if there is a blatant mismatch between the respondent group's abilities/attitudes and item difficulties, or there are anomalies in the item-person interactions. In some cases, the data may adequately fit the model overall without defining a continuum on a trait. Items that vary little in level of difficulty may fail to define a trait, even though they may result in high internal consistency reliability estimates (Linacre, 2007 and Green & Frantom, 2002).

Invariance may fail across subgroups in a sample (Green & Frantom, 2002). While this does not invalidate the measure, it limits it use. Invariance may fail when items have different meanings depending on respondents' gender, ethnicity, socioeconomic status, or other variables (Andrich, 2012 ;Linacre, 2007 and Green & Frantom, 2002). It may also fail due to order effects, item phrasing (such as negation), survey formatting, or other contextual variables (Andrich, 2012 and Green & Frantom, 2002).

The Rasch Model Critics (*Criticisms of the Rasch Model)*

A criticism of the Rasch model is that it is overly restrictive or prescriptive because it does not permit each item to have a different discrimination. A criticism specific to the use of multiple choice items in educational assessment is that there is no provision in the model for guessing because the left asymptote always approaches a zero probability in the Rasch model. These variations are available in models such as the two and three parameter logistic models (Birnbaum, 1968). However, the specification of uniform discrimination and zero left asymptote are necessary properties of the model in order to sustain sufficiency of the simple, unweighted raw score.

The Rasch model is often considered to be an IRT model, namely the one-parameter model, but is in fact a completely different approach to conceptualizing the relationship between data and the theory (Andrich, 1989). IRT attempts to fit a model to observed data (Steinberg, 2000), while the Rasch model specifies requirements for fundamental measurement, and thus requires adequate data-model fit before a test or research instrument can be claimed to measure a trait (Andrich, 2004). Operationally, this means that the IRT approaches adjust model parameters to reflect the patterns observed in the data, while the Rasch approach requires that the data fits the Rasch model before claims regarding the presence of a latent trait can be considered valid. Therefore, under Rasch models, misfitting responses require diagnosis of the reason for the misfit, and may be excluded from the data set if substantive explanations can be made that they do not address the latent trait (Smith, 1990). It is important to note that this Rasch perspective is in contrast to exploratory approaches, which attempt to develop a theory or model to account for observed data, but follows a confirmatory approach, where a theory or model is hypothesized prior to data collection and data-model fit is used to confirm the research hypotheses. As in any confirmatory analysis, care must be taken to avoid confirmation bias.

A major point of contention is the use of the guessing, or pseudo-chance, parameter. The IRT approach recognizes that guessing is present in multiple choice examinations, and will therefore typically employ the guessing parameter to account for this. In contrast, the

Rasch approach assumes guessing adds random noise to the data. As the noise is randomly distributed, provided sufficient items are tested, the rank-ordering of persons along the latent trait by raw score will not change, but will simply undergo a linear rescaling. The presence of random guessing will not therefore affect the relationships between Rasch person measures, although a larger number of items may be needed to achieve the desired level of reliability and separation. A form of guessing correction is available within Rasch measurement by excluding all responses where person ability and item difficulty differ by preset amounts, so persons are not tested on items where guessing or unlucky mistakes are likely to affect results. However, if guessing is not random, arising through poorly written distractors that address an irrelevant trait, for example, then more sophisticated identification of pseudo-chance responses is needed to correct for guessing. Rasch fit statistics allow identification of unlikely responses which may be excluded from the analysis if they are attributed to guessing (Smith, 1990). This obviously assumes that the researcher is able to identify whether a student guessed or not by simply examining the patterns of responses in the data, so is typically used in analysis of distractor effectiveness in pilot administrations of operational tests or validation of research instruments, where exclusion of outlying persons is normal practice, rather than operational testing, where legal concerns typically dictate the use of rescaled raw scores without correction for guessing or misfit. If misfitting responses are retained, the Rasch model typically results in some items misfitting the model, and, if the number of misfitting items is excessive, there is a data-model mismatch, which has been a major criticism of the approach for decades (Goldstein & Blinkhorn, 1982). Three-parameter IRT, by contrast, achieves data-model fit by selecting a model that fits the data. Unsurprisingly, such methods result in better data-model fit, but, as a model is not specified in advance for confirmation, such an exploratory approach sacrifices the use of fit statistics as a diagnostic tool to confirm whether the theorized model is an acceptable description of the latent trait. Two and three-parameter models will still report fit statistics, but the exploratory nature of the analysis means that they are irrelevant as a tool for confirmatory analysis and lack the diagnostic value of Rasch fit statistics.

Of course, then the obvious question is why use the Rasch approach as opposed to IRT models. Rasch (Rasch, 1960 &1980) showed that parameters are separable with measurement in the educational and physical sciences, which provides *fundamental* person-free measurement if applied to the social sciences. But of course the data in the social sciences has considerable noise and error. Accordingly, proponents of Rasch measurement models assert that only data which adequately fit the Rasch model satisfy the requirements of fundamental measurement, where persons and items can be mapped onto the same invariant scale (Wright, 1992). In other words, in IRT models where the discrimination parameter is controlled, the item response curves of different items can cross. This means that the relative difficulties of items are not invariant across the sample of persons, in violation of the assumptions of invariant measurement. The implication of this is that sample independent measurement is only possible in one-parameter models, where the probability of a correct response is a function only of the difference between person ability and item difficulty. By introducing further parameters and allowing item response curves to cross, the measurement scale becomes sample dependent and relative item difficulties vary for different persons. Thus two or three-parameter models are only appropriate for uses where persons and items are not required to be mapped onto a single scale of measurement because the relative difficulty of items is not stable for different persons. The Rasch model thus has major benefits as a tool for validation of research instruments, where extensive piloting is conducted to diagnose poorly performing items, but mapping persons and items onto an invariant scale is of central interest, while two and three-parameter models have value for standardized testing where extensive piloting is impractical and correction for badly performing distractors is required. *Also in Rasch, when data meet the assumption of unidimensionality, the data also meet the assumption of local independent (Hambleton et al, 2000, Golino et al, 2012, and Ahmad et al, 2012). Also, a poorly fitting IRT model will not yield invariant item and ability parameter (Hambleton et al, 1991).*

Another characteristic of the Rasch approach is that estimation of parameters is more straightforward in Rasch models due to the presence of sufficient statistics, which in this application means a one-to-one mapping of raw number-correct scores to Rasch θ estimates (Fischer & Molenaar, 1995). A practical benefit of this is that the validation analysis of a test can be used to produce a "score table" listing Rasch measures (scores) corresponding to raw scores, and this can then be used for operational scoring of the same test form without requiring the item responses of the new administrations to be collected and analyzed. For classroom testing purposes, such procedures greatly improve the practicality of immediate IRT-type scoring with paper-and-pencil administration. However, for some purposes, this existence of sufficient statistics is also a disadvantage because it means that there are only as many Rasch measures on the standard scale as there are raw scores, whereas 2- or 3-parameter models report a person measure for each possible response string, producing finer differentiation among examinees, which is typically the goal of proficiency tests. However, unless the test is extremely long, the measurement error for individual persons will generally exceed the size of differences between Rasch and IRT person measures, so persons will not be substantively affected unless they display extremely misfitting response strings, which would require investigation to determine whether the person was following the same latent trait as other examinees. Also, for other instruments such as diagnostic tests or research instruments, finely calibrated differentiation of ability may not be the primary objective; hence different models are appropriate for different purposes. Two or three parameter models are useful for analyzing large dichotomous datasets where guessing may be significant, extensive pilot-testing to identify poorly discriminating items is not possible, and estimation of person ability is the only requirement, but outside of large-scale standardized testing, the Rasch model's ability to analyze smaller data sets than more complex IRT models and provision of invariant sample-independent measurement has major practical and theoretical benefits.

In the two-parameter logistic model (2PL-IRT; Lord & Novick, 1968) the weighted raw score is theoretically sufficient for person parameters, where the weights are given by model parameters referred to as discrimination parameters. Lord & Novick's one-parameter logistic model, 1PL, appears similar to the Rasch model in that it does not have discrimination parameters, but 1PL has different motivation and subtly different parameterization. The 1PL is a descriptive model which summarizes the sample as a normal distribution. The dichotomous Rasch model is a measurement model which parameterizes each member of the sample individually.

Verhelst & Glas (1995) derive Conditional Maximum Likelihood (CML) equations for a model they refer to as the One Parameter Logistic Model (OPLM). In algebraic form it appears to be identical with the 2PL model, but OPLM contains preset discrimination indexes rather than 2PL's estimated discrimination parameters. As noted by these authors, though, the problem one faces in estimation with estimated discrimination parameters is that the discriminations are unknown, meaning that the weighted raw score "is not a mere statistic, and hence it is impossible to use CML as an estimation method" (Verhelst & Glas, 1995, p. 217). That is, sufficiency of the weighted "score" in the 2PL cannot be used according to the way in which a sufficient statistic is defined. If the weights are imputed instead of being estimated, as in OPLM, conditional estimation is possible and the properties of the Rasch model are retained (Verhelst, Glas & Verstralen, 1995; Verhelst & Glas, 1995). In OPLM, the values of the discrimination index are restricted to between 1 and 15. A limitation of this approach is that in practice, values of discrimination indexes must be preset as a starting point. This means some type of estimation of discrimination is involved when the purpose is to avoid doing so.

The Rasch model for dichotomous data inherently entails a single discrimination parameter which, as noted by Rasch (1960/1980), constitutes an arbitrary choice of the unit in terms of which magnitudes of the latent trait are expressed or estimated. However, the Rasch model requires that the discrimination is uniform across interactions between persons and items within a specified frame of reference (i.e. the assessment context given conditions for assessment).

The Construct Validity of Test Items

Validity of a test has been defined in a number of ways by different scholars at different stages of time. Baghaei (2011) said a valid test is a test which measures what it is intended to measure. Later on, in 1954, the American Psychological Association distinguished four types of validity: content, predictive, concurrent, and construct validity. Content validity is concerned with the extent to which the items included in a test are selected from a universe of items and the extent to which they are representative of the content intended to be tested. Predictive validity is considered as the effectiveness of a test to predict the test takers' future performance and is calculated via correlating the results of the intended test with another test given in some future time; the higher the correlation, the greater the predictive validity of the test would be. Concurrent validity is very similar to predictive validity in that it is concerned with the degree of correlation with another test, the difference being that the criterion test is given at approximately the same time.

Concurrent validity is required to substitute a test for an already existing standard one due to practicality issues. Finally, construct validity is concerned with the extent to which a test is reflective of the underlying construct the test is supposed to assess. Later, predictive and concurrent validity were combined into one type of validity, namely criterion related validity (Smith, 2001). This combination was due to the fact that both predictive and concurrent validity are computed by correlating the test in focus with another test set as a criterion. Thus, four types of validity were reduced to three main types: content, criterion-related and construct validity. Gradually, theorists began to move in the direction of unifying the three types of validity into one type which was the construct validity. For example, Baghaei (2011) said Cronbach (1980) mentioned that "all validation is one", and by "one" he meant construct validity.

Finally, Messick (1989) confirming the unitary nature of validity, extended the definition of construct validity and defined it as "an overall evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions on the basis of test scores or other modes of assessment". For Messick (1989,1996), validity is a unitary concept realized in construct validity and has six facets of content, substantive, structural, generalizability, external, and consequential.

The content aspect of construct validity mainly refers to content relevance, representativeness, and technical quality. The concern of content validity is that all the items or tasks as well as the cognitive processes involved in responding to them be relevant and representative of the construct domain to be assessed. The extent to which test items or tasks are relevant and representative of the construct domain is normally determined by professional judgment of experts. Technical quality of items, referring to issues like "appropriate reading level, unambiguous phrasing and correct keying", (Messick, 1996) is also considered to be part of the content aspect of construct validity.

Substantive aspect of construct validity may be roughly defined as the substantiation of the content aspect. It deals with finding empirical evidence to assure that test-takers are actually engaged with the domain processes provided by the test items or tasks. An obvious example is multiple choice distracter analysis which is carried out to provide empirical evidence for "the degree to which the responses to the distracters are consistent with the intended cognitive processes around which the distracters were developed" (Wolfe & Smith, 2007).

Structural aspect of construct validity is mainly concerned with the scoring profile. It is highly important to take into account the structure of the test when scoring it. It does not seem sound to add up the scores of the different parts of a test, when each part measures a different dimension. While one single score can summarize the performance of an individual on a unidimensional test, scores on different dimensions must be reported separately. In other words, the scoring models should be informed by the structure of the test.

Generalizability aspect of construct validity deals with the extent to which the score meanings and interpretations are generalizable to other tasks and contents which are not included in the test but are part of the broader construct domain. In other words, the generalizability aspect tells us to what extent we can depend on the test scores as broader indicators of a person's ability and not just as an index of the examinee's ability to perform a limited number of tasks included in an assessment device.

The external aspect of construct validity is concerned with the degree to which test scores are related to other test and non-test behaviors. In Messick's (1996) own words: The *external* aspect of validity refers to the extent to which the assessment scores' relationships with other measures and non-assessment behaviors reflect the expected high, low and interactive relations implicit in the theory of the construct being assessed. Thus, the meaning of the scores is substantiated externally by appraising the degree to which empirical relationships with other measures, or the lack thereof, is consistent with that meaning. That is, the constructs represented in the assessment should rationally account for the external pattern of correlations.

Analyses of group differences and responsiveness of scores to experimental treatment (Messick, 1989) are considered to be two important methods which serve as important evidence for the external aspect of construct validity. If a measurement instrument shows sensitivity to changes in the test takers' levels of latent trait as a result of introducing treatment as an external non-assessment behavior, it is said to have external validity. That is, if a test is given before and after a treatment and results indicate that the test-takers did better on the test after the treatment, it is said the test has external validity. Moreover, a test is said to have external validity in case it can differentiate between those who possess the construct and those who do not or between those who possess varying levels of the construct.

The consequential aspect of construct validity, as the name suggests, deals with the intended and unintended (e.g. bias) consequences of the assessment and the implications scores meanings have for action. According to Wolfe and Smith (2007) the consequential aspect of validity focuses on the value implications of score interpretation as a source for action. Evidence concerning the consequential aspect of validity also addresses the actual and potential consequences of test score use, especially in regard to sources of invalidity such as bias, fairness, and distributive justice.

A simple example in which case consequential aspect of validity is violated could be a test which includes items that are biased in favor of a group of test takers and thus results in high scores for one group and low scores for the other.

Item Discrimination, Guessing and Carelessness Asymptotes: Estimating IRT Parameters with Rasch

Fred Lord's three-parameter-logistic Item Response Theory (3-PL IRT) model (Birnbaum, 1968) incorporates an item discrimination parameter, modeling the slope of the item characteristic curve, and a lower asymptote parameter modeling "guessing" or, better, "item guessability". Here is a 3-PL model, written in log-odds format, with c_i as the lower asymptote, a_i as the item discrimination, θ_n as the person ability and b_i as the item difficulty:

$$\log\left(\frac{P_{ni}-c_i}{1-P_{ni}}\right)=a_i(\theta_n-b_i) \qquad \log\left(\frac{P_{ni}}{d_i-P_{ni}}\right)=a_i(\theta_n-b_i)$$

Lord's 4-PL model (Barton & Lord, 1981) incorporates an upper asymptote parameter for item-specific "carelessness". Here is a "carelessness" model, written in log-odds format, with d_i as the upper asymptote:

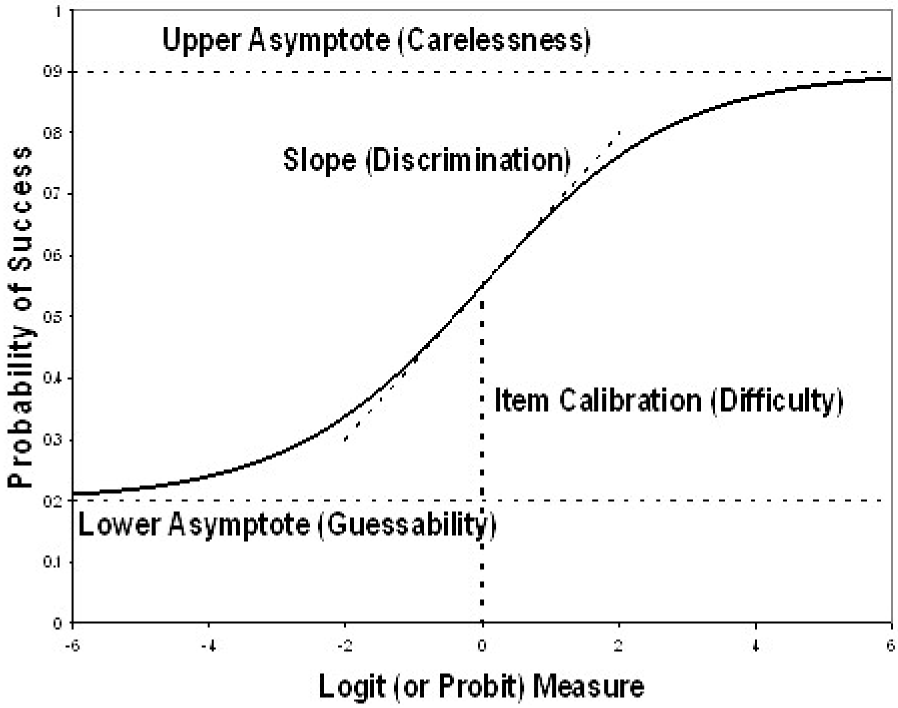


Fig 5: 4-PL IRT Item Characteristic Curve

Upper and lower asymptotes are notoriously difficult to estimate, so it appears that Lord abandoned his 4-PL model, and the value of ci in the 3-PL model is, on occasion, imputed from the number of options in a multiple-choice item, instead of being estimated directly from the data. Even the estimation of item discrimination usually requires constraints, such as "a_i cannot be negative or too big."

The dichotomous Rasch model, however, provides an opportunity to estimate a first approximation to these parameters. These estimates can be useful in diagnosing whether the behavior they reflect could be distorting the Rasch measures.

In the dichotomous Rasch model, c_i=0, d_i=1 and a_i=1. We can, however, treat the Rasch values as starting values in a Newton-Raphson iterative processed apparently intended to find the maximum-likelihood values of each of these parameters, in a context in which all other parameter values are known.

$$\hat{a}_i = 1 + \left[\frac{\sum_n (X_{ni} - P_{ni})(\theta_n - b_i)}{\sum_n P_{ni}(1 - P_{ni})(\theta_n - b_i)^2} \right]$$

$$\hat{a}_{ij} = 1 + \left[\frac{\sum_{X_{ni} \geq j} (\theta_n - b_i - \tau_{ij}) - \sum_n (\theta_n - b_i - \tau_{ik}) \sum_{k=j}^{m} P_{nik}}{\sum_n (\theta_n - b_i - \tau_{ik})^2 \left(\sum_{k=j}^{m} P_{nik} - \left(\sum_{k=j}^{m} P_{nik} \right)^2 \right)} \right]$$

Following Wright & Masters (1982, 72-77), and using the standard approach of first and second derivatives of the log-likelihood of the data with respect to the parameter of interest, we obtain the following Newton-Raphson estimation equations for the first approximations: Item discrimination (ICC slope):

$$\log\left(\frac{P_{nij}}{P_{ni(j-1)}} \right) = a_i (\theta_n - b_i - \tau_{ij}) \qquad \log\left(\frac{P_{nij}}{P_{ni(j-1)}} \right) = a_{ij} (\theta_n - b_i - \tau_{ij})$$

The Rasch expectation of a_i is 1. A corollary is that, when data fit the dichotomous Rasch model, there is zero correlation between the observation residuals and their generating measure differences. There is a similar result for polytomous items. The Generalized Partial Credit can be written:

$$\hat{c}_i = \left[\frac{\sum_{n,x=1} e^{-(\theta_n - b_i)} - \sum_{n,x=0} 1}{\sum_{n,x=1} e^{-2(\theta_n - b_i)} + \sum_{n,x=0} 1} \right]$$

$$\hat{a}_i = 1 + \left[\frac{\sum_n \left(M_{niX_{ni}} - \sum_{k=1}^{m} P_{nik} M_{nik} \right)}{\sum_n \left(\sum_{k=1}^{m} M_{nik}^2 P_{nik} - \left(\sum_{k=1}^{m} M_{nik} P_{nik} \right)^2 \right)} \right]$$

The "generalized" item discrimination (ICC slope) is equivalent to a Rasch item discrimination index. For the discrimination of polytomous inter-category "generalized" thresholds: the "generalized" threshold discrimination is:

$$\left(\sum_{n,x=1} e^{-2(\theta_n - \delta_i)} + \sum_{n,x=0} 1 \right) \qquad \hat{d}_i = 1 - \left[\frac{\sum_{n,x=0} e^{(\theta_n - \delta_i)} - \sum_{n,x=1} 1}{\sum_{n,x=0} e^{2(\theta_n - \delta_i)} + \sum_{n,x=1} 1} \right]$$

Returning to the dichotomous model, the lower asymptote (guessability) is: where $0 <= c_i <= 1$. The Rasch expectation of c_i is 0. The upper asymptote (carelessness) is: where $0 <= d_i <= 1$. The Rasch expectation of d_i is 1. In practice, it is convenient to use only observations in the lower tail for estimating the lower asymptote, in the center for estimating discrimination, and in the upper tail for estimating the upper asymptote.

Overview of Rasch Model Indices

The Rasch model is a mathematical formula that specifies the form of the relationship between persons and the items that operationalize one trait. Specifically, the likelihood of higher scores increases as people have more of the trait and decreases as they have less of the trait, whereby items become more difficult to endorse. The Rasch model assumes that item responses are governed by a person's position on the underlying trait and item difficulty. As implied by the theory's name, item responses are modeled rather than sum total responses. The model makes no allowance for deliberate or unconscious deception, guessing, or any other variable that might impinge on the responses provided. We model only the trait and not minor, peripheral influences. The Rasch model is a one-parameter model, meaning that it models the "one" parameter difference between person position and item difficulty. However, it actually provides two parameter estimates: person position and item difficulty, also referred to as person logit and item logit respectively, where a logit is a translation of the raw score. Equal-interval measures can be constructed using the Rasch model, where persons and items exist on a common scale. In other words, raw scores are nonlinearly transformed into position estimates for items and persons so that the data best fit the model.

Fit statistics provide the indices of fit of the data to the model and usefulness of the measure. Fit statistics include the average fit (mean square and standardized) of persons and items, and fit statistics reflecting the appropriateness of rating scale category use. The fit statistics are calculated by differencing each pair of observed and model-expected responses, squaring the differences, summing over all pairs, averaging, and standardizing to approximate a unit normal (z) distribution. The expected values of the mean square and standardized fit indices are 1.0 and 0.0, respectively, if the data fit the model. Fit is expressed as "infit" (weighted by the distance between the person position and item difficulty) and as "outfit" (an unweighted measure). Infit is less sensitive than outfit to extreme responses.

Person fit to the Rasch model is an index of whether individuals are responding to items in a consistent manner or if responses are idiosyncratic or erratic. Responses may fail to be consistent when people are bored and inattentive to the task, when they are confused, or when an item evokes an unusually salient response from an individual.

Similarly, item fit is an index of whether items function logically and provide a continuum useful for all respondents. An item may "misfit" because it is too complex, confusing, or because it actually measures a different construct. As test developers or survey researchers, we work to use the same language that respondents use and carefully frame items in that language on the test items or the survey. Fit statistics allow us to check whether we truly have a basis for communication. As already noted, “fit” is expressed as a mean square and as a standardized value. Values for differentiating "fit" and "misfit" are arbitrary and should be sufficiently flexible to allow for researcher judgment. Also, some fit values will appear too large or too small by chance.

Person and item separation and reliability of separation assess instrument spread across the trait continuum. “Separation” measures the spread of both items and persons in standard error units. It can be thought of as the number of levels into which the sample of items and persons can be separated. For an instrument to be useful, separation should exceed 1.0, with higher values of separation representing greater spread of items and persons along a continuum. Lower values of separation indicate redundancy in the items and less variability of persons on the trait. To operationalize a variable with items, each item should mark a different amount of the trait, as for instance, the way marks on a ruler form a measure of length. Separation, in turn, determines reliability. Higher separation in concert with variance in person or item position yields higher reliability. Reliability of person separation is conceptually equivalent to Cronbach's alpha, though the formulas are different (Green & Frantom, 2002 and Ahmad et al 2012).

Persons and items can "overfit" or "underfit." Overfit is indicated by a mean square value lower than 1.0, and a negative standardized fit. Overfit is interpreted as too little variation in the response pattern, perhaps indicating the presence of redundant items. Although it provides a guide to refining an instrument, it is otherwise probably of little concern. Underfit ("noise") is indicated by a mean square >1.2 and standardized fit >2.0 (as one rule of thumb) and suggests unusual and/or inappropriate response patterns. These indices can be used to identify and sometimes correct a measurement disturbance. Placement of items and persons on a common scale permits evaluation of scale function relative to the sample. The Rasch model software, WINSTEPS, BILOG MG3, PARSCALE or MULTILOG (Linacre & Wright, 1999-2001) can graph person position with item position. Simultaneous positioning of items and person responses illustrates where responses place each person with respect to those items. This graph is useful in three ways: (1) It can be used to determine the extent to which item positions match person positions. If the positions do not line up, the items are likely inappropriate for the persons (e.g., too easy or too hard). (2) Gaps in the measure can be detected, suggesting where items might be added. (3) Item order can be reviewed to assess the validity of the measure. Logic in the arrangement of items indicates that a researcher understood the construct, adequately operationalized it with the items written, and successfully communicated it to respondents via the items written to define it. The logic of item placement depends on qualitative judgment by the researcher, and is based on his or her knowledge of and experience with the construct.

No definitive rules exist regarding what is considered acceptable and unacceptable fit but some suggestions for acceptable fit are as follows: (1) Mean square (infit or outfit) between .6 and 1.4, for sample as small as 30 and above (Bond & Fox, 2007) (2) Mean square (infit or outfit) between .8 and 1.2 (Bode & Wright, 1999), (3) Mean square less than 1.3 for samples less than 500, 1.2 for 500-1,000 and 1.1 if n>1,000 (Smith, Schumacker, & Bush, 1995, Green & Frantom, 2002 and Ahmed et al, 2012), (4) standardized fit (infit or outfit) between -2 and +2, (5) standardized fit between -3 and +2, and (6) standardized fit less than +2 (Smith, 1992). Infit is a weighted goodness-of-fit statistic, where unexpected responses to items close to the person's logit position are weighted more heavily than unexpected responses to items far away from the person's level (information laden). Outfit is unweighted and so is sensitive to extreme unexpected responses (*out*lier sensitive). The score correlation is the correlation between item score and the measure (as distinct from a total score), and so is an item discrimination index. It should be positive. Therefore, for this study just as Green & Frantom, 2002 and Ahmad et al, 2012 suggested 0.7-1.1 was used for Mean square (infit and outfit) while -2.0 and +2.0 was used for standardized fit (infit and infit) since the sample size is greater than 1000.

Data Requirements for Design and Analysis with the Rasch Model

An instrument can be developed using classical test theory and/or item response theory. In general, the tasks involved are the same. Using the Rasch model, however, provides an opportunity to attend to the anticipated item positions along a continuum of item endorsement difficulty. A panel of experts can be a valuable resource for judging the difficulty level of items through a sorting process (Baghaei & Amrahi 2011 and Green & Frantom, 2002). A hierarchical ordering of items by the panel of experts that is similar to the ordering determined by the primary researchers would suggest that they have a common understanding of the construct. The empirical item order would be expected to conform to a similar pattern. An instrument best defines a trait when the items written to support it, function consistently throughout the instrument development process. Inconsistencies can suggest areas for reconsideration. Note that data collected from instruments that were *not* designed with Rasch analysis in mind can still utilize the Rasch model trait continuum to see how well the construct was understood. An initial requirement, then, is item sorting by the primary researcher and an expert panel.

A sample size of at least 100 and a minimum of at least 20 items are suggested for obtaining stable indices when using Rasch analysis (Green & Frantom, 2002). Analyses can still be conducted, however, with far fewer people and items. Smith, Lawless, Curda and Curda, (1999) used 48 people and 14 items for analysis with the Rasch model. Arguably, Rasch analysis can be informative with small samples since the data are responses to individual items which generates an N x n matrix (Persons x Items). Green & Frantom (2002) said that Rasch model can be used with categorical data, rating scale data, or frequency count data. Logit person and item positions can be adjusted for the influence of extraneous variables (e.g., judge severity, person gender). Also, a large sample size of 30million and a large item size of 10000 can equally be used for obtaining item banking when using Rasch Analysis. A software called WINSTEPS 3.75 can handle as large as that during analysis (www.rasch.org/rmt)/ Website: www.rasch.org/rmt/contents.htm,

Students' Achievement in Mathematics

Students' achievement measurement has been one of the most difficult tasks to deal with in education as no measurement index can give its complete package (Bessong & Obo, 2005). Periodic tests of students on various subjects are likely to give the true ability assessment of the students in various fields of study since performance to a great extent depends on the condition of the examinee's state of mind at that moment. This understanding made the Nigerian Federal Government in total agreement with Ariyo, 2006 in Olaleye (2004) to emphasize the use of continuous assessment (CA) at all educational levels and stipulates the adoption as in the 6-3-3-4 system.

Continuous assessment which has been adopted at all levels of Nigerian educational system is supposed to be a noble. This system has not also been able to determine true performance of the students as a result of defective implementation. This is because the teachers do not develop valid and reliable tests. They quickly scribble some questions and the method of marking the scripts is not objective enough. The teachers should therefore ensure that they develop tests that would measure adequately the required ability traits.

Students are not the same especially when it is observed that the rate at which facts and principles in Mathematics are being assimilated. This means that there is disparity in the ability to perform specific tasks. According to Olaleye & Aliyu (2012) and Adeyemi (2008), all aspects of Mathematics could be said to be problem solving and students have varying abilities when they are confronted with problems to solve. It is the view of Olaleye & Aliyu (2012) and Adeyemi (2008) that problem solving in science and Mathematics depend on student's cognitive ability level. This statement agrees with Aliyu's (2008) finding, that students who were successful in solving chemical problems and those with high proportional reasoning ability tend to use algorithmic reasoning strategies in Mathematics more frequently than non-successful and low proportional reasoning students.

Studies such as Adekola, (2012) and Olaleye (2004) have shown that within the Nigerian environment, learners are qualitatively different in their ability levels and in learning problems. Olaleye & Aliyu (2012); Adekola (2012) and Ariyo (2006) and Ariyo (2006) have shown that method of instruction can influence, the performance of low achieving students. The problem solving instructional technique, either as a teaching strategy by teacher or as a self-learning technique, has been found to be useful in teaching students.

The lack of classroom spaces, furniture, chalkboard, etc. has become a permanent feature in the schools. Population explosion and inadequate teaching personnel have continued to hamper learning incompetence and non-dedication to the duty, on part of the teachers weigh most on student's poor performance. In the good old days, only the best and interested personnels were found in the teaching profession. They were teachers by calling and they held teaching supreme in their entire endeavour. But today, teaching is a launching pad for other more lucrative jobs. The transitional nature of teachers today is summarized by Adekola (2012) who after his interview with dons of some universities in western Nigeria stated that "neglect and deprivation have turned ivory towers into empty towers as teachers are leaving in droves". They further showed that while there were 11,016 lecturers in the Nigerian Universities in 1985/86 session, the number dropped to fewer than 8,000 by 1991. Teachers are no more dedicated to their assignments. They give more time to trading, petty contract, farming and other businesses. Those classes of teachers have no regard for the curriculum and the school calendar. They sneak in and out of the c1ass at will. Topics are selectively and haphazardly taught. Tests and examinations are done on familiar questions/topics and marks generously awarded to give the impression of good teaching. The students feel happy with class performance which is short lived with the result of an external examination. This lack of dedication is expressed by Adegoke (2002) that "all manner of people who are no better than cheats are employed as teachers". Most teachers lack the competence expected of them. Some of these teachers actually went through the same system whose students' performance is being questioned. Preparedness as distinct from maturation is very vital for students' good performance in Mathematics. One of Thorndike's laws of learning (Aliyu, 2002) is the law of readiness which implies preparatory adjustment of the organism confronted with the problem. Preparatory adjustment of the student starts from primary through secondary to tertiary education. The vicious circle most often starts from elementary preparedness. A good preparation of primary children in primary science will produce high quality motivational preparedness for science, technology and Mathematics at higher levels.

The belief that Mathematics is a very difficult abstract subject and therefore is the field of males and that it should not be made compulsory are not encouraging statements.

This belief however, agreed well with the result of Gross National Studies of Educational Attainment (FME, 1989) which states that ‘Universally, girls appear to perform less well than boys in Science and Mathematics. In a communique issued at the end of a seminar organized by the Women Research and Documents Centre (WORDC) at the Institute of African Studies, University of Ibadan, from 4th to 6th November, 1979 (FME, 1989) the participants observed that girls have been systematically discouraged from undertaking training in the Sciences and Mathematics as a result of the bias which perceived these spheres as male disciplines.

There is loss of continuity resulting in poor implementation of educational policies. Teachers and students become aggrieved with unconducive dictates from the government and naturally offer resistance that will retard progress in learning. Resolution of such crisis finally settles the staff and students with rushing to complete the syllabus and take examinations, resulting in poor performance. There is need for students’ performance to be improved upon if all stakeholders in education play their role, work diligently and public examining bodies look into their method of development of tests and system of measuring students’ performance (Olaleye, 2004).

Location and Students’ Achievement in Mathematics

School location over the years has been an important variable that has been found to influence the achievement of students in all school subjects (Charleston, 2004). In almost all the works done in this area especially in science based subjects, point to the fact that location of a school and the testees influence achievement (Reeves, 2011; Williams, 2005). Some of these studies examined the effect of location and their attendant consequences on students’ achievement in the various subjects especially in science subjects. These studies were designed to assist education authorities and stakeholders in education to decide the suitability of where a particular school type should be located to maximize academic achievement. These research works provide government and individuals with information to decide whether a new school should be built or otherwise. The differences in the academic achievement due to location could be as a result of preference by teachers to work in some locations than the others. Johnson (2011) concluded that high qualified teachers prefer to serve the urban to the rural areas. Many teachers do not accept being posted to the rural areas and even if they accept, they do not leave in those rural areas, thereby not being fully committed to their duties. Many rural schools do not have adequate amenities and facilities that can enhance teaching and learning, such like science laboratories, Mathematics laboratories and equipment, libraries and other infrastructural facilities (Williams, 2005). More recently in Nigeria, there are some locations or cities where there are security challenges, qualified teachers or educational personnel do not go there in search of work or posting. This is likely to affect the performance of students in such locations.

Gender and Students' Achievement in Mathematics

Mathematics is a science subject and some gender-based science researchers have reported that what both the 'Feminist empiricists' and the 'liberal critics' seem to agree that female in principle will produce exactly the same scientific knowledge as males provided with sufficient rigour undertaken in scientific inquiry (Howes, 2002, Olaleye, 2004). They also believe that females and males are equal in their approach in science and that inequality in science and science education is caused by political educational and social factors external to science. If these external factors are removed, males and females will achieve equally. There is need therefore to give boys and girls exactly the same opportunities and challenges.

One of the reasons students often find it difficulty to cope with Mathematics examination and some other subject more than others is the extent of interaction between concepts of images (Adegoke, 2002). This by implication, if a students achieve more in Mathematics, it is likely he or she does well or better in others subjects because Mathematics is logical and is involved in everyday life.. It is also believed that learning outcome in Mathematics is gender based, because male children learn about people's emotion and their relationships while girls learn more of people and their personal relationship (Spelk, 2005). The function of gender in the learning of Mathematics in secondary schools cannot be over emphasized. Almost all the empirical studies in this area point to the fact that boys perform better than girls in Mathematics (Olaleye, 2004). Olaleye (2004) in a study on approaches to gender equity in science education concluded that Mathematics are gender based and that male perform better. UNESCO (2003) also confirmed that the poor performance in Mathematics is worsened by gender imbalance.

Some scholars have found that it is only in some aspect of Mathematics that males are performing better (Abiam &Odok, 2006; Hopkins, 2004). This could be as a result of Mathematics being a subject of number and the abstract nature of it required much thinking and demands extra time and attention. Effort should be geared towards helping students to achieve more in Mathematics by assessing them with well designed, structured and secured items in Mathematics. A proper development and validation of Mathematics items will help to remove the impression that achievement is gender based, thereby bridging the gap between male and female achievement in Mathematics.

Social Economy Status and Students' Achievement in Mathematics

Poor academic performance according to Aremu and Sokan (2003) is a performance that is adjudged by the examinee/testee and some other significant as falling below an expected standard. Poor academic performance has been observed in school subjects especially Mathematics and English language among secondary school students (Akanle, 2007 and Ojerinde, 2013). Aremu (2000) stresses that academic failure is not only frustrating to the students and the parents, its effect are equally grave on the society in terms of dearth of manpower in all spheres of the economy and politics. Education of secondary school level is supposed to be the bedrock and the foundation towards higher knowledge in tertiary institutions. It is an investment as well as an instrument that can be used to achieve a more rapid economic, social, political, technological, scientific and cultural development in the country. The National Policy on Education (2004) stipulated that secondary education is an instrument for national development that fosters the worth and development of the individual for further education and development, general development of the society and equality of educational opportunities to all Nigerian children, irrespective of any real or marginal disabilities.

In most African countries and western world, economic status of a family is usually linked with the family's income, parents' educational level, parents' occupation and social status among the kiths and kin and even at the global level. Ford and Harris (1997) followed this logic while examining parental influences on African American students' school environment by focusing on specific socio-demographic factors, including parents' level of education, marital status, and family income. It is generally believed that children from high and middle economic status parents are better exposed to a learning environment at home because of provision and availability of extra learning facilities. The use of data about family possessions may be thought to be connected to economic status, students who used a computer both at home and at school achieved a significantly higher science score than those who only used a computer at school (Thompson and Fleming, 2003). Children from low socio-economic status parent do not have access to extra learning facilities; hence, the opportunity to get to the top of their educational ladder may not be very easy.

Danesty (2004) while discussing the "low-income parents" beliefs about their role in children's academic learning mentioned that a few of these parents indicated that a few of these parents indicated that their responsibilities were limited to meeting children's basic and social/emotional needs, such as providing clothing, emotional support and socializing manners. So these parents's shortsightedness towards their responsibilities in the educational processes of their children and scarcity of fund to intensity such processes could be a challenge to their children's success. In and of themselves such socio-demographic variables do not fully account for the academic successes or failure of minority students (Smith, Schneider and Ruck, 2005). Low economic status children are often left home to fend for themselves and their younger siblings, while their caregiver work long hours; compared with their well-off peers, they spend less time playing outdoors and more time watching television and are less likely to participate in afterschool activities (U.S. Census Bureau, 2000). Unfortunately, children would not get the model for how to develop proper emotions or respond appropriately to others from watching cartoons; they need warm, person-to-person interactions. The failure to form positive relationships with peers inflicts long term socioemotional consequences (Steinberg, Dornbus & Brown, 1999).

Previous studies in the same field have established that other factors in spite of socio-economic status (SES) can boost academic successes among students. Studies which examined African American parents recorded that parents who maintained positive views about the value of education and who hold high academic expectations have children who often experience higher levels of academic achievement (Ford & Harris, 1997 and Steinberg et al, 1992). Children and families living in poverty are at greater risk of hunger, homelessness, sickness, physical and mental disabilities, violence, teen parenthood, family stress, and educational failure. These environmental factors are contributors to children that live in poverty being four times more likely to have learning disabilities than non-poverty students (Apple and Zenk, 1996). According to Casanova, Garcia-Linares, Torre and Carpio (2005), it is a combination of these environmental factors as well as family influence that contributes to student's academic success. If a student has not eaten for days and has clothes that don't fit, however, he/she be expected to maintain focus in a classroom? Children coming from poverty are not provided the same tools as the wealthy; they are entering schools already behind those not living in similar conditions. According to Li-Grining (2007), research suggests that the problem starts with the parents and their lack of education and understanding of the needs of children. Many individuals who might have done this nation proud in different fields have been forced into uninspired careers due to unavailability of finance resources. Such individuals are forced out of school and made to engage in hawking, selling packaged drinking water and the likes so as to save money for their school expenses. Most of the time, they cannot afford instructional materials, and are always at the mercy of examiners during examination period. The persistence of this in life of an individual student may spell doom for his academic success. Tracy and Walter (1998) corroborate this when they submit that individuals at the lowest economic level are often the least well-served by the school system. A considerable number of researches repeatedly have shown that low socio-economic status is linked to a range of indicators of child and adolescent well-being, including students' academic achievement (Tracy & walter, 1998). Aremu (2004) claimed that poverty contributes towards educational failure, simply because poor children are all "culturally disadvantaged", but because their health and nutritional status is inadequate to allow for the maximum mental development and for the realization of their educational potential. The likelihood that the poor children would end up being at risk in terms of deficient development is a reality that could begin even before birth. In that regard, Birch and Gussow (1979) emphasized that society should concern itself more with the full range of factors contributing to the educational failure, among which the health of the child is a variable of potential primary importance.

Other factors according to Dantesy (2004), complementing environmental and socio-economic factors to produce high academic achievement and performance include good teaching, counseling, good administration, good seating arrangement and good building. Dilapidating buildings, lacking mental stimulating facilities that are characterized with low or no seating arrangement will also be destructive. Dantesy (2004) however lamented that the innovative environment do stimulate head start learning and mental perception, not only that, it has also been proved that students that come from simulative environment with laboratory equipments or those that are taught with rich instructional aids, pictures, and allowed to demonstrate using their functional peripheral nerves like eyes, hands, and sense of taste performed better than those trained under theoretical and canopy of abstraction. Thus, teaching and learning should be done under organized, planned, and fortified environment with learning instructional aids to stimulate students' sense of conception, perception and concentration to facilitate systematic understanding and acquisition of knowledge in them. In sum, a combination of a healthy family background, living in a good environment plus the child being educated in a conclusive environment with fortified learning or instructional aids or motivational incentives will prompt academic performance and lack of it will retard academic performance.

Empirical Studies Relating to The Rasch Analysis Model

Wistner (2012) in applying the Rasch model in the validation of a test of matalinguistics knowledge designed a ten-item multiple choice test for 100 Japanese university students. The result of the Rasch analysis indicated that the calibrated item estimates were sufficiently accurate and precise. All the items exhibited good fit with the Rasch model (M=0.18, SD= 1.29) for average person measures which was slightly higher than the item estimates(M= 0.00, SD= 1.91)

Ahmad et al (2012) in advance in educational measurement: a Rasch model analysis of Mathematics Proficiency Test (MPT) used a self- developed 50 items in MPT on 588 students of grade 8. The Rasch model was used for the analysis with reliability of items difficulty measures of 0.99 with Cronbach's alpha of 0.90 indicating that it can be replicated. The MNSQ for both infit and outfit are 0.76 and 1.30 respectively while the ZSTD for both infit and outfit are +2 and -2 respectively. The data fit the Rasch model in all.

Chuesathuchon & Waugh (2008) conducted a study in Thailand on item banking with Rasch measurement with the intention to create a Mathematics items bank. First, 290 multiple-choice test items were created for the item bank. Through review, 40 items were deleted and 250 items were administered to 3062 students at the final year. 172 out of the 250 were deleted as not fitting the Rasch measurement model, only 78 items were stored. The person separation index was 0.83 while item and persons reliability are 0.98 and 0.57 respectively. This might be as a result of the narrow spread of the person on the person-map-item. In all, the items showed a model fit.

Golino et al (2012) carried out a study that involved the used of the Rasch model. The sample was 167 and 188 Brazilian people with 48 items. Rasch model was applied, the result showed item reliability of 0.97, infit mean of 0.87 (SD=0.28, max= 1.69, min=0.39) and person reliability of 0.98 while for 188 people, the item reliability was 0.99, infit mean was 0.94 (SD=0.22, max=1.46, min= 0.56). The person reliability was equally 0.95. A sample t-test showed a significant with small effect size for the two analyses.

Andrich (2012) carried out a study that involved the use of the Rasch dichotomous model calculated by using the software winsteps. Out of the 48 items, 5 were responded to correctly by all participants. The reliability for the forty three (43) items (non- extreme) was 0.99 and for the full scale (48 items), the reliability was 0.97. The infit means was 0.87 (SD=0.28, max=1.69, min=0.39) falling within the acceptable range. The person reliability was 0.95, which is estimated to indicate the degree to which a person's response pattern conforms to the difficulty structure of the measure. The principle contrast showed that the raw variance explained by measures (modeled) is 70.6% and the unexplained variance in the first contrast (modeled) is 10.4%, suggesting that the instrument can be thought of as unidimensional. A one-sample t- test with 95% confidence interval was used which showed that there is no significant differences between the mean scores of the gender, SES, school location and school type conducted in the work. The effect size of the cohen's (d) was large for the study.

Osman, Naam & Jaafar (2012) carried out a research on 64 final year testees in 2010/2011 in the department of Civil and Structural Engineering using the Rasch analysis of Winsteps. The person mean value for the analysis is 3.02 which was higher above the threshold value mean item = 0. Besides, totally 53 testees (82.8%) were found to be above the mean item and the highest person managed to scored 9.09logit while 11 testees (17.2%) were located below mean item. The Cronbach $\propto = 0.66$ value was showed on the summary statistics table which was higher than the acceptable level of 0.6. This validate that the model is acceptable. From the analysis, the person reliability is 0.57 while the item reliability is 0.00 which is rather low and will need further inspection. Students' separation 1.16 was equally low. This was enough to separate them into different performances level.

Nkpone (2001) applied two latent trait models which are one and two-parameter logistic models of the item response theory and the classical test in the development and standardization of a physics achievement test (theory paper) for senior secondary students. The sample used was 2215 (1349 males and 866 females) students who sat for the Senior School Certificate Examination during the May/June 1999, school year in Rivers State. Their ages ranges from 16-20years with a mean of 18years and standard deviation of 2.6

years. Approximately, 66.0 percent (1452) of the sample came from urban schools while approximately 34.0 percent (763) of the sample came from rural schools in the state. The 60 multiple-choice item physics achievement test was administered to the subjects and scored. The resulting data analysed. The author tested for the reliability and validity for each item and the whole test, the unidimensionality of the instrument, the item parameter estimation (the difficulty and discrimination indices) of the items, the person parameter estimation (the magnitude of the ability estimates), each item's standard error measurement and the analysis of fit test. The specific item parameter that was employed in comparing the classical test theory with the item response theory was the difficulty and discrimination parameters.

The analysis involved the use of PROX and regression techniques on the Microsoft Excel Visual BASIC computer programme. The chi-square goodness of fit test was used and also factor analysis was used to establish the unidimensionality of the test items. The results showed that the overall reliability coefficient (KR_{20}) of the instrument was 0.89 and the fit to the model was good and that there was no significant relationship between easiness values obtained from two contrasting score groups.

Opasina (2009) in a study on development and validation of alternative to practical Physics used 3PL model of IRT. She used 11 local government areas of Oyo state, 160 secondary schools and 1545 sample (boys= 867; girls= 678). 1395 of the samples are from public schools while 150 samples are from private schools. The items were trial tested using 60 students who were not part of the samples used. It was observed that the intercorrelation of some subgroups were high and the correlation values were between -0.48 to 0.63. The b parameter ranged between -0.92 to 0.99 and the standard error of measurement was from 0.03 to 0.09 which showed a good precision. The present study aimed at developing and establishing the psychometric properties of a Mathematics Achievement Test (MAT) using the Rasch model.

Appraisal of Literature Review

The literature review covered both empirical and theoretical background of Achievement Test and measurement in its historical perspectives. The reviewed literature also explored the historical and conceptual background of the item response theory models and precisely focused on the Rasch model. Item analysis utilising latent trait models were reviewed with regard to their capacity for provisional items selection.

Both theoretical and empirical evidence of instrumentation research were reviewed and the need for valid assessment instruments highlighted. Many authors based their studies on the classical test model. This is primarily due to its simplicity, both conceptually and computationally. Unfortunately the classical test model has some serious disadvantages, which made its use in developing and validating achievement tests questionable. For instance, in the classical test model, item statistics change from a sample whose mean ability is high to one whose mean ability is low. In addition, the classical test model typically provides only one overall standard error of measurement for all items on the entire test. Due to these limitations of the classical test model, the present study aimed at developing and validating Mathematics test, using the Rasch model of one parameter logistic model of the item response theory.

Previous studies such as those carried out: compared the classical item analysis with the latent trait model; investigation of the pretest-posttest validation of a criterion-referenced test and test information function of the latent trait model which did not establish the unidimensionality, single trait for ability of data and also disregarded the issue of validity of instruments by not exploring the fit of the latent models to data. Also, some studies carried out made use of some of the Messick's validity six facet procedure to establish Rasch model but this present study made use of the six facet procedures of Messick's validity in establishing the Rasch model in this single study with unidimensionality trait. The content aspect, substantive aspect, structural aspect, generalizability aspect, the external aspect and the consequential aspect were all established in the study carried out by the researcher using the Rasch model.

The previous studies too made use of small sample sizes with small item sizes, which limited the partitioning necessary in the model control. A further look at previous studies indicated that no empirical study has been carried out using Mathematics Achievement Test for senior school precisely. Also, no study of this type has been carried out in Nigeria as known to the researcher. This is going to be a great asset to the government, examining bodies, classroom educators, measurement experts and educational sector if they could explore the procedure in this study in other areas too.

In summary, existing problems in previous studies highlighted above were considered in the present study, utilizing the developed and validated Mathematics Achievement Test on a sample of 1499 senior secondary school III students applying the

Rasch model. This will ensure that the examinee's pattern of responses to items could be used to estimate his/her trait value. If items are developed and calibrated to cover the continuum of all possible levels of a given well defined trait, then a testee's performances over time could be determined and compared as an indicator of the amount of learning that has taken place within the time. Thus, this study fill the gap in literature created by the absence of Mathematics test whose item parameters are sampled dependent in the measurement of achievement in Senior School Certificate Examination in Nigeria.

CHAPTER THREE

RESEARCH METHOD AND PROCEDURE

This chapter presents how the study was conducted, how the data for the study was also collected and the means by which the data was collected and how the instrument that was used for data collection was developed and used. The various aspects that this chapter looked at are:

i) Design of the study
ii) Population of the study
iii) Sample and Sampling Techniques
iv) Instrumentation
v) Validity of the Instrument
vi) Reliability of the Instrument
vii) Method of data Collection
viii) Method of data Analysis

Design of the Study

The study was designed to develop and validate Mathematics Achievement Test (MAT) for senior secondary three students in Oyo and Delta State. Instrumentation research design was deemed appropriate for this study. According to the International Centre for Educational Evaluation (ICEE, 2003), it is a study which aims at introducing new contents, procedures, technologies or instruments for educational practices. This study has thus made use of the Rasch model of Item Response Theory techniques in analyzing Mathematics test items. Also, Alli (1996) and Nworgu (2011) defined instrumental research as a study, which was geared towards the development and validation of measurement instruments in education.

Population of the Study

The population size of the study comprised the entire 100,470 students in Senior Secondary School three (SSS 3) students in Oyo and Delta states who enrolled for the Senior School Certificate Examination (SSCE) in 2013. The population was chosen because it was assumed they had covered the required syllabus for Mathematics. (source: internet, appendix xvii)

Sample and Sampling Techniques

One thousand five hundred (1500) testees formed the sample for this study. The multi-stage sampling technique was employed for the study. Simple random sampling technique was used to select five (5) local governments from each state. Delta state has 25 LGAs while Oyo state has 33 LGAs.

The selected LGAs were cluster into Urban and Rural areas while purposive sampling technique was employed to select 2 private schools and 2 public schools from each of the five selected LGAs making a total of 20 schools from each state. Therefore, the total schools used for the study were 40 secondary schools altogether. The testees were selected using the non-proportionate stratisfied random sampling technique from the selected schools to arrive at the 1500 testees which formed the sample for the study (20 testees each from 5 schools and 40 testees each from the remaining 35 schools making a sum of 1500 samples).

Instrumentation

The instruments used in this study were a self-developed 100-Mathematics Achievement Test (MAT) items with the proportion of the schemes/ topics and a 25-validated Questionnaire on Social Economy Status (QSES), adapted from Odili (2005) and Akinbile (2007). In the development of the MAT, three fundamental test parameters were identified: test content, learning outcomes and item difficulty or item unidimensionality using the PCA. The first two parameters were outlined in the curriculum specifications (Federal Ministry of Education, 2009) whereas the difficulty for every item was conceptually determined by the researcher and other experts in the Measurement and Evaluation department. Test content included the scheme, topics and subtopics. Learning outcomes identified the knowledge, skill and abilities that students needed to demonstrated at the end of the every topic or subtopic. Items were developed to operationalize these learning outcomes in terms of item scores. Three levels of difficulty (easy, moderate and difficult) were applied to target the learning outcomes. The difficulty levels were related to the reliability of the scored items.

At inception, the researcher generated a test of 165 multiple-choice items, each with option A, B, C, D while the QSES which were 3-point scale were graded and scored as 3, 2, 1 marks in a descending order. The items were constructed from all aspects of

Mathematics curriculum as stated by the Ministry handbook, WASSCE and NECO examining bodies' syllabi.

Validity of the Instrument

The instrument consisted of three sections with a 150-Mathematics Achievement Test (MAT) items. Section A consisted of the students' bio-data, section B consisted of 150 items of MAT developed by the researcher with Test Blue Print (TBP) which were categorized under the six behaviour levels of cognitive of knowledge, comprehensive, application, and higher thinking which are analysis, synthesis and evaluation while section C consisted of a 25-validated questionnaire relating to SES of the testees adapted from Odili, 2005 and Akinbile, 2007. The process of MAT content and face validation involved a discussion with three experienced teacher regarding the items of Mathematics test that were drawn from the secondary school curricula for Mathematics using table of specification and item difficulty. The weight of the topics in the table of specifications was given prime attention at the initial stage of the discussion. Several criteria were employed during the process. The criteria included: (i) the coverage given in the curriculum specifications, (ii) the suitability of the topics for multiple-choice format (iii) the rating of the topics by the panel. Topics with the weight of 3 were considered to be important topics since they were given the widest coverage in the curriculum specification of the Ministry, WASSCE and NECO (Osadebe, 2001, & 2014).

The next stage was to develop test items according to the table of specifications. In this study, items for the MAT were developed through various procedures as stated above. After the items had been developed, they were given again to the same panel of experienced teachers for validation. It was equally verified by the researcher's supervisors, WASSCE and NECO's chief examiners, two college Mathematics lecturers and two graduate Mathematics teachers who were members of Mathematical Association of Nigeria (MAN) at both Oyo and Delta state chapters. Out of the initially developed (165) items, some were discarded and deleted while some were reframed, eventually 150 items scaled through with the use of TBP. This was necessary to ensure both content and face validity. Hence, empirical data from the pilot or try out test was used to provide more meaningful information on item reliability of the MAT.

The 150 items that emerged from the vetting were trial tested. They were administered to 60 testees/students (30 boys and 30 girls) who were not part of the sample used. Factor analysis of SPSS was used to select items that were valid and that were measuring the same construct of achievement.

The principle component analysis of factor analysis used generated 16 major components. The 16 components were based on eigenvalue standard of not less than 1.0. This agrees with Kaiser Rule which states that eigenvalue that is less than 1.0 should be regarded as being insignificant. The total variances based on significant eigenvalues are categorized into 16 components as shown in the appendix III. Component Matrix, Rotated Component Matrix and Scree plot were used to select the 100 items which are to be used in the final administration of the test. In order to obtain conceptually similar and significant clusters of issues of the variables, principal components analysis with varimax rotation and Kaiser normalization were conducted. Eigenvalues equal to or greater than 1.00 were extracted. With regards to the 150 items used as variables, orthogonal rotation of variables yielded 16 factors, accounting for 16.274, 15.393, 11.130, 10.962, 10.395, 9.863, 8.607, 4.926, 3.344, 1.334, 1.271, 0.936, 0.897, 0.865, 0.815 and 0.812 per cent of the total variance respectively, a total of 97.824 percent of the variance. The factor loadings are presented in table in the appendix. To enhance the interpretability of the factors (Okorodudu, 2012), only variables with factor loadings as follows were selected for inclusion in their respective factors: ≥0.970 (factor 1), 21 items; =0.995 (factor 2), 22 items; ≥0.942 (factor 3), 12 items; ≥0.947 (factor 4), 10 items; =0.990 (factor 5), 13 items; ≥0.942 (factor 6), 6 items; ≥0.943 (factor 7), 9 items; ≥0.973 (factor 8), 2 items; and =0.970 (factor 9), 5 items. The total item altogether is one hundred (100). The 7 other factors were not considered based on the scree plot and the number of items needed by the researcher.

The range of rotated component matrix for each selected item is between 0.942 and 0.999 ($0.942 \leq h^2 \leq 0.999$). Factor Analysis using the principle component and rotation analysis was used to test for the unidimensionality of the MAT instrument i.e the extent to which MAT measured a single trait (Okorodudu, 2012).

Reliability of the Instrument

A KR_{20} reliability method was employed in order to estimate the reliability coefficient of the instrument. The value obtained was 0.85. On the basis of the calculated reliability coefficient, the instrument (MAT) was considered reliable for the study.

Method of Data Collection

The One hundred (100) developed items that constituted the Mathematics Achievement Test (MAT) from the PCA and Rotation analysis were administered to the 1500 testees in the two states- Oyo and Delta states. The testees were adequately briefed about the purpose of the test. The test was administered for 2½ hours under a favourable examination conditions with the aid of the test developer and the Mathematics teachers. Mathematics teachers were used to assist in the administration and collection of the data.

Method of Data Analysis

All data were analyzed using WINSTEPS Rasch software version 3.75.0. To determine item difficulty and person ability, the Rasch measure, in logits and wit were calculated for all items and persons measures (Rasch, 1960). The Rasch measure is calculated. The standard error for each item is also provided by the WINSTEPS output and when doubled and subtracted or added to the Rasch measure value, represents a 95% confidence interval of the true difficulty of this item. Item strata are calculated by the software to provide evidence for representativeness by the following formula: Item strata = ($4G_{item}$ + 1/3), where *Gitem* is the Rasch item separation value of the software (derived by dividing the item standard deviations by the average measurement error) (Beglar, 2010). Item strata identify the statistically distinct item difficulty levels (Wright & Masters, 2002). Smith (2001) requires a minimum of two difficulty level in order to be able to deem items representative of the assessed content.

A person-item map, a graphical representation of person-abilities and item difficulties is drawn. Since the Rasch model transforms raw item difficulties and raw person scores into equal interval measures (logits), it is possible to map these results linearly. The person-item map contains a wealth of arguments applicable to nearly each of Messick's (1989/1995) six facets (Beglar, 2010). The spread of items, the spread of abilities, gaps or overlaps in item difficulty or ability are all illustrated. The item measure correlations are calculated by a fairly complex formula with the software, the results of which are influenced by the predictability of the data, the item targeting on the sample of test-takers as well as the distribution of the person measures (Linacre, 2004). They can be used as evidence towards technical quality or unidimensionality.

Item correlations close to zero indicate that the item is either extremely easy, extremely difficult, or suggests that the item may be measuring the construct in a different way to other items (Wolfe & Smith, 2007). In fact, negative item measure correlations suggest that the item is opposing the direction of measurement. As part of a check for content, structural and substantive validity, two types of fit statistics were calculated. Infit statistics reflect unexpected response patterns where the test is targeting ability, essentially where the test is log[$p_{ni}/(1 - p_{ni})$]= B_n- D_i(1) where B_n is the ability of a person *n* and D_i is the difficulty of item *i*, while outfit statistics are more sensitive to guessing or mistakes, perhaps when a test taker gets an answer wrong that they should have gotten right or vice versa (Linacre, 2007). Fit statistics consist of mean-square values (MNSQ) and z-standardized scores (ZSTD) for both infit and outfit. According to Linacre (2007), acceptable values for multiple choice tests for MNSQs range from 0.7 to 1.1 for a sample above 1000 and -2.0 to +2.0 for ZSTDs (Green & Frantom, 2002).

The independent t-test was also used to test the significant of the stated hypotheses to determine the invariance of the items across the subgroups with a 95% confidence interval. To verify the adjustment of the data to the model, the infit (information-weighted fit) means-square statistics is used. It represents "the amount of distortation of the measurement system" (Linacre 2007). Values between 0.7 and 1.1logits are considered productive for measurement and < 0.5 and 1.2 and 2.0 are not productive for measurement but do not degrade it for sample <1000 (Green & Frantom, 2002). The unidimensionality of the instrument was checked by a number of procedures, each one complementing the other (Messick, 1998 and Tennant, 2000). Here, unidimensionality was addressed using the model fit statistics that is if the data fit the model, one of the consequences is the linearity of the measure, its unidimensionality and the principal contrast, which can be verified through the percentage (%) of variance explained by the measures and by the % of unexplained variance in the first contrast. The former should be closer to or greater than 60% (Peter & Stone, 2009) while the latter should be closer to or less than 5%. Each item of MAT in the testee's answer sheet was scored dichotomously either right (1) or wrong (0). Also, for the analysis of QSES, the first option was allotted with 3 marks, the second option with 2 marks while the third option with one mark. The testees who scored 45 marks and above in QSES were grouped under HSES while testees who scored below 45 marks were grouped under LSES since the total obtainable scores for the QSES was 75 marks. This was so because the average score of 75 marks is 37.5 marks. The scores were organized, analyzed and used to answer the various research questions and tested the stated hypotheses. However, in the course of analyzing the data with WINSTEPS software, person/testee 255 was observed to be incorrectly entered into the software thereby not identifying it as an object/ a testee leaving 1499 testees for analysis.

CHAPTER FOUR

PRESENTATION OF RESULTS AND DISCUSSION OF FINDINGS

The purpose of the study was to develop and validate Mathematics Achievement Test (MAT) suitable for measuring the achievement of testees using the Rasch model. The results obtained in this study are presented and discussed in this chapter. Winsteps 3.75 of Rasch model was used to answer the research questions while t-test statistic of SPSS was equally used to test the stated hypotheses. The results have been presented according to the research questions and hypotheses.

Preliminary Observations

Prior to interpretation of the item and person (position) scores in logit/wit from a Rasch analysis, appraisal of whether the data fit the model reasonably well is required. This is to serve as a precursor to data presentation. Summary tables for the Rasch model based on logit and wit are presented. Table 1-4 present overall information about whether the data showed acceptable fit to the model.

Table 1 - level of item data fit to the Rasch model in wit

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	915.1	1499.0	50.00	.57	1.00	.1	1.00	.0
S.D.	233.7	.0	7.21	.07	.03	2.5	.04	2.6
MAX.	1377.0	1499.0	64.52	.95	1.10	9.9	1.11	9.9
MIN.	426.0	1499.0	30.53	.52	.93	-7.1	.93	-7.1

REAL RMSE .58 TRUE SD 7.19 SEPARATION 12.42 Item RELIABILITY .99
MODEL RMSE .58 TRUE SD 7.19 SEPARATION 12.48 Item RELIABILITY .99
S.E. OF Item MEAN = .72

UMEAN=50.0000 USCALE=10.0000

Table 1 showed the level of item data fit to the Rasch model in wit. From the table, the mean square infit is 1.0 while the mean squares outfit is 1.0. On the other hand, the mean standardized scores for infit is 0.1 while that of the outfit is 0.0. For a fit to the model, mean squares for infit and outfit should be 1.0 respectively while the mean standardized scores (ZSTD) for infit and outfit should also be 0.0. The mean ZSTD scores for infit of 0.1 indicates that the data does not perfectly fit. It is an indication that some items in the test should be re-worked or dropped for a fit to the model based on the item individual MNSQ and ZSTD score.

The table also revealed that the separation statistics is 12.42 compared with 12.48 expected of the model. A value of 1.0 and below indicates a non-fit. In terms of separation factor, the data fit the model. The above statistics in wit is compared with similar statistics in logit.

TABLE 2 - level of item data fit to the Rasch model in logit

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	915.1	1499.0	.00	.06	1.00	.1	1.00	.0
S.D.	233.7	.0	.72	.01	.03	2.5	.04	2.6
MAX.	1377.0	1499.0	1.45	.09	1.10	9.9	1.11	9.9
MIN.	426.0	1499.0	-1.95	.05	.93	-7.1	.93	-7.1

REAL RMSE	.06	TRUE SD	.72	SEPARATION 12.42	Item	RELIABILITY .99
MODEL RMSE	.06	TRUE SD	.72	SEPARATION 12.48	Item	RELIABILITY .99
S.E. OF Item MEAN = .07						

The figure in table 2 revealed that they are the same with table 1. This reveals that the data can be analysed using the logit and wit data.

TABLE 3 - level of person data fit to the Rasch model in wit

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	61.0	100.0	55.11	2.19	1.00	.0	1.00	.0
S.D.	6.1	.0	2.91	.07	.09	1.2	.12	1.2
MAX.	79.0	100.0	64.68	2.56	1.28	3.7	1.42	3.7
MIN.	42.0	100.0	46.38	2.12	.71	-4.0	.65	-3.8

REAL RMSE	2.23	TRUE SD	1.87	SEPARATION 1.84	Person	RELIABILITY .51
MODEL RMSE	2.19	TRUE SD	1.91	SEPARATION 1.87	Person	RELIABILITY .53
S.E. OF Person MEAN = .08						

Person RAW SCORE-TO-MEASURE CORRELATION = 1.00
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .53

Table 3 showed the level of person data fit to the Rasch model in wit. From the table, the mean square infit is 1.0 while the mean squares outfit is 1.0. On the other hand, the mean standardized scores for infit is 0.0 while that of the outfit is 0.0. For a fit to the model, mean squares for infit and outfit should be 1.0 respectively while the mean standardized scores (ZSTD) for infit and outfit should also be 0.0. The mean MNSQ and ZSTD scores for infit and outfit of 0.0 indicates that the data does perfectly fit the model. The person mean here is 61.0, which suggests that these items were not difficult, on average. If the person mean is positive, the items would on the average be easy.

The separation statistics is an index of how the person spread across the latent scale. An index of 1.84 is closed enough to the maximum of 1.87. If separation is 1.0 or below, the test may not have sufficient breadth in position with the testees.

TABLE 4 – level of person data fit to the Rasch model in logit

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	61.0	100.0	.51	.22	1.00	.0	1.00	.0
S.D.	6.1	.0	.29	.01	.09	1.2	.12	1.2
MAX.	79.0	100.0	1.47	.26	1.28	3.7	1.42	3.7
MIN.	42.0	100.0	-.36	.21	.71	-4.0	.65	-3.8

```
 REAL RMSE    .22 TRUE SD    1.19  SEPARATION  1.84  Person RELIABILITY  .51
MODEL RMSE    .22 TRUE SD    1.19  SEPARATION  1.87  Person RELIABILITY  .53
 S.E. OF Person MEAN = .01
```

Person RAW SCORE-TO-MEASURE CORRELATION = 1.00
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .53

The figure in table 4 revealed that they are the same with table 3. This reveals that the data can be analyzed using the logit and wit data. Hence, it has been established that the data fit the Rasch model.

In the following section, the data is presented according to the research questions and hypotheses in the work.

Research Question One

(1) What is the validity of the Mathematics Achievement Test (MAT) items using the Rasch model?

The data that was used to answer research question one was presented in tables 5, 6 and 7 respectively. This provided information for the validity of the test.

Table 5 – validity of MAT using the Principle Component Analysis for STANDARDIZED RESIDUAL variance (in Eigenvalue units) in Rasch

		Empirical			Modeled
Total raw variance in observations	=	113.4	100.0%		100.0%
Raw variance explained by measures	=	80.0	71.8%		71.8%
Raw variance explained by persons	=	11.2	11.1%		11.1%
Raw Variance explained by items	=	22.2	17.1%		17.1%
Raw unexplained variance (total)	=	100.0	28.2%	100.0%	28.2%
Unexplned variance in 1st contrast	=	1.4	11.8%	2.8%	
Unexplned variance in 2nd contrast	=	1.5	11.4%	2.4%	
Unexplned variance in 3rd contrast	=	1.6	11.0%	2.3%	
Unexplned variance in 4th contrast	=	1.7	10.0%	2.1%	
Unexplned variance in 5th contrast	=	1.8	9.6%	1.9%	

The table 5 was interpreted by comparing the empirical values of the entries with the modeled value. It revealed that the total raw variance in observation agreed with the model value of 100%, raw variance explained by measures of 71.8% agreed with the model value of 71.8%, raw variance explained by persons of 11.1% agreed with the model value of 11.1%, and raw variance explained by items of 17.1% agreed with the model value of 17.1%. These values confirmed that the test has content and construct validity. The ratio of 11.8% of unexplained variance in 1st contrast to 17.1% of raw variance explained by items is 1.40. This seemed good since the 1st, 2nd and the 3rd were not supposed to be more than

2.0 if they were indicating unidimensionality according to Wright, 1997 and Linacre, 2009. This was further confirmed by a scree plot that is presented in figure 1.

Figure 1 - Standardized Residual Variance Scree Plot for MAT

```
        VARIANCE COMPONENT SCREE PLOT
       +--+--+--+--+--+--+--+--+--+--+--+
   100%+  T                             +
       |     M                          |
 V  63%+                                +
 A     |                                |
 R  40%+                                +
 I     |             U                  |
 A  25%+                                +
 N     |          I                     |
 C  16%+                 1 2            +
 E     |        P             3         |
    10%+                         4  5   +
 L     |                                |
 O   6%+                                +
 G     |                                |
 |   4%+                                +
 S     |                                |
 C   3%+                                +
 A     |                                |
 L   2%+                                +
 E     |                                |
 D   1%+                                +
       |                                |
   0.5%+                                +
       +--+--+--+--+--+--+--+--+--+--+--+
          TV MV PV IV UV U1 U2 U3 U4 U5

 VARIANCE COMPONENTS
```

The T is referred to the total raw variance in observation, M is raw variance explained by measure, U is the raw unexplained variance (total), I is raw variance explained by item on the plot graph, P is raw variance explained by person while 1, 2, 3, 4, 5, represented the unexplained variance in 1st, 2nd, 3rd, 4th and 5th contrast on the plot graph. Therefore, this confirmed that the data has both content and construct validity which indicate unidimensionality trait.

Table 6 also presented information for the respected items in wit using the infit and outfit of MNSQ and ZSTD indices.

Table 6- Validity of MAT using infit and outfit of MNSQ and ZSTD indices in wit

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
45	823	1499	53.10	.52	1.10	9.9	1.11	9.9	A-.20	.14	48.9	56.9	I0045
29	1217	1499	40.23	.66	1.03	.8	1.08	1.8	B-.03	.11	81.2	81.2	I0029
100	1123	1499	43.96	.60	1.04	1.3	1.07	2.2	C-.03	.12	74.9	74.9	I0100
32	1034	1499	46.96	.56	1.06	2.5	1.07	2.9	D-.07	.13	68.9	69.0	I0032
37	873	1499	51.71	.53	1.06	5.6	1.07	5.7	E-.08	.14	54.1	58.9	I0037
27	542	1499	60.90	.54	1.05	2.9	1.06	3.3	F-.03	.14	62.7	64.0	I0027
99	1155	1499	42.77	.62	1.03	.8	1.05	1.5	G .00	.12	77.1	77.0	I0099
26	493	1499	62.38	.55	1.04	2.2	1.05	2.5	H-.02	.14	67.1	67.1	I0026
7	427	1499	64.49	.58	1.04	1.5	1.05	1.9	I-.01	.13	71.6	71.5	I0007
93	1033	1499	46.99	.56	1.03	1.4	1.05	2.0	J .01	.13	68.8	68.9	I0093
50	805	1499	53.59	.52	1.04	5.1	1.05	5.2	K .00	.14	53.4	56.3	I0050
49	674	1499	57.17	.52	1.04	4.1	1.04	4.1	L .02	.14	53.8	57.3	I0049
2	1002	1499	47.96	.55	1.03	1.8	1.04	1.8	M .01	.13	66.8	66.8	I0002
92	1056	1499	46.25	.57	1.03	1.1	1.04	1.5	N .03	.13	70.4	70.4	I0092
23	750	1499	55.09	.52	1.03	4.7	1.04	4.6	O .02	.14	49.8	55.8	I0023
31	889	1499	51.26	.53	1.03	2.6	1.04	2.8	P .03	.14	58.7	59.8	I0031
74	1040	1499	46.77	.57	1.02	.9	1.03	1.4	Q .05	.13	69.4	69.4	I0074
35	626	1499	58.50	.53	1.03	2.5	1.03	2.8	R .04	.14	57.9	59.3	I0035
6	597	1499	59.31	.53	1.03	2.2	1.03	2.4	S .04	.14	60.1	60.8	I0006
17	587	1499	59.60	.53	1.03	2.0	1.03	2.2	T .05	.14	61.5	61.3	I0017
25	728	1499	55.69	.52	1.03	3.9	1.03	3.9	U .04	.14	51.8	55.9	I0025
28	578	1499	59.86	.54	1.03	2.0	1.03	2.0	V .04	.14	60.9	61.8	I0028
36	663	1499	57.47	.53	1.03	3.1	1.03	3.1	W .04	.14	53.5	57.6	I0036
16	780	1499	54.27	.52	1.03	3.5	1.03	3.4	X .05	.14	52.5	55.9	I0016
91	1043	1499	46.67	.57	1.02	.7	1.03	1.1	Y .07	.13	69.6	69.6	I0091
24	612	1499	58.89	.53	1.03	2.1	1.02	2.0	Z .06	.14	56.8	60.0	I0024
39	774	1499	54.44	.52	1.02	2.8	1.02	3.1	.07	.14	54.4	55.8	I0039
15	792	1499	53.95	.52	1.02	3.0	1.02	2.8	.06	.14	51.8	56.1	I0015
43	662	1499	57.50	.53	1.02	2.4	1.02	2.3	.07	.14	55.0	57.7	I0043
21	725	1499	55.77	.52	1.02	2.4	1.02	2.4	.08	.14	53.0	56.0	I0021
	BETTER	FITTING	OMITTED										
18	700	1499	56.45	.52	.98	-2.3	.98	-2.1	.20	.14	61.9	56.5	I0018
60	1068	1499	45.86	.58	.98	-.7	.97	-1.0	z .20	.13	71.2	71.2	I0060
30	712	1499	56.13	.52	.98	-2.4	.98	-2.4	y .21	.14	59.3	56.2	I0030
65	1150	1499	42.96	.62	.98	-.6	.97	-.9	x .19	.12	76.7	76.7	I0065
72	1145	1499	43.15	.61	.98	-.6	.96	-1.0	w .20	.12	76.4	76.4	I0072
59	1055	1499	46.28	.57	.98	-.9	.97	-1.4	v .21	.13	70.4	70.4	I0059
70	1122	1499	44.00	.60	.98	-.7	.97	-1.0	u .20	.12	74.8	74.8	I0070
42	426	1499	64.52	.58	.98	-.9	.98	-.9	t .21	.13	71.6	71.6	I0042
11	986	1499	48.44	.55	.98	-1.3	.98	-1.2	s .21	.13	66.2	65.8	I0011
55	1200	1499	40.96	.65	.98	-.6	.95	-1.2	r .21	.11	80.1	80.0	I0055
56	1143	1499	43.23	.61	.98	-.7	.96	-1.1	q .21	.12	76.3	76.2	I0056
83	1048	1499	46.51	.57	.97	-1.1	.97	-1.4	p .22	.13	70.0	69.9	I0083
62	1094	1499	44.98	.59	.97	-.9	.96	-1.3	o .22	.13	73.0	73.0	I0062
68	1135	1499	43.52	.61	.97	-.8	.95	-1.5	n .23	.12	75.7	75.7	I0068
69	1131	1499	43.67	.60	.97	-.9	.96	-1.2	m .22	.12	75.5	75.4	I0069
63	1103	1499	44.67	.59	.97	-1.0	.96	-1.3	l .22	.12	73.6	73.6	I0063
64	1163	1499	42.46	.62	.97	-.8	.94	-1.6	k .23	.12	77.6	77.6	I0064
38	652	1499	57.77	.53	.96	-4.2	.96	-4.1	j .28	.14	63.0	58.1	I0038
19	584	1499	59.68	.53	.96	-3.1	.96	-3.0	i .28	.14	64.6	61.5	I0019
20	974	1499	48.80	.55	.96	-2.5	.95	-2.9	h .29	.14	65.3	65.0	I0020
40	508	1499	61.92	.55	.96	-2.3	.95	-2.5	g .28	.14	66.2	66.1	I0040
47	775	1499	54.41	.52	.95	-6.1	.95	-6.1	f .30	.14	62.2	55.8	I0047
46	704	1499	56.34	.52	.95	-6.0	.95	-6.0	e .30	.14	63.4	56.4	I0046
4	753	1499	55.01	.52	.95	-6.8	.95	-6.9	d .31	.14	61.6	55.7	I0004
34	558	1499	60.43	.54	.95	-3.3	.95	-3.3	c .31	.14	66.0	63.0	I0034
41	666	1499	57.39	.53	.94	-7.1	.93	-7.1	b .36	.14	66.4	57.5	I0041
33	631	1499	58.36	.53	.93	-6.4	.93	-6.4	a .38	.14	66.0	59.0	I0033
MEAN	915.1	1499.0	50.00	.57	1.00	.1	1.00	.0			67.3	67.2	
S.D.	233.7	.0	7.21	.07	.03	2.5	.04	2.6			9.1	8.8	

The data were analyzed using winsteps Rasch software version 3.75. First of all, fit indices were examined closely to check the relevance of the items as part of content validity. Table 6 & 7 show the fit indices for some of the items. The items are arranged from difficult to easy in table 7. The first column, "entry number", indicates the number given to each item in the test (ranging from 1 to 100). The second column, labeled as "total score", represents the total score for each item (i.e. the number of testees who have responded correctly to that item). The number of testees who have attempted each item is given in the third column which is labeled as "total count". The difficulty estimates for the items are given in the fourth column labeled as "measure". The fifth column, "model S.E.", shows the standard error of the item difficulty measures. "MNSQ" and "ZSTD" are abbreviations for "mean square" and "z standardized distribution" respectively, and are provided for "outfit" as well as "infit" columns. A similar data under logit was presented in table 7

Table 7- Validity of MAT using infit and outfit of MNSQ and ZSTD indices in logit

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
42	426	1499	1.45	.06	.98	-.9	.98	-.9	.21	.13	71.6	71.6	I0042
7	427	1499	1.45	.06	1.04	1.5	1.05	1.9	-.01	.13	71.6	71.5	I0007
26	493	1499	1.24	.06	1.04	2.2	1.05	2.5	-.02	.14	67.1	67.1	I0026
40	508	1499	1.19	.06	.96	-2.3	.95	-2.5	.28	.14	66.2	66.1	I0040
27	542	1499	1.09	.05	1.05	2.9	1.06	3.3	-.03	.14	62.7	64.0	I0027
3	555	1499	1.05	.05	.98	-1.2	.98	-1.2	.20	.14	64.4	63.2	I0003
34	558	1499	1.04	.05	.95	-3.3	.95	-3.3	.31	.14	66.0	63.0	I0034
22	566	1499	1.02	.05	1.00	.1	1.00	.2	.13	.14	63.1	62.5	I0022
28	578	1499	.99	.05	1.03	2.0	1.03	2.0	.04	.14	60.9	61.8	I0028
8	579	1499	.98	.05	1.00	.3	1.01	.4	.12	.14	61.6	61.8	I0008
19	584	1499	.97	.05	.96	-3.1	.96	-3.0	.28	.14	64.6	61.5	I0019
17	587	1499	.96	.05	1.03	2.0	1.03	2.2	.05	.14	61.5	61.3	I0017
6	597	1499	.93	.05	1.03	2.2	1.03	2.4	.04	.14	60.1	60.8	I0006
24	612	1499	.89	.05	1.03	2.1	1.02	2.0	.06	.14	56.8	60.0	I0024
35	626	1499	.85	.05	1.03	2.5	1.03	2.8	.04	.14	57.9	59.3	I0035
33	631	1499	.84	.05	.93	-6.4	.93	-6.4	.38	.14	66.0	59.0	I0033
38	652	1499	.78	.05	.96	-4.2	.96	-4.1	.28	.14	63.0	58.1	I0038
43	662	1499	.75	.05	1.02	2.4	1.02	2.3	.07	.14	55.0	57.7	I0043
36	663	1499	.75	.05	1.03	3.1	1.03	3.1	.04	.14	53.5	57.6	I0036
41	666	1499	.74	.05	.94	-7.1	.93	-7.1	.36	.14	66.4	57.5	I0041
9	674	1499	.72	.05	.99	-1.0	.99	-1.0	.17	.14	57.6	57.3	I0009
49	674	1499	.72	.05	1.04	4.1	1.04	4.1	.02	.14	53.8	57.3	I0049
14	698	1499	.65	.05	1.00	.5	1.00	.6	.13	.14	55.6	56.5	I0014
18	700	1499	.65	.05	.98	-2.3	.98	-2.1	.20	.14	61.9	56.5	I0018
12	702	1499	.64	.05	.99	-1.7	.99	-1.6	.19	.14	58.9	56.4	I0012
46	704	1499	.63	.05	.95	-6.0	.95	-6.0	.30	.14	63.4	56.4	I0046
30	712	1499	.61	.05	.98	-2.4	.98	-2.4	.21	.14	59.3	56.2	I0030
21	725	1499	.58	.05	1.02	2.4	1.02	2.4	.08	.14	53.0	56.0	I0021
25	728	1499	.57	.05	1.03	3.9	1.03	3.9	.04	.14	51.8	55.9	I0025
5	738	1499	.54	.05	.99	-1.1	.99	-1.0	.17	.14	59.0	55.8	I0005
1	746	1499	.52	.05	.99	-.9	.99	-1.1	.17	.14	54.2	55.8	I0001
23	750	1499	.51	.05	1.03	4.7	1.04	4.6	.02	.14	49.8	55.8	I0023
4	753	1499	.50	.05	.95	-6.8	.95	-6.9	.31	.14	61.6	55.7	I0004
39	774	1499	.44	.05	1.02	2.8	1.02	3.1	.07	.14	54.4	55.8	I0039
47	775	1499	.44	.05	.95	-6.1	.95	-6.1	.30	.14	62.2	55.8	I0047
16	780	1499	.43	.05	1.03	3.5	1.03	3.4	.05	.14	52.5	55.9	I0016
10	785	1499	.41	.05	1.01	1.2	1.01	1.0	.11	.14	54.4	56.0	I0010
15	792	1499	.39	.05	1.02	3.0	1.02	2.8	.06	.14	51.8	56.1	I0015
44	792	1499	.39	.05	.99	-1.6	.99	-1.6	.19	.14	57.4	56.1	I0044
13	804	1499	.36	.05	1.00	-.4	1.00	-.5	.15	.14	56.0	56.3	I0013
50	805	1499	.36	.05	1.04	5.1	1.05	5.2	.00	.14	53.4	56.3	I0050
45	823	1499	.31	.05	1.10	9.9	1.11	9.9	-.20	.14	48.9	56.9	I0045
37	873	1499	.17	.05	1.06	5.6	1.07	5.7	-.08	.14	54.1	58.9	I0037
31	889	1499	.13	.05	1.03	2.6	1.04	2.8	.03	.14	58.7	59.8	I0031
48	958	1499	-.07	.05	.99	-.3	.99	-.4	.16	.14	63.6	64.0	I0048
20	974	1499	-.12	.05	.96	-2.5	.95	-2.9	.29	.14	65.3	65.0	I0020
11	986	1499	-.16	.05	.98	-1.3	.98	-1.2	.21	.13	66.2	65.8	I0011
80	995	1499	-.18	.06	1.00	.1	1.01	.3	.12	.13	66.4	66.4	I0080
2	1002	1499	-.20	.06	1.03	1.8	1.04	1.8	.01	.13	66.8	66.8	I0002
79	1002	1499	-.20	.06	.99	-.7	.98	-.9	.19	.13	67.0	66.8	I0079
76	1027	1499	-.28	.06	1.00	.0	1.00	-.1	.14	.13	68.6	68.5	I0076
93	1033	1499	-.30	.06	1.03	1.4	1.05	2.0	.01	.13	68.8	68.9	I0093
32	1034	1499	-.30	.06	1.06	2.5	1.07	2.9	-.07	.13	68.9	69.0	I0032
74	1040	1499	-.32	.06	1.02	.9	1.03	1.4	.05	.13	69.4	69.4	I0074
94	1042	1499	-.33	.06	1.02	.8	1.02	1.0	.06	.13	69.6	69.5	I0094
81	1043	1499	-.33	.06	1.00	-.2	1.00	-.2	.14	.13	69.5	69.6	I0081
91	1043	1499	-.33	.06	1.02	.7	1.03	1.1	.07	.13	69.6	69.6	I0091
83	1048	1499	-.35	.06	.97	-1.1	.97	-1.4	.22	.13	70.0	69.9	I0083
88	1049	1499	-.35	.06	.99	-.3	.99	-.4	.15	.13	69.9	70.0	I0088
98	1053	1499	-.37	.06	1.02	.7	1.02	.9	.07	.13	70.2	70.2	I0098

59	1055	1499	-.37	.06	.98	-.9	.97	-1.4	.21	.13	70.4	70.4	I0059
92	1056	1499	-.37	.06	1.03	1.1	1.04	1.5	.03	.13	70.4	70.4	I0092
73	1057	1499	-.38	.06	1.00	.2	1.01	.2	.11	.13	70.5	70.5	I0073
95	1065	1499	-.40	.06	1.00	.0	1.00	.0	.13	.13	71.0	71.0	I0095
60	1068	1499	-.41	.06	.98	-.7	.97	-1.0	.20	.13	71.2	71.2	I0060
86	1070	1499	-.42	.06	.99	-.2	.99	-.4	.15	.13	71.4	71.4	I0086
58	1072	1499	-.43	.06	.99	-.5	.98	-.7	.18	.13	71.5	71.5	I0058
87	1073	1499	-.43	.06	.99	-.5	.98	-.6	.17	.13	71.6	71.6	I0087
75	1081	1499	-.46	.06	.99	-.4	.99	-.4	.16	.13	72.1	72.1	I0075
82	1081	1499	-.46	.06	.99	-.3	.99	-.3	.16	.13	72.1	72.1	I0082
61	1082	1499	-.46	.06	.99	-.6	.97	-1.0	.19	.13	72.2	72.2	I0061
78	1090	1499	-.49	.06	1.00	-.1	.99	-.2	.14	.13	72.7	72.7	I0078
85	1090	1499	-.49	.06	.98	-.6	.98	-.8	.18	.13	72.7	72.7	I0085
62	1094	1499	-.50	.06	.97	-.9	.96	-1.3	.22	.13	73.0	73.0	I0062
57	1095	1499	-.51	.06	.99	-.3	.98	-.6	.16	.13	73.0	73.0	I0057
77	1099	1499	-.52	.06	1.01	.4	1.02	.6	.08	.12	73.3	73.3	I0077
63	1103	1499	-.53	.06	.97	-1.0	.96	-1.3	.22	.12	73.6	73.6	I0063
90	1105	1499	-.54	.06	1.01	.2	1.01	.3	.10	.12	73.7	73.7	I0090
97	1110	1499	-.56	.06	1.01	.4	1.01	.4	.09	.12	74.0	74.0	I0097
84	1116	1499	-.58	.06	.99	-.2	.99	-.4	.15	.12	74.4	74.4	I0084
70	1122	1499	-.60	.06	.98	-.7	.97	-1.0	.20	.12	74.8	74.8	I0070
100	1123	1499	-.60	.06	1.04	1.3	1.07	2.2	-.03	.12	74.9	74.9	I0100
69	1131	1499	-.63	.06	.97	-.9	.96	-1.2	.22	.12	75.5	75.4	I0069
96	1133	1499	-.64	.06	1.01	.5	1.02	.6	.07	.12	75.6	75.6	I0096
68	1135	1499	-.65	.06	.97	-.8	.95	-1.5	.23	.12	75.7	75.7	I0068
89	1137	1499	-.66	.06	1.00	.1	1.01	.2	.11	.12	75.9	75.8	I0089
56	1143	1499	-.68	.06	.98	-.7	.96	-1.1	.21	.12	76.3	76.2	I0056
72	1145	1499	-.68	.06	.98	-.6	.96	-1.0	.20	.12	76.4	76.4	I0072
65	1150	1499	-.70	.06	.98	-.6	.97	-.9	.19	.12	76.7	76.7	I0065
99	1155	1499	-.72	.06	1.03	.8	1.05	1.5	.00	.12	77.1	77.0	I0099
67	1160	1499	-.74	.06	.98	-.5	.97	-.9	.19	.12	77.4	77.4	I0067
64	1163	1499	-.75	.06	.97	-.8	.94	-1.6	.23	.12	77.6	77.6	I0064
71	1165	1499	-.76	.06	.99	-.2	.99	-.2	.14	.12	77.7	77.7	I0071
55	1200	1499	-.90	.07	.98	-.6	.95	-1.2	.21	.11	80.1	80.0	I0055
66	1205	1499	-.92	.07	.98	-.4	.96	-.9	.18	.11	80.4	80.4	I0066
29	1217	1499	-.98	.07	1.03	.8	1.08	1.8	-.03	.11	81.2	81.2	I0029
54	1280	1499	-1.28	.07	.98	-.3	.96	-.8	.18	.10	85.4	85.4	I0054
53	1319	1499	-1.51	.08	.99	-.2	.96	-.7	.15	.09	88.0	88.0	I0053
52	1351	1499	-1.73	.09	1.00	.0	1.00	.0	.09	.08	90.1	90.1	I0052
51	1377	1499	-1.95	.09	.99	-.1	.95	-.5	.12	.08	91.9	91.9	I0051
MEAN	915.1	1499.0	.00	.06	1.00	.1	1.00	.0			67.3	67.2	
S.D.	233.7	.0	.72	.01	.03	2.5	.04	2.6			9.1	8.8	

The acceptable values range from 0.7 to 1.1 for 'MNSQ' since the sample used was greater than (>) 1000 and -2 and +2 for 'ZSTD'. The table 7 shows that item 7 and 42 are the most difficulty item on the test. From 1499 testees who have attempted these items. Only 427 and 426 could get it right respectively. The difficulty index of the items is estimated to be 64.52wit or 1.45logit with standard error of .58 or .060 respectively. This means that one can be 95% sure that the true value for the difficulty of this item lies somewhere between -1.95logit and 1.45logit, i.e two S.E's below and above the observed measure.

The infit and outfit indices for both MNSQ and ZSTD are within the acceptable range, thus not causing a serious problem. Table 7 indicates that 33 items should either be deleted or revised because of lack of fit to the model. Such items are 4, 6, 15, 16, 17, 18, 19, 20, 21, 23, 24, 25, 26, 27, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 43, 45, 46, 47, 49, 50 and 100. These items are measuring something other than the intended content and construct. That is, they are construct irrelevant. Hence, the 67 items have both construct and content validity. Therefore using the Rasch model, it was found that the MAT items have content and construct validity.

Research Question 2

What is the reliability of the Mathematics Achievement Tests (MAT) items using the Rasch Model?

To answer the research question, the level of item data fit to the Rasch model of 1499 measured person with that of 100 and 65 measured items were both considered here. The person [Cronbach's alpha (KR-20)] reliability estimate of the 100 test was 0.53 (table 3 & 4) which was slightly moderate. This moderate reliability was due to the narrow spread of the person's ability in the analysis. Table 3 showed the raw score standard deviation of the sample 6.1 out of 100, which was a very narrow spread of person indeed. When the item with negative point-measure correlation indices were deleted from the test, the reliability increased to 0.78 for 65 test items (table 8). The items with negative PT measure correlation indices are eight in numbers. They are 7, 26, 27, 29, 32, 37, 45 and 100. The table 7 showed the items with negative PTMC while the reliability table 8 & 9 showed the increased in the person reliability when the negative PTMC were removed.

Table 8 – Reliability table of 65 MAT ITEMS (Person – units in wit)

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	45.2	65.0	52.11	2.79	1.00	.0	1.00	.0
S.D.	34.9	.0	5.81	.19	.10	.9	.15	1.0
MAX.	62.0	65.0	67.19	4.76	1.34	5.3	1.86	5.3
MIN.	27.0	65.0	40.59	2.56	.70	-2.8	.55	-2.7

REAL RMSE 2.85 TRUE SD 2.53 SEPARATION 2.89 Person RELIABILITY .74
MODEL RMSE 2.79 TRUE SD 2.59 SEPARATION 2.93 Person RELIABILITY .78
S.E. OF Person MEAN = .10

Person RAW SCORE-TO-MEASURE CORRELATION = 1.00
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .78

The separation index of the person is 2.93 which translates to a person strata index of 5.30. Person strata index indicates the number of distinct ability levels which can be identified by the test. The minimum person strata index is 2, which means that the test is capable of distinguishing at least 2 strata of persons, namely, high-ability and low-ability persons. For a strata index of 2, a separation index of at least 1.0 is needed. A reliability index of at least 0.00logit or 50.0wit is required for a separation index of 1.0.

Table 9 – Reliability table of 65 MAT ITEMS (Person – units in logit)

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	45.2	65.0	.51	.28	1.00	.0	1.00	.0
S.D.	34.9	.0	.38	.02	.10	.9	.15	1.0
MAX.	62.0	65.0	2.72	.48	1.34	5.3	1.86	5.3
MIN.	27.0	65.0	-.44	.26	.70	-2.8	.55	-2.7

REAL RMSE	.28	TRUE SD	2.25	SEPARATION	2.89	Person RELIABILITY	.74
MODEL RMSE	.28	TRUE SD	2.26	SEPARATION	2.93	Person RELIABILITY	.78
S.E. OF Person MEAN = .10							

Person RAW SCORE-TO-MEASURE CORRELATION = 1.00
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .78

It should be noted that the moderate reliability, separation, and strata indices for 100 test is due to the low standard deviation of persons' abilities (2.91wit or 0.29logit) while the high reliability, separation and strata indices for the 65 test items was due to the high standard deviation of person's ability (5.81wit or .38logit) . If another sample with a wider spread of abilities were to be tested, these statistics would improve still. Also, the result presented in logit form in table 9 shows the same with the one presented in table 8 in wit.

Table 10 – Reliability table of the 65 MAT items (item – units in wit)

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	1011.1	1499.0	50.00	.57	1.00	.0	1.00	.0
S.D.	209.1	.0	7.21	.07	.03	2.5	.05	1.6
MAX.	1377.0	1499.0	64.52	.95	1.10	8.7	1.11	8.7
MIN.	426.0	1499.0	30.53	.52	.93	-7.1	.93	-7.1

REAL RMSE	.58	TRUE SD	7.19	SEPARATION	13.11	Item RELIABILITY	.99
MODEL RMSE	.58	TRUE SD	7.19	SEPARATION	13.17	Item RELIABILITY	.99
S.E. OF Item MEAN = .72							

The table 10 showed the summary statistics of the 65 measured items. This investigated the representativeness of the items by checking the value given for item strata, item separation and item reliability. The item strata is 8.7, item separation is 13.17 while item reliability is 0.99. The reliability for the items was very good. That is, the chances that the difficulty ordering of the items be repeated if the test items were given to another group is extremely high. Thus, one can rely on the representativeness and reliability of the test items. Therefore, the reliability of the MAT items using the Rasch model was 0.99

Table 11 - Reliability table of the 65 MAT items (items - units in logit)

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	1011.1	1499.0	.00	.06	1.00	.0	1.00	.0
S.D.	209.1	.0	.67	.01	.03	1.3	.05	1.6
MAX.	1377.0	1499.0	1.76	.09	1.07	8.7	1.11	8.7
MIN.	426.0	1499.0	-1.67	.05	.95	-6.6	.90	-6.6

REAL RMSE	.06	TRUE SD	.67	SEPARATION 13.11	Item	RELIABILITY .99
MODEL RMSE	.06	TRUE SD	.67	SEPARATION 13.17	Item	RELIABILITY .99
S.E. OF Item MEAN = .08						

The table 11 has equally reported the same value of result for the 65 items in logit as reported in table 10 and there are correspondence of information between the two tables..

Research Question three

What are the difficulty indices of the Mathematics achievement Test (MAT) items using the Rasch model?

To answer the Research question, the table 12 and 13 are considered. Table 12 was expressed in wit while table 13 was expressed in logit. The difficulty estimates or indices for the 65 items of MAT were given in the fourth column labeled as **"measure"**.

TABLE 12- Difficulty Indices of 65 MAT Items (MEASURE ORDER- unit in wit)

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
99	1155	1499	42.77	.62	1.03	.8	1.05	1.5	G .00	.12	77.1	77.0	I0099
93	1033	1499	46.99	.56	1.03	1.4	1.05	2.0	J .01	.13	68.8	68.9	I0093
2	1002	1499	47.96	.55	1.03	1.8	1.04	1.8	M .01	.13	66.8	66.8	I0002
92	1056	1499	46.25	.57	1.03	1.1	1.04	1.5	N .03	.13	70.4	70.4	I0092
74	1040	1499	46.77	.57	1.02	.9	1.03	1.4	Q .05	.13	69.4	69.4	I0074
28	578	1499	59.86	.54	1.03	2.0	1.03	2.0	V .04	.14	60.9	61.8	I0028
91	1043	1499	46.67	.57	1.02	.7	1.03	1.1	Y .07	.13	69.6	69.6	I0091
	BETTER	FITTING	OMITTED										
60	1068	1499	45.86	.58	.98	-.7	.97	-1.0	z .20	.13	71.2	71.2	I0060
65	1150	1499	42.96	.62	.98	-.6	.97	-.9	x .19	.12	76.7	76.7	I0065
72	1145	1499	43.15	.61	.98	-.6	.96	-1.0	w .20	.12	76.4	76.4	I0072
59	1055	1499	46.28	.57	.98	-.9	.97	-1.4	v .21	.13	70.4	70.4	I0059
70	1122	1499	44.00	.60	.98	-.7	.97	-1.0	u .20	.12	74.8	74.8	I0070
42	426	1499	64.52	.58	.98	-.9	.98	-.9	t .21	.13	71.6	71.6	I0042
11	986	1499	48.44	.55	.98	-1.3	.98	-1.2	s .21	.13	66.2	65.8	I0011
55	1200	1499	40.96	.65	.98	-.6	.95	-1.2	r .21	.11	80.1	80.0	I0055
56	1143	1499	43.23	.61	.98	-.7	.96	-1.1	q .21	.12	76.3	76.2	I0056
83	1048	1499	46.51	.57	.97	-1.1	.97	-1.4	p .22	.13	70.0	69.9	I0083
62	1094	1499	44.98	.59	.97	-.9	.96	-1.3	o .22	.13	73.0	73.0	I0062
68	1135	1499	43.52	.61	.97	-.8	.95	-1.5	n .23	.12	75.7	75.7	I0068
69	1131	1499	43.67	.60	.97	-.9	.96	-1.2	m .22	.12	75.5	75.4	I0069
63	1103	1499	44.67	.59	.97	-1.0	.96	-1.3	l .22	.12	73.6	73.6	I0063
64	1163	1499	42.46	.62	.97	-.8	.94	-1.6	k .23	.12	77.6	77.6	I0064
MEAN	1011.1	1499.0	50.00	.57	1.00	.0	1.00	.0			67.2	67.2	
S.D.	209.1	.0	7.21	.07	.03	2.5	.04	2.6			9.1	8.8	

Table 13 showed the difficulty indices of 65 MAT items (measure order) in logit unit. The difficulty estimates or indices for the items MAT were given in the fourth column labeled as **"measure"** too**.** The MAT items were arranged from difficult to easy in table 13.

TABLE 13- Difficulty Indices of 65 MAT Items (MEASURE ORDER- unit in logit)

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
42	426	1499	1.45	.06	.98	-.9	.98	-.9	.21	.13	71.6	71.6	I0042
3	555	1499	1.05	.05	.98	-1.2	.98	-1.2	.20	.14	64.4	63.2	I0003
22	566	1499	1.02	.05	1.00	.1	1.00	.2	.13	.14	63.1	62.5	I0022
28	578	1499	.99	.05	1.03	2.0	1.03	2.0	.04	.14	60.9	61.8	I0028
8	579	1499	.98	.05	1.00	.3	1.01	.4	.12	.14	61.6	61.8	I0008
9	674	1499	.72	.05	.99	-1.0	.99	-1.0	.17	.14	57.6	57.3	I0009
14	698	1499	.65	.05	1.00	.5	1.00	.6	.13	.14	55.6	56.5	I0014
12	702	1499	.64	.05	.99	-1.7	.99	-1.6	.19	.14	58.9	56.4	I0012
5	738	1499	.54	.05	.99	-1.1	.99	-1.0	.17	.14	59.0	55.8	I0005
1	746	1499	.52	.05	.99	-.9	.99	-1.1	.17	.14	54.2	55.8	I0001
10	785	1499	.41	.05	1.01	1.2	1.01	1.0	.11	.14	54.4	56.0	I0010
44	792	1499	.39	.05	.99	-1.6	.99	-1.6	.19	.14	57.4	56.1	I0044
13	804	1499	.36	.05	1.00	-.4	1.00	-.5	.15	.14	56.0	56.3	I0013
48	958	1499	-.07	.05	.99	-.3	.99	-.4	.16	.14	63.6	64.0	I0048
11	986	1499	-.16	.05	.98	-1.3	.98	-1.2	.21	.13	66.2	65.8	I0011
80	995	1499	-.18	.06	1.00	.1	1.01	.3	.12	.13	66.4	66.4	I0080
2	1002	1499	-.20	.06	1.03	1.8	1.04	1.8	.01	.13	66.8	66.8	I0002
79	1002	1499	-.20	.06	.99	-.7	.98	-.9	.19	.13	67.0	66.8	I0079
76	1027	1499	-.28	.06	1.00	.0	1.00	-.1	.14	.13	68.6	68.5	I0076
93	1033	1499	-.30	.06	1.03	1.4	1.05	2.0	.01	.13	68.8	68.9	I0093
74	1040	1499	-.32	.06	1.02	.9	1.03	1.4	.05	.13	69.4	69.4	I0074
94	1042	1499	-.33	.06	1.02	.8	1.02	1.0	.06	.13	69.6	69.5	I0094
81	1043	1499	-.33	.06	1.00	-.2	1.00	-.2	.14	.13	69.5	69.6	I0081
91	1043	1499	-.33	.06	1.02	.7	1.03	1.1	.07	.13	69.6	69.6	I0091
83	1048	1499	-.35	.06	.97	-1.1	.97	-1.4	.22	.13	70.0	69.9	I0083
88	1049	1499	-.35	.06	.99	-.3	.99	-.4	.15	.13	69.9	70.0	I0088
98	1053	1499	-.37	.06	1.02	.7	1.02	.9	.07	.13	70.2	70.2	I0098
59	1055	1499	-.37	.06	.98	-.9	.97	-1.4	.21	.13	70.4	70.4	I0059
92	1056	1499	-.37	.06	1.03	1.1	1.04	1.5	.03	.13	70.4	70.4	I0092
73	1057	1499	-.38	.06	1.00	.2	1.01	.2	.11	.13	70.5	70.5	I0073
95	1065	1499	-.40	.06	1.00	.0	1.00	.0	.13	.13	71.0	71.0	I0095
60	1068	1499	-.41	.06	.98	-.7	.97	-1.0	.20	.13	71.2	71.2	I0060
86	1070	1499	-.42	.06	.99	-.2	.99	-.4	.15	.13	71.4	71.4	I0086
58	1072	1499	-.43	.06	.99	-.5	.98	-.7	.18	.13	71.5	71.5	I0058
87	1073	1499	-.43	.06	.99	-.5	.98	-.6	.17	.13	71.6	71.6	I0087
75	1081	1499	-.46	.06	.99	-.4	.99	-.4	.16	.13	72.1	72.1	I0075
82	1081	1499	-.46	.06	.99	-.3	.99	-.3	.16	.13	72.1	72.1	I0082
61	1082	1499	-.46	.06	.99	-.6	.97	-1.0	.19	.13	72.2	72.2	I0061
78	1090	1499	-.49	.06	1.00	-.1	.99	-.2	.14	.13	72.7	72.7	I0078
85	1090	1499	-.49	.06	.98	-.6	.98	-.8	.18	.13	72.7	72.7	I0085
62	1094	1499	-.50	.06	.97	-.9	.96	-1.3	.22	.13	73.0	73.0	I0062
57	1095	1499	-.51	.06	.99	-.3	.98	-.6	.16	.13	73.0	73.0	I0057
77	1099	1499	-.52	.06	1.01	.4	1.02	.6	.08	.12	73.3	73.3	I0077
63	1103	1499	-.53	.06	.97	-1.0	.96	-1.3	.22	.12	73.6	73.6	I0063
90	1105	1499	-.54	.06	1.01	.2	1.01	.3	.10	.12	73.7	73.7	I0090
97	1110	1499	-.56	.06	1.01	.4	1.01	.4	.09	.12	74.0	74.0	I0097
84	1116	1499	-.58	.06	.99	-.2	.99	-.4	.15	.12	74.4	74.4	I0084
70	1122	1499	-.60	.06	.98	-.7	.97	-1.0	.20	.12	74.8	74.8	I0070
69	1131	1499	-.63	.06	.97	-.9	.96	-1.2	.22	.12	75.5	75.4	I0069
96	1133	1499	-.64	.06	1.01	.5	1.02	.6	.07	.12	75.6	75.6	I0096
68	1135	1499	-.65	.06	.97	-.8	.95	-1.5	.23	.12	75.7	75.7	I0068
89	1137	1499	-.66	.06	1.00	.1	1.01	.2	.11	.12	75.9	75.8	I0089
56	1143	1499	-.68	.06	.98	-.7	.96	-1.1	.21	.12	76.3	76.2	I0056
72	1145	1499	-.68	.06	.98	-.6	.96	-1.0	.20	.12	76.4	76.4	I0072
65	1150	1499	-.70	.06	.98	-.6	.97	-.9	.19	.12	76.7	76.7	I0065
99	1155	1499	-.72	.06	1.03	.8	1.05	1.5	.00	.12	77.1	77.0	I0099
67	1160	1499	-.74	.06	.98	-.5	.97	-.9	.19	.12	77.4	77.4	I0067
64	1163	1499	-.75	.06	.97	-.8	.94	-1.6	.23	.12	77.6	77.6	I0064
71	1165	1499	-.76	.06	.99	-.2	.99	-.2	.14	.12	77.7	77.7	I0071
55	1200	1499	-.90	.07	.98	-.6	.95	-1.2	.21	.11	80.1	80.0	I0055
66	1205	1499	-.92	.07	.98	-.4	.96	-.9	.18	.11	80.4	80.4	I0066
54	1280	1499	-1.28	.07	.98	-.3	.96	-.8	.18	.10	85.4	85.4	I0054
53	1319	1499	-1.51	.08	.99	-.2	.96	-.7	.15	.09	88.0	88.0	I0053
52	1351	1499	-1.73	.09	1.00	.0	1.00	.0	.09	.08	90.1	90.1	I0052
51	1377	1499	-1.95	.09	.99	-.1	.95	-.5	.12	.08	91.9	91.9	I0051
MEAN	1011.1	1499.0	.00	.06	1.00	.0	1.00	.0			67.2	67.2	
S.D.	209.1	.0	.72	.01	.03	2.5	.04	2.6			9.1	8.8	

Therefore, using the Rasch model, the difficulty index ranges between the value of -1.95logits to 1.45logits.

Research Question four

What items are the distracters of Mathematics Achievement Test (MAT) using the Rasch Model?

Table 14, the "Item Category/Option/Distracter Table" reveals helpful information regarding **substantive aspect** of construct validity. Item Distracter Table for 9 items is given below.

Table 14-Item CATEGORY/OPTION/DISTRACTOR FREQUENCIES: MISFIT ORDER unit in wit

ENTRY NUMBER	DATA CODE	SCORE VALUE	DATA COUNT	%	AVERAGE ABILITY	S.E. MEAN	OUTF MNSQ	PTMEA CORR.	Item
45 A	0	0	676	45	55.73	.12	1.1	.20	I0045
	1	1	823	55	54.59*	.09	1.1	-.20	
29 B	0	0	282	19	55.28	.18	1.1	.03	I0029
	1	1	1217	81	55.06*	.08	1.0	-.03	
100 C	0	0	376	25	55.26	.16	1.1	.03	I0100
	1	1	1123	75	55.05*	.09	1.0	-.03	
32 D	0	0	465	31	55.39	.12	1.1	.07	I0032
	1	1	1034	69	54.98*	.10	1.0	-.07	
37 E	0	0	626	42	55.37	.11	1.1	.08	I0037
	1	1	873	58	54.91*	.10	1.1	-.08	
27 F	0	0	957	64	55.18	.09	1.0	.03	I0027
	1	1	542	36	54.98*	.13	1.1	-.03	
99 G	0	0	344	23	55.08	.16	1.1	.00	I0099
	1	1	1155	77	55.11	.09	1.0	.00	
26 H	0	0	1006	67	55.14	.09	1.0	.02	I0026
	1	1	493	33	55.03*	.13	1.1	-.02	
7 I	0	0	1072	72	55.12	.09	1.0	.01	I0007
	1	1	427	28	55.06*	.13	1.1	-.01	
93 J	0	0	466	31	55.05	.14	1.1	-.01	I0093
	1	1	1033	69	55.13	.09	1.0	.01	
50 K	0	0	694	46	55.12	.11	1.1	.00	I0050
	1	1	805	54	55.09*	.10	1.0	.00	
49 L	0	0	825	55	55.07	.10	1.0	-.02	I0049
	1	1	674	45	55.15	.11	1.0	.02	
2 M	0	0	497	33	55.05	.12	1.0	-.01	I0002
	1	1	1002	67	55.13	.10	1.0	.01	

- Average ability does not ascend with category score

The first column shows the entry number for each item. The second column, "data code", indicates the codes given to the item. That is, 0 represented the failed choice, 1 the passed choice. (.) represented the missing items, i.e. the cases wherein none of the items was chosen. "score value" column shows the correct option (pass) by coding it as 1 and the other incorrect options (fail) as 0. The fourth column, "data count", indicated the number as well as the percentage of the participants who had chosen a particular option, be it right/pass or wrong/fail. "average measure/ability" or choice mean shows the mean of the ability estimates of all the testees who had chosen a particular option. It is expected that the value for average ability be the highest for the correct choice items (1) and lower for incorrect choice items (0). There is an asterisk placed above the average ability for correct items (1) in cases where this expectation is not met. "S. E. mean", is an abbreviation for standard error of the mean which is the standard error of the mean of the ability estimates of those testees who had chosen a particular option. Finally, "PTMEA CORR." showed the correlation between the occurrence and non-occurrence of each option and the ability estimates of the testees choosing a particular option.

Although all the information provided by the table is helpful, special attention is given to average ability. Items whose correct options are marked with asterisks (as a sign of flagging unacceptable values) should be checked. Those which have good fit indices and also the average measures for their wrong options are smaller than the average measure for their correct option are kept. However, those manifesting poor fit and those with greater average measures for wrong options than the correct option should be revised or deleted. Putting aside the items which do not fit the model, items 45, **29**, 100, 32, 37, 27, 26, **7** and 50 have the asterisk above their correct options. This means that the mean of the persons who have chosen the right option is not greater than the means of those who have chosen the wrong options. This indicates that these distracters do not function in the expected fashion.

Distracter analysis showed that the distracters of most items acted in the intended way, i.e. elicited responses consistent with the intended cognitive processes, to a great degree. This is how multiple choice distracter analysis provides empirical evidence for the substantive aspect of construct validity. Therefore, using the Rasch model, the items such as 7, 26, 27, 29, 32, 37, 45, 50 and 100 with asterick above their correct options are the poordistracters of MAT items. They did not fit into the model properly.

Research Question five

What is the person-item-map of Mathematics Achievement Test (MAT) items using the Rasch model?

Figure 2 showed the person-item-map in logit. Person-item-map served to provide evidence for the representativeness of the test items to show whether all part of the construct are well covered by the test. It was equally to indicate that the items were spread over the entire range of the scale i.e all parts of the construct were well covered by the test. The numbers on the right indicated items and # on the left indicated persons. Rasch item difficulty and person ability measures were therefore computed. This figure plotted person ability against item difficulty. The distribution of persons was consistent, making a curve-like shape which peaked around the mean. The person ability and item difficulty mean estimates were 0.51logit and 0.00logit while the SD indices for persons and items were 0.25logit and 0.72logit respectively.

Figure 2 – Person-item-Map

```
MEASURE(logit) Person - MAP - Item
              <more>|<rare>
    2               +
                    |
                    |
                    |
                    |
                    |
                    |
                    |
                 .  |T I0007  I0042
                 .  |
                    |
                    |  I0026
                 .  |  I0040
               .##  |
                .# T|  I0003  I0027  I0034
    1           .#  +  I0008  I0019  I0022  I0028
               .##  |  I0006  I0017
            .#####  |  I0024  I0033  I0035
              .### S|  I0038
             .####  |S I0009  I0036  I0041  I0043  I0049
        .#########  |  I0012  I0014  I0018  I0046
             .####  |  I0021  I0025  I0030
     .############ M|  I0001  I0004  I0005  I0023
           .######  |  I0039  I0047
            .#####  |  I0010  I0015  I0016  I0044
       .##########  |  I0013  I0045  I0050
              .###  |
         .######## S|  I0037
              .###  |  I0031
               ###  |
    0           .#  +M
                .# T|  I0048
                 .  |  I0011  I0020
                 .  |  I0002  I0079  I0080
                    |  I0076
                 .  |  I0032  I0074  I0081  I0083  I0088  I0091  I0093  I0094
                       I0098
                 .  |  I0058  I0059  I0060  I0073  I0086  I0087  I0092  I0095
                    |  I0061  I0075  I0078  I0082  I0085
                 .  |  I0057  I0062  I0063  I0077  I0090  I0097
                    |  I0069  I0070  I0084  I0100
                 .  |  I0056  I0068  I0072  I0089  I0096
                    |S I0064  I0065  I0067  I0071  I0099
                    |
                    |
                 .  |  I0055  I0066
   -1               +  I0029
                 .  |
                    |                I00= Item code
                    |                S= 1 standard deviation away from mean
                    |  I0054         S= 2 standard deviation away from mean
                    |                M= Mean Achievement/difficulty
                    |T
                    |  I0053         EACH "#" IS 16 testees
                    |                EACH "." IS 1 to 15 testees
                    |
                    |
                    |  I0052
                    |
                    |
                    |  I0051
   -2               +
              <less>|<frequent>
 EACH "#" IS 16. EACH "." IS 1 TO 15
```

The map shows that some of the items on the right lower part did not match to the persons on the left, indicating that the items were not appropriate for this group of testees, though they indicated good fit to the model. Four (4) of these items at the lower part may be omitted since they were too easy for the testees and in fact useless since there were no testees at that ability level.

Items and persons placed on top of the scale were more difficulty and more competent respectively. As one can see, all testees were clustered towards the centre of the scale and the items were spread all over the scale. The map showed that there were enough items in the region of the scale where testees lie and this part of the scale was pretty well covered by items. Therefore, the person abilities were estimated quite precisely as was evident from the low root mean square standard error of the persons which was 2.19wit or 0.22logit. Therefore, the accurate reliability of the test was due to an actual homogeneity in the persons with respect to the item difficulty. Overall, the items showed acceptable degree of representativeness. The person-item-map of MAT items using the Rasch model showed the items were spread all over the scale with the testees ability clustered towards the centre of the scale.

Research Question 6

What are the difficulty indices of the Mathematics Achievement Test (MAT) items for male and female testees using the Rasch model?

The table 15 showed the difficulty indices of the Mathematics Achievement test items for the male and female using the Rasch model. The difficulty indices of the male ranges from -1.98logits to 1.55logit while that of female ranges from -1.92logits to 1.58logits. The third column from the subgroup indicates the difficulty indices which is named DIF measure.

Table 15 Difficulty indices of the 65 MAT items for Gender

Person CLASS	Obs-Exp Average	DIF MEASURE	DIF S.E.	Person CLASS	Obs-Exp Average	DIF MEASURE	DIF S.E.	DIF CONTRAST	JOINT S.E.	Welch t	d.f.	Prob.	Mantel-Haenszel Chi-squ	Prob.	Size CUMLOR	Item Number	Name
F	.01	.48	.07	M	-.01	.56	.07	-.08	.10	-.76	INF	.4470	.3263	.5679	-.07	1	I0001
F	.01	-.24	.08	M	-.01	-.18	.08	-.06	.11	-.54	INF	.5897	.0389	.8436	-.03	2	I0002
F	.02	.98	.08	M	-.02	1.12	.08	-.14	.11	-1.26	INF	.2082	1.6042	.2053	-.15	3	I0003
F	-.01	.59	.07	M	.01	.49	.07	.10	.10	.91	INF	.3605	.4315	.5113	.08	5	I0005
F	-.03	1.58	.08	M	.03	1.33	.08	.25	.12	2.20	INF	.0282	3.9471	.0470	.24	7	I0007
F	-.01	1.01	.08	M	.01	.96	.08	.05	.11	.45	INF	.6521	.3293	.5661	.07	8	I0008
F	.01	.69	.07	M	-.01	.74	.07	-.05	.10	-.51	INF	.6110	.4572	.4989	-.08	9	I0009
F	.01	.36	.07	M	-.01	.47	.07	-.11	.10	-1.05	INF	.2949	.6808	.4093	-.09	10	I0010
F	-.02	-.07	.08	M	.02	-.24	.08	.17	.11	1.56	INF	.1183	1.3784	.2404	.14	11	I0011
F	.02	.57	.07	M	-.02	.70	.07	-.13	.10	-1.26	INF	.2090	1.3953	.2375	-.13	12	I0012
F	.00	.34	.07	M	.00	.36	.07	-.02	.10	-.20	INF	.8441	.0718	.7888	-.03	13	I0013
F	.03	.52	.07	M	-.03	.78	.07	-.25	.10	-2.40	INF	.0163	4.0721	.0436	-.22	14	I0014
F	.03	.89	.08	M	-.03	1.15	.08	-.26	.11	-2.39	INF	.0170	4.9901	.0255	-.25	22	I0022
F	.02	.89	.08	M	-.02	1.08	.08	-.19	.11	-1.75	INF	.0811	2.9200	.0875	-.19	28	I0028
F	-.02	-.82	.09	M	.02	-1.15	.10	.33	.13	2.44	INF	.0147	5.1137	.0237	.31	29	I0029
F	.02	1.36	.08	M	-.02	1.55	.08	-.19	.12	-1.67	INF	.0950	1.6536	.1985	-.16	42	I0042
F	.00	.39	.07	M	.00	.39	.07	.00	.10	.00	INF	1.000	.0002	.9890	.00	44	I0044
F	-.01	-.05	.08	M	.01	-.10	.08	.05	.11	.48	INF	.6330	.2029	.6524	.06	48	I0048
F	.00	-1.92	.13	M	.00	-1.98	.13	.05	.19	.29	INF	.7717	.1209	.7281	.09	51	I0051
F	-.01	-1.62	.12	M	.01	-1.85	.13	.23	.17	1.33	INF	.1832	1.1827	.2768	.21	52	I0052
F	.00	-1.49	.11	M	.00	-1.54	.11	.05	.16	.29	INF	.7718	.0575	.8105	.05	53	I0053
F	-.01	-1.24	.10	M	.01	-1.33	.11	.10	.15	.66	INF	.5066	.1677	.6822	.07	54	I0054
F	-.01	-.85	.09	M	.01	-.96	.09	.12	.13	.91	INF	.3655	1.0049	.3161	.14	55	I0055
F	-.01	-.63	.09	M	.01	-.72	.09	.09	.12	.71	INF	.4796	.0408	.8399	.03	56	I0056
F	-.02	-.42	.08	M	.02	-.60	.08	.18	.12	1.53	INF	.1255	2.7602	.0966	.20	57	I0057
F	.00	-.45	.08	M	.00	-.41	.08	-.04	.12	-.38	INF	.7035	.2466	.6195	-.06	58	I0058
F	.02	-.45	.08	M	-.02	-.30	.08	-.15	.11	-1.33	INF	.1823	1.1071	.2927	-.13	59	I0059
F	.01	-.46	.08	M	-.01	-.37	.08	-.10	.12	-.84	INF	.4027	.4665	.4946	-.09	60	I0060
F	.00	-.46	.08	M	.00	-.46	.08	.00	.12	.00	INF	1.000	.0874	.7675	.04	61	I0061
F	.00	-.50	.08	M	.00	-.50	.08	.00	.12	.00	INF	1.000	.0032	.9546	.00	62	I0062
F	.00	-.51	.08	M	.00	-.55	.08	.04	.12	.36	INF	.7217	.0188	.8911	.02	63	I0063
F	.00	-.73	.09	M	.00	-.78	.09	.06	.12	.46	INF	.6475	.3689	.5436	.09	64	I0064
F	.01	-.74	.09	M	-.01	-.67	.09	-.07	.12	-.59	INF	.5581	.2497	.6173	-.07	65	I0065
F	-.01	-.88	.09	M	.01	-.97	.09	.09	.13	.71	INF	.4774	.6235	.4297	.11	66	I0066
F	.00	-.74	.09	M	.00	-.74	.09	.00	.12	.00	INF	1.000	.0070	.9332	-.02	67	I0067
F	-.01	-.62	.09	M	.01	-.68	.09	.06	.12	.46	INF	.6422	.4599	.4977	.09	68	I0068
F	.00	-.63	.09	M	.00	-.63	.08	.00	.12	.00	INF	1.000	.3446	.5572	-.08	69	I0069
F	-.02	-.49	.08	M	.02	-.71	.09	.22	.12	1.85	INF	.0648	3.4380	.0637	.23	70	I0070
F	-.01	-.69	.09	M	.01	-.84	.09	.15	.13	1.21	INF	.2277	.9820	.3217	.13	71	I0071
F	-.01	-.63	.09	M	.01	-.74	.09	.10	.12	.83	INF	.4069	.8047	.3697	.12	72	I0072
F	-.03	-.25	.08	M	.03	-.51	.08	.27	.11	2.32	INF	.0206	5.0994	.0239	.27	73	I0073
F	.01	-.35	.08	M	-.01	-.30	.08	-.05	.11	-.46	INF	.6427	.5593	.4545	-.09	74	I0074
F	.00	-.46	.08	M	.00	-.46	.08	.00	.12	.00	INF	1.000	.0018	.9659	.00	75	I0075
F	-.04	-.10	.08	M	.04	-.47	.08	.37	.11	3.30	INF	.0010	10.2462	.0014	.37	76	I0076
F	-.01	-.48	.08	M	.01	-.56	.08	.08	.12	.71	INF	.4770	.5942	.4408	.10	77	I0077
F	-.02	-.40	.08	M	.02	-.58	.08	.19	.12	1.59	INF	.1122	1.8264	.1766	.17	78	I0078
F	-.01	-.15	.08	M	.01	-.26	.08	.11	.11	1.01	INF	.3124	.7830	.3762	.11	79	I0079
F	.00	-.20	.08	M	.00	-.18	.08	-.02	.11	-.19	INF	.8483	1.0101	.3149	-.12	80	I0080
F	-.01	-.27	.08	M	.01	-.40	.08	.13	.11	1.18	INF	.2401	.9334	.3340	.12	81	I0081
F	.00	-.48	.08	M	.00	-.46	.08	-.02	.12	-.17	INF	.8624	.3602	.5484	-.08	82	I0082
F	-.02	-.27	.08	M	.02	-.43	.08	.17	.11	1.46	INF	.1447	1.6603	.1976	.16	83	I0083
F	.01	-.66	.09	M	-.01	-.51	.08	-.15	.12	-1.26	INF	.2078	1.0142	.3139	-.13	84	I0084
F	-.01	-.46	.08	M	.01	-.52	.08	.06	.12	.54	INF	.5912	.0876	.7672	.04	85	I0085
F	.01	-.45	.08	M	-.01	-.39	.08	-.06	.12	-.49	INF	.6215	.2548	.6137	-.07	86	I0086
F	.00	-.43	.08	M	.00	-.43	.08	.00	.12	.00	INF	1.000	.1256	.7230	-.05	87	I0087
F	.02	-.46	.08	M	-.02	-.25	.08	-.20	.11	-1.78	INF	.0754	2.3566	.1248	-.18	88	I0088
F	.03	-.82	.09	M	-.03	-.51	.08	-.31	.12	-2.58	INF	.0101	5.6236	.0177	-.30	89	I0089
F	.01	-.59	.09	M	-.01	-.49	.08	-.10	.12	-.83	INF	.4082	.2598	.6102	-.07	90	I0090
F	.00	-.33	.08	M	.00	-.33	.08	.00	.11	.00	INF	1.000	.0020	.9641	.00	91	I0091
F	.02	-.46	.08	M	-.02	-.29	.08	-.17	.11	-1.51	INF	.1320	2.3750	.1233	-.18	92	I0092
F	-.01	-.24	.08	M	.01	-.36	.08	.12	.11	1.06	INF	.2882	.8378	.3600	.11	93	I0093
F	.01	-.38	.08	M	-.01	-.28	.08	-.10	.11	-.92	INF	.3584	.3977	.5283	-.08	94	I0094
F	-.04	-.23	.08	M	.04	-.59	.08	.36	.12	3.13	INF	.0018	8.5699	.0034	.35	95	I0095
F	.01	-.67	.09	M	-.01	-.61	.08	-.06	.12	-.50	INF	.6144	.0334	.8550	-.03	96	I0096
F	-.03	-.42	.08	M	.03	-.71	.09	.29	.12	2.43	INF	.0152	6.0435	.0140	.30	97	I0097
F	.01	-.39	.08	M	.00	-.34	.08	-.05	.11	-.42	INF	.6748	.0039	.9504	-.01	98	I0098
F	.00	-.72	.09	M	.00	-.72	.09	.00	.12	.00	INF	1.000	.0231	.8793	-.03	99	I0099

Research Question 7

What are the difficulty indices of the Mathematics Achievement Test (MAT) items for rural and urban testees using the Rasch model?

The table 16 showed the difficulty indices of the Mathematics Achievement test items for the rural and urban using the Rasch model. The difficulty indices of the rural ranges from -1.99logits to 1.40logit while that of urban ranges from -1.95logits to 1.49logits. The third column from the subgroup indicates the difficulty indices which is named DIF measure.

Table 16   Difficulty indices of the 65 MAT items for Location

Person CLASS	Obs-Exp Average	DIF MEASURE	DIF S.E.	Person CLASS	Obs-Exp Average	DIF MEASURE	DIF S.E.	DIF CONTRAST	JOINT S.E.	Welch t	Welch d.f.	Welch Prob.	Mantel-Haenszel Chi-squ	Mantel-Haenszel Prob.	Size CUMLOR	Item Number	Name
R	.00	.52	.11	U	.00	.52	.06	.00	.13	.00	657	1.000	.0001	.9905	.01	1	I0001
R	.00	-.18	.12	U	.00	-.20	.06	.02	.14	.15	658	.8799	.0039	.9505	.00	2	I0002
R	.01	1.01	.12	U	.00	1.05	.06	-.04	.13	-.29	660	.7735	.1059	.7448	-.05	3	I0003
R	-.01	.60	.11	U	.00	.54	.06	.05	.13	.42	657	.6773	.3739	.5409	.09	5	I0005
R	.03	1.29	.12	U	-.01	1.49	.07	-.20	.14	-1.45	673	.1462	1.0150	.3137	-.15	7	I0007
R	.00	.98	.12	U	.00	.98	.06	.00	.13	.00	658	1.000	.0006	.9803	-.01	8	I0008
R	-.01	.74	.12	U	.00	.72	.06	.02	.13	.19	657	.8482	.0173	.8955	.03	9	I0009
R	-.01	.46	.11	U	.00	.41	.06	.05	.13	.38	658	.7005	.0565	.8121	.04	10	I0010
R	-.01	-.10	.12	U	.00	-.16	.06	.06	.14	.44	660	.6584	.5518	.4576	.12	11	I0011
R	.00	.64	.12	U	.00	.64	.06	.00	.13	.00	657	1.000	.0031	.9553	.00	12	I0012
R	.02	.29	.12	U	.00	.36	.06	-.07	.13	-.55	656	.5839	.1603	.6889	-.06	13	I0013
R	-.02	.73	.12	U	.00	.63	.06	.10	.13	.75	656	.4512	.1713	.6789	.06	14	I0014
R	.01	.96	.12	U	.00	1.02	.06	-.06	.13	-.47	661	.6399	.3472	.5557	-.09	22	I0022
R	.03	.85	.12	U	-.01	1.02	.06	-.17	.13	-1.32	664	.1882	2.6239	.1053	-.22	28	I0028
R	.00	-.95	.15	U	.00	-.98	.07	.03	.16	.19	660	.8487	.0057	.9400	.03	29	I0029
R	.01	1.40	.13	U	.00	1.45	.07	-.05	.14	-.39	664	.6992	.0248	.8748	-.03	42	I0042
R	.00	.39	.12	U	.00	.39	.06	.00	.13	.00	657	1.000	.0856	.7698	.05	44	I0044
R	.01	-.10	.12	U	.00	-.07	.06	-.02	.13	-.18	655	.8604	.0001	.9936	-.01	48	I0048
R	.00	-1.99	.21	U	.00	-1.95	.11	-.04	.24	-.16	648	.8725	.1919	.6614	-.14	51	I0051
R	.00	-1.70	.19	U	.00	-1.73	.10	.03	.21	.14	660	.8856	.0000	.9948	.03	52	I0052
R	.03	-1.82	.20	U	-.01	-1.45	.09	-.37	.22	-1.71	607	.0880	3.2443	.0717	-.43	53	I0053
R	.00	-1.25	.16	U	.00	-1.28	.08	.03	.18	.19	661	.8465	.1235	.7252	.08	54	I0054
R	.00	-.90	.14	U	.00	-.90	.07	.00	.16	.00	655	1.000	.0092	.9234	.03	55	I0055
R	.00	-.68	.14	U	.00	-.68	.07	.00	.15	.00	654	1.000	.0007	.9793	-.01	56	I0056
R	.00	-.51	.13	U	.00	-.51	.07	.00	.15	.00	657	1.000	.0305	.8615	.04	57	I0057
R	-.02	-.34	.13	U	.00	-.45	.07	.12	.14	.83	665	.4083	.7851	.3756	.14	58	I0058
R	.01	-.40	.13	U	.00	-.37	.06	-.03	.14	-.19	654	.8457	.0472	.8280	-.04	59	I0059
R	-.01	-.35	.13	U	.00	-.41	.06	.06	.14	.44	662	.6572	.3427	.5583	.10	60	I0060
R	.00	-.46	.13	U	.00	-.46	.07	.00	.14	.00	654	1.000	.0328	.8562	-.04	61	I0061
R	-.01	-.46	.13	U	.00	-.50	.07	.04	.14	.26	660	.7941	.0882	.7665	.06	62	I0062
R	-.01	-.46	.13	U	.00	-.53	.07	.07	.14	.48	663	.6343	.2487	.6180	.08	63	I0063
R	.01	-.82	.14	U	.00	-.75	.07	-.07	.16	-.44	647	.6573	.1711	.6791	-.08	64	I0064
R	.02	-.84	.14	U	-.01	-.67	.07	-.17	.16	-1.10	639	.2713	.8871	.3463	-.16	65	I0065
R	.04	-1.20	.16	U	-.01	-.86	.07	-.34	.17	-1.94	619	.0534	2.9451	.0861	-.32	66	I0066
R	-.03	-.57	.13	U	.01	-.79	.07	.23	.15	1.51	677	.1310	2.3325	.1267	.24	67	I0067
R	.00	-.67	.13	U	.00	-.65	.07	-.02	.15	-.16	653	.8735	.0033	.9543	.00	68	I0068
R	.01	-.67	.13	U	.00	-.63	.07	-.04	.15	-.26	651	.7975	.0736	.7862	-.05	69	I0069
R	.01	-.65	.13	U	.00	-.60	.07	-.05	.15	-.35	650	.7227	.0803	.7769	-.06	70	I0070
R	-.02	-.67	.13	U	.00	-.79	.07	.11	.15	.75	667	.4531	.6491	.4204	.14	71	I0071
R	.00	-.71	.14	U	.00	-.68	.07	-.02	.15	-.16	653	.8769	.0016	.9684	-.02	72	I0072
R	.02	-.50	.13	U	-.01	-.35	.06	-.15	.14	-1.03	645	.3034	.9978	.3178	-.16	73	I0073
R	.02	-.40	.13	U	.00	-.32	.06	-.08	.14	-.54	650	.5916	.9058	.3412	-.15	74	I0074
R	.00	-.46	.13	U	.00	-.46	.07	.00	.14	.00	657	1.000	.0010	.9743	.01	75	I0075
R	.01	-.34	.13	U	.00	-.28	.06	-.05	.14	-.38	652	.7005	.2665	.6057	-.08	76	I0076
R	-.02	-.42	.13	U	.01	-.55	.07	.13	.14	.92	667	.3568	.5954	.4404	.12	77	I0077
R	.00	-.46	.13	U	.00	-.49	.07	.02	.14	.17	658	.8681	.1175	.7318	.06	78	I0078
R	.01	-.23	.12	U	.00	-.20	.06	-.02	.14	-.18	654	.8607	.0031	.9558	-.02	79	I0079
R	.00	-.18	.12	U	.00	-.18	.06	.00	.14	.00	656	1.000	.0099	.9206	.02	80	I0080
R	.02	-.45	.13	U	-.01	-.30	.06	-.14	.14	-1.00	646	.3162	.7432	.3887	-.14	81	I0081
R	-.02	-.37	.13	U	.00	-.48	.07	.11	.14	.81	665	.4179	.6454	.4218	.13	82	I0082
R	-.03	-.23	.12	U	.01	-.38	.06	.15	.14	1.11	666	.2691	1.9469	.1629	.21	83	I0083
R	-.02	-.46	.13	U	.01	-.61	.07	.15	.15	1.01	668	.3145	.5660	.4519	.12	84	I0084
R	.00	-.49	.13	U	.00	-.49	.07	.00	.14	.00	655	1.000	.0053	.9420	.02	85	I0085
R	-.03	-.27	.12	U	.01	-.46	.07	.19	.14	1.34	670	.1815	1.1076	.2926	.16	86	I0086
R	.01	-.46	.13	U	.00	-.43	.06	-.03	.14	-.23	653	.8164	.0131	.9088	-.03	87	I0087
R	.01	-.38	.13	U	.00	-.35	.06	-.03	.14	-.22	653	.8260	.0619	.8035	-.05	88	I0088
R	.04	-.86	.14	U	-.01	-.61	.07	-.26	.16	-1.64	632	.1022	2.3887	.1222	-.25	89	I0089
R	.02	-.65	.13	U	-.01	-.51	.07	-.14	.15	-.94	644	.3458	.5962	.4400	-.13	90	I0090
R	.03	-.48	.13	U	-.01	-.30	.06	-.18	.14	-1.28	643	.1999	2.3064	.1288	-.23	91	I0091
R	-.02	-.29	.12	U	.00	-.40	.06	.11	.14	.78	664	.4365	.3991	.5275	.10	92	I0092
R	.00	-.30	.12	U	.00	-.30	.06	.00	.14	.00	657	1.000	.0030	.9563	.00	93	I0093
R	-.04	-.12	.12	U	.01	-.39	.06	.26	.14	1.91	673	.0568	2.7397	.0979	.23	94	I0094
R	-.05	-.15	.12	U	.01	-.47	.07	.32	.14	2.32	678	.0206	4.7545	.0292	.32	95	I0095
R	.01	-.69	.14	U	.00	-.64	.07	-.05	.15	-.33	650	.7430	.1491	.6994	-.07	96	I0096
R	-.01	-.48	.13	U	.00	-.58	.07	.10	.15	.67	664	.5006	.1126	.7372	.06	97	I0097
R	-.01	-.30	.12	U	.00	-.37	.06	.06	.14	.43	661	.6665	.1042	.7468	.06	98	I0098
R	.03	-.92	.14	U	-.01	-.67	.07	-.25	.16	-1.56	631	.1190	3.2124	.0731	-.30	99	I0099

Research Question 8

What are the difficulty indices of the Mathematics Achievement Test (MAT) items for public and private testees using the Rasch model?

The table 17 showed the difficulty indices of the Mathematics Achievement test items for the public school and private school using the Rasch model. The difficulty indices of the public school ranges from -1.92logits to 1.42logit while that of private school ranges from -2.02logits to 1.73logits. The third column from the subgroup indicates the difficulty indices which is named DIF measure.

Table 17 Difficulty indices of the 65 MAT items for School type

Person CLASS	Obs-Exp Average	DIF MEASURE	DIF S.E.	Person CLASS	Obs-Exp Average	DIF MEASURE	DIF S.E.	DIF CONTRAST	JOINT S.E.	Welch t	Welch d.f.	Welch Prob.	Mantel-Haenszel Chi-squ	Mantel-Haenszel Prob.	Size CUMLOR	Item Number	Name
A	.00	.52	.06	B	.01	.49	.09	.03	.11	.24	INF	.8084	.0064	.9361	.02	1	I0001
A	.00	-.23	.07	B	-.01	-.16	.10	-.07	.12	-.59	INF	.5565	.1445	.7038	-.05	2	I0002
A	.01	1.01	.06	B	-.02	1.15	.10	-.14	.12	-1.20	INF	.2297	2.1574	.1419	-.19	3	I0003
A	.00	.54	.06	B	.00	.54	.09	.00	.11	.00	INF	1.000	.0088	.9253	.02	5	I0005
A	.02	1.33	.07	B	-.05	1.73	.11	-.40	.13	-3.08	INF	.0021	4.3485	.0370	-.28	7	I0007
A	.01	.96	.06	B	-.01	1.05	.10	-.09	.12	-.77	INF	.4388	.1258	.7228	-.05	8	I0008
A	-.01	.77	.06	B	.03	.59	.09	.18	.11	1.60	INF	.1089	1.9183	.1660	.17	9	I0009
A	-.01	.45	.06	B	.02	.32	.09	.13	.11	1.15	INF	.2510	1.1322	.2873	.13	10	I0010
A	.01	-.19	.07	B	-.02	-.09	.10	-.10	.12	-.84	INF	.4007	.4058	.5241	-.09	11	I0011
A	.00	.64	.06	B	.00	.64	.09	.00	.11	.00	INF	1.000	.0001	.9909	.01	12	I0012
A	.00	.36	.06	B	.00	.36	.09	.00	.11	.00	INF	1.000	.0002	.9885	.01	13	I0013
A	-.01	.70	.06	B	.03	.55	.09	.15	.11	1.33	INF	.1827	1.6097	.2045	.15	14	I0014
A	.00	1.02	.06	B	.00	1.02	.10	.00	.12	.00	INF	1.000	.0090	.9246	-.02	22	I0022
A	.01	.96	.06	B	-.01	1.04	.10	-.07	.12	-.62	INF	.5335	.4546	.5001	-.09	28	I0028
A	.01	-1.02	.08	B	-.01	-.88	.12	-.14	.14	-.98	INF	.3285	.2687	.6042	-.08	29	I0029
A	.01	1.42	.07	B	-.01	1.52	.11	-.10	.13	-.80	INF	.4252	.4978	.4805	-.10	42	I0042
A	.00	.39	.06	B	.00	.39	.09	.00	.11	.00	INF	1.000	.0114	.9149	.02	44	I0044
A	.01	-.11	.07	B	-.02	.02	.10	-.13	.12	-1.13	INF	.2602	.8170	.3661	-.11	48	I0048
A	.00	-1.92	.11	B	.01	-2.02	.18	.10	.21	.48	INF	.6279	.0011	.9734	.02	51	I0051
A	-.01	-1.67	.10	B	.01	-1.90	.17	.23	.20	1.16	INF	.2459	.9108	.3399	.21	52	I0052
A	.01	-1.60	.10	B	-.02	-1.33	.14	-.27	.17	-1.63	INF	.1041	3.5633	.0591	-.33	53	I0053
A	-.01	-1.20	.09	B	.02	-1.51	.14	.31	.17	1.83	INF	.0683	2.5347	.1114	.29	54	I0054
A	.00	-.90	.08	B	.00	-.92	.12	.02	.14	.14	INF	.8879	.0134	.9077	.03	55	I0055
A	-.01	-.60	.07	B	.03	-.88	.12	.28	.14	2.07	INF	.0387	3.3657	.0666	.27	56	I0056
A	.00	-.51	.07	B	.01	-.55	.11	.05	.13	.36	INF	.7166	.1335	.7149	.06	57	I0057
A	-.02	-.34	.07	B	.04	-.64	.11	.29	.13	2.25	INF	.0244	4.0861	.0432	.28	58	I0058
A	.01	-.40	.07	B	-.01	-.30	.10	-.11	.12	-.86	INF	.3902	1.2287	.2677	-.15	59	I0059
A	.00	-.41	.07	B	.01	-.46	.11	.05	.13	.37	INF	.7120	.0662	.7969	.04	60	I0060
A	.01	-.53	.07	B	-.03	-.31	.10	-.22	.12	-1.78	INF	.0751	3.3105	.0688	-.24	61	I0061
A	-.02	-.41	.07	B	.04	-.72	.11	.31	.13	2.34	INF	.0196	4.2405	.0395	.29	62	I0062
A	.00	-.51	.07	B	.01	-.59	.11	.08	.13	.59	INF	.5556	.0303	.8617	.03	63	I0063
A	-.01	-.68	.07	B	.03	-.95	.12	.28	.14	1.97	INF	.0488	2.6978	.1005	.25	64	I0064
A	-.01	-.64	.07	B	.02	-.85	.12	.21	.14	1.53	INF	.1259	1.7708	.1833	.19	65	I0065
A	-.01	-.87	.08	B	.02	-1.07	.12	.21	.15	1.41	INF	.1590	1.4587	.2271	.19	66	I0066
A	-.01	-.66	.07	B	.03	-.95	.12	.29	.14	2.09	INF	.0370	4.1301	.0421	.31	67	I0067
A	-.01	-.60	.07	B	.02	-.76	.11	.16	.13	1.18	INF	.2376	.7775	.3779	.13	68	I0068
A	-.01	-.60	.07	B	.02	-.72	.11	.13	.13	.94	INF	.3471	.1613	.6879	.07	69	I0069
A	-.01	-.57	.07	B	.02	-.68	.11	.12	.13	.90	INF	.3706	.6033	.4373	.12	70	I0070
A	.00	-.73	.07	B	.01	-.83	.12	.09	.14	.67	INF	.5047	.2321	.6299	.08	71	I0071
A	.00	-.68	.07	B	.00	-.68	.11	.00	.13	.00	INF	1.000	.0083	.9274	-.02	72	I0072
A	.00	-.38	.07	B	-.01	-.35	.10	-.03	.12	-.21	INF	.8307	.0055	.9407	-.02	73	I0073
A	.00	-.32	.07	B	.00	-.32	.10	.00	.12	.00	INF	1.000	.0015	.9693	.00	74	I0074
A	.01	-.52	.07	B	-.03	-.33	.10	-.19	.12	-1.49	INF	.1368	2.2086	.1372	-.20	75	I0075
A	.00	-.30	.07	B	-.01	-.24	.10	-.07	.12	-.54	INF	.5869	.6159	.4326	-.10	76	I0076
A	-.01	-.48	.07	B	.02	-.61	.11	.13	.13	1.00	INF	.3181	.6383	.4243	.11	77	I0077
A	.00	-.49	.07	B	.00	-.49	.11	.00	.13	.00	INF	1.000	.0042	.9481	.00	78	I0078
A	.00	-.20	.07	B	.00	-.20	.10	.00	.12	.00	INF	1.000	.0346	.8525	-.03	79	I0079
A	.01	-.24	.07	B	-.03	-.07	.10	-.17	.12	-1.41	INF	.1583	1.6710	.1961	-.17	80	I0080
A	.00	-.33	.07	B	-.01	-.30	.10	-.03	.12	-.28	INF	.7814	.0440	.8338	-.03	81	I0081
A	.00	-.46	.07	B	.00	-.46	.11	.00	.13	.00	INF	1.000	.2131	.6443	-.07	82	I0082
A	-.01	-.28	.07	B	.03	-.51	.11	.22	.13	1.76	INF	.0795	2.5191	.1125	.22	83	I0083
A	-.01	-.50	.07	B	.03	-.77	.11	.27	.13	2.04	INF	.0419	3.4543	.0631	.26	84	I0084
A	-.01	-.45	.07	B	.02	-.59	.11	.14	.13	1.08	INF	.2790	.3728	.5415	.09	85	I0085
A	-.02	-.32	.07	B	.05	-.67	.11	.35	.13	2.70	INF	.0070	6.9385	.0084	.36	86	I0086
A	-.01	-.40	.07	B	.01	-.49	.11	.09	.13	.71	INF	.4783	.2762	.5992	.08	87	I0087
A	.00	-.35	.07	B	.01	-.39	.10	.04	.12	.34	INF	.7324	.0400	.8415	.03	88	I0088
A	.00	-.66	.07	B	.00	-.66	.11	.00	.13	.00	INF	1.000	.0008	.9773	-.01	89	I0089
A	.01	-.59	.07	B	-.02	-.43	.10	-.16	.13	-1.29	INF	.1970	1.2689	.2600	-.15	90	I0090
A	.01	-.39	.07	B	-.03	-.21	.10	-.18	.12	-1.52	INF	.1301	2.5632	.1094	-.21	91	I0091
A	.00	-.37	.07	B	.00	-.35	.10	-.02	.12	-.19	INF	.8511	.0061	.9380	.02	92	I0092
A	.01	-.34	.07	B	-.02	-.22	.10	-.12	.12	-1.01	INF	.3114	.7699	.3802	-.11	93	I0093
A	.00	-.33	.07	B	.01	-.37	.10	.04	.12	.35	INF	.7260	.5193	.4711	.10	94	I0094
A	-.01	-.34	.07	B	.03	-.56	.11	.22	.13	1.76	INF	.0791	1.6662	.1968	.18	95	I0095
A	.01	-.68	.07	B	-.01	-.56	.11	-.11	.13	-.85	INF	.3937	.3346	.5630	-.08	96	I0096
A	-.01	-.52	.07	B	.02	-.66	.11	.14	.13	1.10	INF	.2733	.6865	.4074	.12	97	I0097
A	.00	-.37	.07	B	.00	-.37	.10	.00	.12	.00	INF	1.000	.0222	.8815	.03	98	I0098
A	.02	-.83	.08	B	-.04	-.49	.11	-.34	.13	-2.57	INF	.0102	5.1602	.0231	-.30	99	I0099

Research Question 9

What are the difficulty indices of the Mathematics Achievement Test (MAT) items for HSES and LSES testees using the Rasch model?

The table 18 showed the difficulty indices of the Mathematics Achievement test items for the high social economic status and low social economic status using the Rasch model. The difficulty indices of the high SES ranges from -1.99logits to 1.70logit while that of low SES ranges from -1.92logits to 1.43logits. The third column from the subgroup indicates the difficulty indices which is named DIF measure.

Table 18 Difficulty indices of the 65 MAT items for SES

Person CLASS	Obs-Exp Average	DIF MEASURE	DIF S.E.	Person CLASS	Obs-Exp Average	DIF MEASURE	DIF S.E.	DIF CONTRAST	JOINT S.E.	t	Welch d.f.	Prob.	Mantel-Haenszel Chi-squ	Prob.	Size CUMLOR	Item Number	Name
H	.01	.50	.08	L	.00	.52	.07	-.02	.11	-.21	INF	.8317	.1251	.7235	-.04	1	I0001
H	.00	-.20	.09	L	.00	-.20	.07	.00	.11	.00	INF	1.000	.0242	.8763	.02	2	I0002
H	-.02	1.16	.08	L	.02	.98	.07	.19	.11	1.69	INF	.0920	3.4589	.0629	.22	3	I0003
H	.00	.54	.08	L	.00	.54	.07	.00	.11	.00	INF	1.000	.0331	.8556	-.03	5	I0005
H	-.05	1.70	.09	L	.03	1.28	.07	.41	.12	3.45	INF	.0006	7.0117	.0081	.33	7	I0007
H	-.01	1.01	.08	L	.01	.96	.07	.05	.11	.49	INF	.6253	.0999	.7520	.04	8	I0008
H	.03	.59	.08	L	-.02	.81	.07	-.21	.11	-1.98	INF	.0482	2.6930	.1008	-.18	9	I0009
H	.04	.25	.08	L	-.03	.53	.07	-.28	.11	-2.65	INF	.0081	6.2406	.0125	-.27	10	I0010
H	.01	-.19	.09	L	-.01	-.13	.07	-.06	.11	-.55	INF	.5806	.5433	.4611	-.09	11	I0011
H	.01	.60	.08	L	-.01	.67	.07	-.07	.11	-.62	INF	.5343	.4551	.4999	-.08	12	I0012
H	.00	.36	.08	L	.00	.36	.07	.00	.11	.00	INF	1.000	.2509	.6165	-.06	13	I0013
H	.02	.56	.08	L	-.02	.72	.07	-.16	.11	-1.54	INF	.1235	1.1832	.2767	-.12	14	I0014
H	.00	1.02	.08	L	.00	1.02	.07	.00	.11	.00	INF	1.000	.0026	.9595	.00	22	I0022
H	-.01	1.04	.08	L	.01	.95	.07	.10	.11	.88	INF	.3805	.8669	.3518	.11	28	I0028
H	.00	-.98	.10	L	.00	-.98	.09	.00	.13	.00	INF	1.000	.0136	.9070	.03	29	I0029
H	-.01	1.49	.09	L	.00	1.43	.08	.06	.12	.48	INF	.6284	.2546	.6139	.07	42	I0042
H	.00	.39	.08	L	.00	.39	.07	.00	.11	.00	INF	1.000	.0009	.9763	.01	44	I0044
H	-.01	-.01	.08	L	.01	-.12	.07	.10	.11	.92	INF	.3593	.3551	.5513	.07	48	I0048
H	.00	-1.99	.15	L	.00	-1.92	.12	-.07	.19	-.35	INF	.7267	.0022	.9626	-.01	51	I0051
H	.00	-1.78	.14	L	.00	-1.70	.11	-.08	.18	-.47	INF	.6401	.0610	.8049	-.06	52	I0052
H	-.03	-1.27	.11	L	.02	-1.71	.11	.44	.16	2.74	INF	.0062	8.7223	.0031	.49	53	I0053
H	.00	-1.28	.11	L	.00	-1.28	.10	.00	.15	.00	INF	1.000	.0960	.7567	-.06	54	I0054
H	-.03	-.73	.10	L	.02	-1.04	.09	.32	.13	2.42	INF	.0156	6.4461	.0111	.35	55	I0055
H	.01	-.71	.10	L	.00	-.66	.08	-.05	.12	-.42	INF	.6742	.2063	.6497	-.07	56	I0056
H	-.02	-.41	.09	L	.01	-.58	.08	.16	.12	1.38	INF	.1689	1.1423	.2852	.14	57	I0057
H	.01	-.46	.09	L	.00	-.41	.08	-.06	.12	-.47	INF	.6372	.1612	.6880	-.06	58	I0058
H	-.04	-.16	.09	L	.03	-.53	.08	.37	.11	3.22	INF	.0013	9.9462	.0016	.38	59	I0059
H	.00	-.41	.09	L	.00	-.41	.08	.00	.12	.00	INF	1.000	.0004	.9832	.01	60	I0060
H	-.04	-.24	.09	L	.03	-.63	.08	.39	.12	3.31	INF	.0010	11.0582	.0009	.40	61	I0061
H	.02	-.60	.09	L	-.01	-.43	.08	-.17	.12	-1.41	INF	.1595	1.4575	.2273	-.16	62	I0062
H	.00	-.53	.09	L	.00	-.53	.08	.00	.12	.00	INF	1.000	.1213	.7277	.05	63	I0063
H	.01	-.82	.10	L	-.01	-.71	.08	-.11	.13	-.89	INF	.3732	.3853	.5348	-.09	64	I0064
H	-.01	-.63	.09	L	.01	-.76	.08	.13	.12	1.06	INF	.2872	1.7268	.1888	.18	65	I0065
H	.00	-.95	.10	L	.00	-.92	.09	-.03	.13	-.20	INF	.8406	.0671	.7956	-.04	66	I0066
H	.01	-.78	.10	L	.00	-.71	.08	-.07	.13	-.54	INF	.5882	.2160	.6421	-.07	67	I0067
H	.01	-.73	.10	L	-.01	-.59	.08	-.13	.12	-1.07	INF	.2831	.8407	.3592	-.12	68	I0068
H	-.01	-.59	.09	L	.01	-.67	.08	.08	.12	.69	INF	.4924	.7395	.3898	.12	69	I0069
H	.02	-.70	.10	L	-.01	-.53	.08	-.17	.12	-1.36	INF	.1753	1.4862	.2228	-.16	70	I0070
H	.01	-.84	.10	L	-.01	-.71	.08	-.13	.13	-1.04	INF	.2984	.6989	.4031	-.12	71	I0071
H	.00	-.68	.09	L	.00	-.68	.08	.00	.12	.00	INF	1.000	.0804	.7768	.04	72	I0072
H	.01	-.40	.09	L	.00	-.38	.07	-.03	.12	-.23	INF	.8196	.0109	.9168	-.02	73	I0073
H	-.02	-.24	.09	L	.01	-.38	.07	.14	.11	1.21	INF	.2247	1.9619	.1613	.17	74	I0074
H	.02	-.55	.09	L	-.01	-.39	.08	-.16	.12	-1.33	INF	.1848	1.3713	.2416	-.15	75	I0075
H	-.01	-.24	.09	L	.01	-.31	.07	.07	.11	.59	INF	.5565	.3748	.5404	.08	76	I0076
H	.02	-.60	.09	L	-.01	-.46	.08	-.14	.12	-1.16	INF	.2445	1.0607	.3031	-.13	77	I0077
H	.03	-.64	.09	L	-.02	-.39	.07	-.25	.12	-2.07	INF	.0383	4.9261	.0265	-.28	78	I0078
H	-.02	-.09	.08	L	.02	-.29	.07	.20	.11	1.76	INF	.0779	3.1927	.0740	.21	79	I0079
H	-.02	-.10	.08	L	.01	-.25	.07	.15	.11	1.32	INF	.1872	.8172	.3660	.11	80	I0080
H	-.03	-.17	.09	L	.03	-.46	.08	.29	.11	2.51	INF	.0123	4.9643	.0259	.27	81	I0081
H	.00	-.46	.09	L	.00	-.46	.08	.00	.12	.00	INF	1.000	.0137	.9069	-.02	82	I0082
H	.01	-.40	.09	L	-.01	-.32	.07	-.08	.12	-.70	INF	.4857	.7649	.3818	-.11	83	I0083
H	.03	-.73	.10	L	-.02	-.48	.08	-.25	.12	-2.01	INF	.0442	3.2091	.0732	-.23	84	I0084
H	.02	-.59	.09	L	-.01	-.42	.08	-.16	.12	-1.36	INF	.1727	1.8054	.1791	-.17	85	I0085
H	.03	-.56	.09	L	-.02	-.33	.07	-.23	.12	-1.97	INF	.0492	3.3275	.0681	-.23	86	I0086
H	.02	-.56	.09	L	-.02	-.34	.07	-.22	.12	-1.83	INF	.0680	2.9480	.0860	-.22	87	I0087
H	.00	-.35	.09	L	.00	-.35	.07	.00	.12	.00	INF	1.000	.0077	.9300	-.02	88	I0088
H	.01	-.70	.10	L	-.01	-.62	.08	-.07	.12	-.60	INF	.5464	.2215	.6379	-.07	89	I0089
H	.00	-.52	.09	L	.00	-.54	.08	.02	.12	.19	INF	.8524	.0703	.7910	.04	90	I0090
H	-.03	-.17	.09	L	.03	-.46	.08	.29	.11	2.51	INF	.0123	6.9685	.0083	.31	91	I0091
H	.01	-.44	.09	L	-.01	-.33	.07	-.10	.12	-.89	INF	.3722	1.1933	.2747	-.13	92	I0092
H	-.01	-.26	.09	L	.01	-.33	.07	.07	.11	.65	INF	.5171	.6254	.4291	.10	93	I0093
H	.02	-.42	.09	L	-.01	-.27	.07	-.15	.12	-1.33	INF	.1850	1.5644	.2110	-.15	94	I0094
H	.02	-.50	.09	L	-.01	-.34	.07	-.16	.12	-1.39	INF	.1643	1.7490	.1860	-.16	95	I0095
H	.02	-.74	.10	L	-.01	-.58	.08	-.16	.12	-1.30	INF	.1948	1.1276	.2883	-.14	96	I0096
H	.02	-.66	.09	L	-.01	-.49	.08	-.18	.12	-1.46	INF	.1434	1.3333	.2482	-.15	97	I0097
H	.00	-.39	.09	L	.00	-.37	.07	-.02	.12	-.21	INF	.8375	.0000	.9974	-.01	98	I0098
H	-.01	-.68	.09	L	.01	-.75	.08	.07	.13	.58	INF	.5625	.3476	.5555	.08	99	I0099

Hypotheses

The Hypotheses tested below were used to establish if item parameter index (b-parameter value) significantly differ among the subpopulation of the test takers. This was done using t-test at .05 level of significance.

Hypothesis One

There is no significant difference between the mean difficulty index (b-parameter value) of male and female testees in Mathematics Achievement Test (MAT)

The independent-sample t-test indicated that the b-parameter values were slightly higher for male (N=757, M= .6763, SD= .07135) than for female (N=742, M=.6742, SD= .07697), t(1497)= .561, p> .05.

Table 19 Summary table of independent samples t-test showing the invariance of the mean difficulty index of gender in the MAT item

	Gender	N	Mean	SD	df	t_{cal}	t_{crit}	Decision	Remarks
TEST	**Female**	**742**	**.6742**	**.07697**	**1497**	**.547**	**1.96**	**accept**	**Not significant**
	Male	**757**	**.6763**	**.07135**					

Table 19 showed that the critical t-value of 1.96 was greater than the calculated t-value of 0.547 at .05 level of significance. The null hypothesis was therefore accepted. This implies that there was no significant difference between the mean difficulty index (b-parameter) of male and female testees in MAT

Hypothesis Two

There is no significant difference between the mean difficulty index (b-parameter value) of Low Social Economic Status and High Social Economic Status testees' in Mathematics Achievement Test (MAT).

In the group statistics, the independent-sample t-test indicated that b-parameter values were slightly insignificantly higher for HSES (N=629, M= .6777, SD= .07637) than for LSES (N=870, M=.6735, SD= .07253), t(1497)= 1.071, p> .05,

Table 20 Summary table of independent samples t-test showing the invariance of the mean difficulty index of SES in the MAT item

	SES	N	Mean	SD	df	t_{cal}	t_{crit}	Decision	Remarks
TEST	**Low**	**870**	**.6735**	**.07253**	**1497**	**.340**	**1.96**	**accept**	**Not significant**
	High	**629**	**.6777**	**.07637**					

Table 20 showed that the critical t-value of 1.96 was greater than the calculated t-value of 0.340 at .05 level of significance. The null hypothesis was therefore accepted. This means that there was no significant difference between the mean difficulty index (b-parameter) of high social economic status and low social economic status testees in MAT**.**

Hypothesis Three

There is no significant difference between the mean difficulty index (b-parameter value) of the public school testees and the private school testees in MAT.

In the group t-test statistics, the independent-sample t-test of the group statistics indicated that the b-parameter values were slightly higher for private schools (N=459, M= .6848, SD= .07354) than for public schools (N=1040, M=.6871, SD= .07409), t(1497)= 3.328, p> .05,. This means that the M and SD of the private schools are slightly higher but it was insignificant.

Table 21 Summary table of independent samples t-test showing the invariance of the mean difficulty index of School type in the MAT item

	School	N	Mean	SD	df	t_{cal}	t_{crit}	Decision	Remarks
TEST	**Public**	**1040**	**.6871**	**.07409**	**1497**	**.557**	**1.96**	**accept**	**Not significant**
	Private	**459**	**.6848**	**.07354**					

Table 21 showed that the critical t-value of 1.96 was greater than the calculated t-value of 0.557 at .05 level of significance. The null hypothesis was therefore accepted. This means that there was no significant difference between the mean difficulty index (b-parameter) of public and private school testees in MAT

Hypothesis Four

There is no significant difference between the mean difficulty index (b-parameter value) of the urban testees and the rural testees in Mathematics Achievement Test (MAT).

In the group t-test statistics below, the independent-sample t-test of the group statistics indicated that the b-parameter values were higher insignificantly for Rural (N=309, M= .6782, SD= .07066) than for Urban (N=1190, M=.6745, SD= .07506), t(1497)= .781, p> .05.

Table 22 Summary table of independent samples t-test showing the invariance of the mean difficulty index of Location in the MAT item

	Location	N	Mean	SD	df	t_{cal}	t_{crit}	Decision	Remarks
TEST	**Rural**	**309**	**.6782**	**.07066**	**1497**	**.809**	**1.96**	**accept**	**Not significant**
	Urban	**1190**	**.6745**	**.07506**					

Table 22 showed that the critical t-value of 1.96 was greater than the calculated t-value of 0.809 at .05 level of significance. The null hypothesis was therefore accepted. This implies that there was no significant difference between the mean difficulty index (b-parameter) of urban and rural testees in MAT

Discussion of Findings

The discussion of the findings of this study was done under the following subheadings:

i. Validity of the MAT items using the Rasch model
ii. Reliability of the MAT items Using the Rasch model
iii. Difficulty indices of the MAT items using the Rasch model
iv. The spread of the MAT items using person-item-map of the Rasch model
v. Stability of difficulty index (b-parameter) of MAT items

Validity of the MAT items using the Rasch model

The means of the infit and outfit MNSQ was 1.00 for both and the means of the infit and outfit ZSTD of .1 and 0.0 respectively, were close to the value expected by the model (1.00 for MNSQ and .0 for ZSTD). This suggests that the amount of distortion of the measurement was minimal. Although the standard deviation of both the infit and outfit MNSQ (.03 and .04, respectively) were slightly insignificant compared with the expected value, these little or no difference discrepancies were too small and showed that the data demonstrated fitness from the Rasch Model expectation. Also, to assess the fit of the Rasch model to the data, we equally examined infit mean-square (information-weighted mean-square statistics which is more sensitive to the unexpected behaviour of items closer to persons' measures) and outfit (unweighted mean-square sensitive to outliers). Mean-square (MNSQ) is computed as the chi-square value divided by the degree of freedom. MNSQ fit indices show useful, as opposed to perfect, fit of the data to the model. An infit MNSQ of, say, 1.2 means 1 unit of modeled information is observed and 0.2 units of unmodeled noise sneaks in (Linacre, 2004). The t-test significance (ZSTD) is used to investigate the perfect fit of the data to the model (acceptable range:|2|). Therefore, individual items demonstrated infit MNSQ values from 0.97 to 1.03, while outfit MNSQ were between 0.94 to 1.08, which were within the acceptable range of 0.7–1.1 for a sample greater than 1000 while the items demonstrated infit ZSTD values from -1.7 to 2.0, while outfit ZSTD were between -1.6 to 2.0, which were within the acceptable range of -2.0 to +2.0. This was in congruent with Green and Frantom, 2002; Bond and Fox, 2001 who suggested 0.7-1.1 and -2+2 for both infit and outfit of MNSQ and ZSTD respectively. This established the structural aspect of construct validity.

Item difficulty measures spread approximately .00*logits* (from -1.95logit to +1.45logit), while testees ability measures spanned approximately .51*logits* (from -.36logit to +1.47logit). The mean for item difficulty was 915.1(standard error = .07logit), while the mean for testees ability was 61.0 (standard error = .01logit). The small difference in mean measures of the testees and the items indicated that the MAT targeted the testees well.

The result of the Principle Component Analysis of Rasch (PCAR) of factor analysis was found to be of statistical significance and of practical importance since the standardised residual coefficient was not larger than 2.0 which showed unidimensionality (Green & Frantom, 2002). This was in congruent with Wright, 1997, Messick, 1998 and Tennants', 2000 findings whose PCAR standardized residual coefficients were not also greater than 2.0. This was used to establish substantive and content aspect of the construct validity of the six facet Messick's principle.

Reliability of the MAT items Using the Rasch model

Reliability of item difficulty measures was very high (.99), suggesting that the ordering of item difficulty was highly replicable with another comparable sample of testees. This was in support of the findings of Ahmad and Nordin 2012, Golino et al, 2012 and Andrich, 2012 with reliability that ranges between 0.97-0.99. This established both structural and content validity.

Internal consistency of the student ability measure was also high at .78, when the items with negative point measure correlation were removed, indicating that it was likely that the ordering of testees ability could be replicated because most of the variance in the measured scores was attributed to true variance of the Mathematics Achievement Test (MAT) construct. The standard error of measurement (SEM) associated with the b-parameter of each of the MAT item is used to estimate its reliability. The SE of the item of MAT ranged from 0.05 for item 27 to 0.09 for item 51. Low SE (0.50 and below) indicate high reliability whereas high SE (0.5 and above) indicate low reliability (Nworgu & Agah, 2012 and Obinne et al, 2011). Therefore, all items had SE within the range of 0.05 and 0.09 and the mean SE of the MAT was 0.06 with SD of 0.01. This accounted for the high item reliability of 0.99. This supported Ahmad and Nordin 2012, Golino et al, 2012 and Andrich, 2012 whose item reliability ranges between 0.97-0.99. Thus, it can be concluded that the TEST was adequate in measuring the Mathematics Achievement construct. This established the content validity

As depicted in Table 6 & 7, the point measure correlation (PTMEA CORR.) ranged from -.01 to .38, with eight (8) items containing negative values. These negative items when removed improved the reliability of person from 0.53 to 0.78. This correlation indicated that almost all items were working together in the same way in defining the Mathematics Achievement Test items. This was in consonant with the findings of Bond and Fox 2007 that when negative point measure correlation values were removed, they helped the items to work together in the same way thereby enhancing the reliability of the test. This established content aspect of the six facet Messick’s construct validity.

Difficulty indices of the MAT items using the Rasch model

The difficulty level of the items ranges from -1.95logit to 1.45logit. Results of the PCA of the residuals of Rasch (PCAR) indicated that the largest factor extracted from the residuals was 1.4 units, which has the strength of about 2 items and is well below the 5 items needed for consideration as a second factor (Linacre, 2007). In addition, no gaps of .5*logits* or more (Linacre, 2004) between item distributions on the Achievement scale showed that the items were inadequate in accessing important features of the Mathematics Achievement Test construct.

The spread of the MAT items using person-item-map of the Rasch model

The results of item difficulty and student ability measures provided initial information as to the adequacy of the MAT. In Figure 4.2, both the items and the testees were located along the achievement scale. The items on the top were more difficult, and the testees at the top displayed higher ability. As we went down the line, the items became easier, and the testees displayed less ability. This supported what Bagheai and Amrahi, 2011 said that a close examination of the person-item-map revealed the amount of information on the basis of which decisions for action could be taken. This provided helpful evidence to decide about the external and consequential aspect of construct validity of a test and the testees.

Another important finding was that all items SD - .72logit, model error - .01, S.E of item mean = .07, separation = 12.48 on person-map-item graph spread over the entire range of the scale which indicate that all parts of the construct are well covered by the test - spread of item and person (standard deviations SD of 0.72logit and 0.29logit respectively). Person had a smaller spread (SD = 0.29logit, separation = 1.84) compared with item SD = .72logit, separation = 12.48. This supported Green and Frantom, 2002; Bond and Foxs', 2001, findings. They were of the opinion that for a test item to spread across the continuum indicated the coverage in content of such test. This established the external, content and substantive validity.

One important observation from the inspection of the individual item in the person-item-map was that the testees involved in this study were able to answer items related to understanding information that was explicitly stated (e.g., the terminology found in the diagram [item 45], or from the list of answers option [item 46]). This understanding was expected since the items only required a lower level of understanding in which testees only need to locate the information that is explicitly stated. Similarly, some testees were able to answer items that require use of some straightforward procedures (e.g., items 54 and 55). However, testees had difficulty with items that required them to use their prior knowledge to solve new problems, particularly, making connections between topics (e.g., items 3, 7, 28 and 42). That trend is troublesome because making connections between various forms of mathematical knowledge, especially between concept and procedure and between mathematics and real-life experience, is important to effective Mathematics learning and teaching (Ahmad & Nordin, 2012).

Stability of difficulty index (b-parameter) of MAT items

The result of study generally showed no significant different between the difficulty index (b-parameter) of the gender, socio-economic status, school type and the location of testees in the MAT. The calculated t-value for gender was 0.547, for SES was 0.34, for school type was 0.557 while for location was 0.809. These results showed that the critical t-value of 1.96 was greater that these calculated t-values at 0.05 level of significance. Therefore, their stated hypotheses were all accepted. This was in congruent with the findings of Andrich, 2012 whose result showed no significant difference between the mean score of the variables used in his work. However, the findings of Golino et al, 2012 did not support this finding. This could be as a result the effect size of his sample or some other factors could have accounted for this. Therefore, these findings showed the stability of the item parameter (b-parameter) of the MAT and it was a major shift from the CTT. Hence, the findings were in support of the principle of invariance in IRT of the Rasch model. The independent t-test at 95% confidence interval showed there were no significant differences between the mean difficulties indexes (b-parameter) of the variables used in this study. This established the generalizability and consequential aspect of the six facets of Messick construct validity.

Research Findings

From the above discussion, this implied that:

(1) The MAT items have the six facet Messick's construct validity. They are the content, substantive, structural, external, consequential and generalizability.

(2) The infit and outfit MNSQ and ZSTD values ranges from 0.7 to 1.1 and -2 to +2 respectively.

(3) The reliability of the MAT items was 0.99 which was high enough for literature.

(4) The difficulty indices ranges from -1.95logit to 1.45logit.

(5) All the 65 MAT items demonstrated good distracter qualities because they were functioning in the intended ways.

(6) The person-item-map showed that the test is quite broad because the items spread across the continuum from the top to the bottom of the graph.

(7) The items showed invariance across the subgroups used in the study e.g SES, gender, location and school type. Also, the independent t-test with 95% confidence interval shows that there was no significant difference between the mean difficulty index (b-parameter value) of the gender, ses, school location

CHAPTER FIVE

SUMMARY, CONCLUSION AND RECOMMENDATIONS

This last chapter briefly gives attention to the summary of the study, conclusion, recommendations, and contribution to knowledge, limitations and as well as suggestions for further studies.

Summary of the study

This research aimed at the development and validation of Mathematics Achievement Test using the Rasch model. Five research questions and four research hypotheses were formulated to guide the study. Literature relevant to the study were reviewed especially procedure for test development and validation using the Rasch model of item response theory. Messick's six facet construct validity were similarly reviewed and checked. Moreover, empirical studies on test development, validation and related factors affecting testees' Achievement in Mathematics were critically looked upon.

In the method of study, the researcher used the instrumentation research design. A large population size of 100470 SSS 3 testees from 25 and 33 LGAs of DELTA and OYO states respectively were considered. A large sample of 1499 testees were used from the selected 20 schools each making a total of 40 schools picked from the two states through a multi-stage sampling technique. Two instruments were used: Mathematics Achievement Test (MAT) and Questionnaire for Social Economic Status (QSES). The researcher generated 165 questions/items using the table of specification which cut across all the topics required by WASSCE and NECO syllabi. The items were prone to 150 through vetting by specialists. Factor analysis using the PCA for item analysis was carried out on the 150 items to select, review, re-write and edit the final test to be administered. The test contained 100 Mathematics Test Items. The test blue print and the specialists in Mathematics and measurement and evaluation were used to establish the content validity and the face validity of the generated items. The KR-20 as well was used to establish the reliability of the test items. The 100 items were used to gather data from the field. The data gathered were then subjected to Rasch analysis.

The summaries of findings adopted for determining the dimensionality of the MAT items are:

i. Empirical variance explained (71.8%) is equivalent to the modeled variance (71.8%)

ii. Person measure variance (11.1%) < item difficulty variance (17.1%), person measure SD (.29logit) < item difficulty SD (.72logit).

iii. Unexplained variance 1st contrast, the eigenvalue is not greater than 2.0 (indication of how many items there might be) = 1.4

iv. Variance in items is at least 4 times more than the variance in the 1st contrast 17.1%/2.8% =6.1.

v. Variance of the measures (71.8%) is greater than 60%.

vi. Unexplained variance in the 1st contrast eigenvalue 1.4 < 3.0, < 1.5 is excellent and 2.8% < 5% is more excellent.

vii. Variance unexplained by the 1st contrast (11.8%) < (17.1%) variance of item difficulties

viii. 1st constrast variance < simulated Rasch data 1st constrast variance 1% - 5% difference between the simulated and actual data is okay, 10% is a red flag (i.e 17.1%-11.8= 5.1).

ix. Total raw variance explained by measures and the total raw variance of unexplained by measure was very high, which is 71.8 to 28.8.

x. Dimensionality plot showed item spread from the top to the bottom- look at the person-item-map from top to bottom.

All the above ten qualities indicate that the MAT has unidimensionality trait.

The following evidences were used in establishing the validity of MAT in this study; in other words, it is known that the observed measures met the minimal requirement of the construct validity required by the Rasch model and Messick's six faceth validity construct:

i. KR-20 analysis of .78 for person when the items were reduced.

ii. Infit and outfit of MNSQ and Z-standardization (ZSTD) scores of the 100 items

iii. Rasch reliability indices of .99 for the items.

iv. Item misfit order.

v. Invariance analysis using t-test analysis.

vi. Person-map-item

vii. Point measure correlation

viii. PCAR and Standardized residual variance in Rasch

ix Rasch measures of item distracter procedure

The sixty-five items of MAT were able to scale through the nine stages stated above in order to establish the construct validity and also met the Messick's six faceth construct validity stages.

Conclusion of the findings

On the basis of these research findings, with the use of Rasch Model framework, it is evident that the MAT was adequate in measuring the Achievement construct regarding the following findings: (a) the individual item provided enough contribution to the overall measurement of Mathematics Achievement Test items and equally established unidimensionality trait and local independence of the items. (b) the MAT items fit the requirement of the Rasch measurement model and demonstrated the six facets of Messick's construct validity such as content, substantive, structural, generalizability, external, and consequential evidences of construct validity, and as such, the calibrated items were useful in measuring students' achievement in Mathematics, (c) the Mathematics Achievement Test measured using the test did confound with other related constructs, and (d) threats to construct validity, such as construct-irrelevance variances and construct under-representation, were kept to a minimum. In addition, the study found out that all indicators in the development and validation of the test items hypothesized in some past studies had strong agreement with the model proposed. Also, the principles of invariance was upheld in the study – there was no significant difference between the mean difficulty index (b-parameter values) of the gender, SES, school type and school location.

Implications of the Study

The study also has several practical implications for teachers or classroom educators, measurement and evaluation experts and testees. First, assessments should be planned by teachers to promote greater mathematics ability rather than ranking testees. Testees will find it useful for self assessment to know their level of preparedness for WASSCE and NECO or any external examinations. The result of the test can be useful for selection and promotion of testees. Test should be in the form of criterion-referenced, where information about what testees know and are able to do are available so that they can learn from the achievement. Teachers and measurement experts can use the information to make effective instruction decisions since because feedback generally leads to clearer and more effective instruction while measurement experts can use the information for counseling both the teacher and the students as the case may be.

Recommendations

As documented throughout the findings, the MAT was developed, validated and scrutinized for empirical evidence of adequacy in measuring the mathematics achievement test. Based on this, it is practically recommended that:

i. Test writers should ensure that all the observed measures of unidimensionality meet the minimum requirement of the construct validity required by the Rasch model and Messick's six faceth validity construct.

ii. The stability of difficulty index (b- parameter) be established across any subpopulation used in a study.

iii. Items that came out as bad items should be removed from the test or re-constructed. This will enhance the qualities of items production and aid effective evaluation.

iv. Threats to construct validity such as construct-iirelevance and construct under-representation should be kept to a minimal.

v. Workshops and seminars on test development and validation should be organized for classroom teachers and test developers who are not familiar with the Rasch model.

Contribution to Knowledge

Therefore, the study has contributed to knowledge in the following ways:

i. The study provided information that helped to establish sample independent item parameter inherent in the Rasch model of Item Response Theory of measurement.

ii. The study has shown a developed MAT item whose items have been calibrated based on Rasch model.

iii. The study will guide item writers in writing test items that measures unit trait in Mathematics.

iv. The study provided an item bank which could allow test users to pick items that would select his ability of interest.

v. The test provided b-parameter (item difficulty) values which confirm a position of Item Response Theory as such values are invariance across some subpopulations of test takers.

vi. The study was the only one that expressed its output result in both wit and logit units in Nigeria. Other analyses have been in wit or logit and not both. Therefore, this study will be a model for anyone who intends to use either wit or logit unit.

Limitations to the study

The following limitations were observed in the course of undertaking the study:

i. The predictive validity of the MAT has not been verified in this study. Further research needs to be done to investigate the relationship between the student's achievement and their successes in their prospective courses.

ii. In this study, the person reliability was somehow moderate, which might be the result of the homogenous grouping. A replication of this study with a sample of wider range ability and heterogeneity could still be very informative.

iii. The study did not take a cutoff score into account. A study or research can be conducted to determine a cutoff score for the MAT, identifying more competent testees from less competent ones.

iv. There was no research carried out with the use of the Rasch model using both logit and wit in Nigeria that could serve as a model for the researcher. It is first of its kind as known to the researcher as at the time it was conducted.

v. The software: Winsteps, Bilog-MG3, Parscale, Multilog and X-calibre are not readily available. The easy of use, the knowledge of specific software and familiarity with output interpretation was almost a limitation.

Suggestions for Further Study

This research was based on the development and validation of Mathematics Achievement Test (MAT) using the Rasch model in Oyo and Delta states. Hence, the researcher hereby suggests that other researchers should look into the following areas when validating:

i. A study should be carried out to cover the six geo-political zones in Nigeria with a larger population sample for better criterion reference since the software for analysis can handle 30million testees and 10,000 items at once.

ii. The items logits should be observed prior to and after removal of the responses in subsequent investigation or research study.

iii. The predictive validity of the instrument and the WASSCE and NECO examinations should be investigated.

iv. The relationship between factor analysis of PCA used in selecting the hundred items of SPSS for final administration and PCAR of the Rasch model should be investigated.

v. The Differential Item Function (DIF) of each item of the MAT should be investigated for subsequent research study.

vi. Other subject areas other than Mathematics should be developed and validated using the same procedure and model.

REFERENCES

Abiam, P. O. & Odok, J. K. (2006). Factors in Students' Achievement in different branches of Senior Secondary School Mathematics. *Journal of Education and Techonology (1) 1, 161-168*

Adedoyin,O.O., Nenty, H. J. & Chilisa, B. (2008). Investigating the invariance of item difficulty parameter estimates based on CTT and IRT. *International Journal Educational science Research and Review vol. 3(2), 83-93*

Adedoyin,O.O. (2010). Investigating the invariance of person parameter estimates Based on CTT and IRT. *International Journal Educational science Research and Review vol. 2(2), 117-113*

Adegoke, A. B., (2002). Teachers influence as a determinant of dependent- prone-students learning outcomes in senior secondary school Geometry in Ibadan-South- East Nigeria. *An unpublished Ph.D Thesis University of Ibadan.*

Adekola, B.O. (2012). Home and School factors as Determinants of students' achievement in Senior Secondary School English Comprehension in four South Western States. *Research Journal in organizational Psychology & educational Studies 1(5), 280-283.*

Adeyemi, T.O. (2008). Predicting students' performance in Senior Secondary Certificate Examinations from performance in Junior Secondary Certificate Examinations in Ondo State, Nigeria. Idosi Publications. *Humanity and Social Sciences Journal 3(1), 26-36.*

Agwagah, V. V. (1985). The development and preliminary validation of an achievement test in Mathematics methods. *M.Sc. Dissertation. University of Nigeria, Nsukka*.

Ahmad, Z. K. & Nordin, M. S. (2011). The development and construct validation of the Mathematics proficiency test for 14- years old student. *Asia Pacific Journal of educators and education 26(1), 33-50*

Ahmad, Z. K. & Nordin, R. (2012). Advance in educational measurement: A Rasch model analysis of Mathematics proficiency test. *International journal of social science and humanity 2(3), 248- 251.*

Ahmad, Z. K. & Nordin, R. & Hasni, S. (2012). Modelling students' Mathematics ability and items' difficulty parameters: application of the Rasch measurement model. *International journal of science and engineering research 3(8): 2229- 5518.*

Akanle, O. B.(2007). Social Economic Factor influencing Students Academic performance in Nigeria. *A journal of social science 14(3,: 201-216.*

Akinbile, L. A. (2007). Standardization of social Economic Status (SES) scale for farm families in south west Nigeria. *A journal of social science 14(3,: 221-227*

Akindele, B. P. (2004). The development of an item bank for selection tests into Nigerian Universities: *An exploratory study. Ph.D Thesis. institute of Education, University of Ibadan.*

Alli, A. (1996). Fundamentals of research in education AWKA: Meks Publishers (Nig.).

Aliyu, R. T. (2002). Role of Parents in the Education of the Disabled Children. *An unpublished PGDE dissertation,* University of Ado-Ekiti (UNAD).

Aliyu, R. T. (2008). Relationship between parental educational level and senior secondary school student's achievement in Mathematics. *An unpublished M. Ed Dissertation, Delta state university*

Aliyu, R. T., & Ocheli, E. U. (2013). Development and validation of college Mathematics Test using Item Response Theory (IRT) models. A Delta Journal of Educational Research and Development 12(1), 130-140.

American Psychological Association (APA) (2012). Concise Rules of APA Style. *Washington,DC:APA publications.*

Anderson, L. A., & Zimmerman, M. A. (1993). Patient and physician perceptions of their relationship and patient satisfaction: A study of chronic disease management. *Patient Education and Counseling, 20, 27-36*.

Andrich, D. (2012). The t-test for the comparison of two estimates of a person. *Journal of measurement USA. University of Western Australia 3(12):207-219.*

Andrich, D. (2004). Controversy and the Rasch model: A characteristic of incompatible paradigms? In Smith E. V. & Smith, R. M (Eds.), *Introduction to Rasch measurement* (pp. 143-166). Maple Grove, MN: JAM Press.

Andrich, D. (2010). Sufficiency and Conditional Estimation of person parameter in the polytomous Rasch model. *Psychometrika parameter. 75(2), 292-308.*

Andrich, D. (1989), Distinctions between assumptions and requirements in measurement in the Social sciences", in Keats, J.A, Taft, R., Heath, R.A, Lovibond, S (Eds), *Mathematical and Theoretical Systems*, Elsevier Science Publishers, North Holland, Amsterdam, pp.7-16.

Andrich, D. (1982). An index of person separation in latent trait theory, the traditional KR.20 index, and the Guttman scale response pattern. *Education Research and Perspectives*, 9, 95-104.

Andrich, D. (1978a). Application of a psychometric model to ordered categories which are scored with successive integers. *Applied Psychological Measurement, 2*, 581-594.

Andrich, D. (1978b). A rating formulation for ordered response categories. *Psychometrika, 43*, 561-573.

Andrich, D. (1980). Using latent trait measurement models to analyse attitudinal data. In Spearits, D. (Ed). *The improvement of Measurement in Education and Psychology: Contributions of Latent Trail Theories. Australia: Australian Council for Educational Research.*

Anene, G.U. & Ndubuisi, O.G. (1992). Test development process in Nworgu, B.G. (Ed.). *Educational Measurement and Evaluation: Theory and Practice. Nsukka: Hallman publishers.*

Angoff, W. H. (1993). Perspectives on differential item functioning methodology. In P. W. Holland and H. Wainer (Eds.), *Differential item functioning (pp 3-23). Hillsdale, NJ: Lawrence Erlbaum*

Apple, M. and Zenk, C, (1996). American realities: Poverty, economy, and education. Cultural Politics and Education. 68-90.

Aremu, A.O and Sokan, B.O. (2003). A multi-casual evaluation of academic performance of Nigerian learner, issues and implications for national development. and Ayodele, B.O, Nwazuoke and Oladiran, O. Education this millennium Ibadan: *Macmillan Nig. Ltd.*

Aremu, A.O. (2004). Psychological and sociological determinant of academic achievement of Nigeria adolescents in Ife Psychologia. *An international journal of psychology in Africa 12(2), 149-161.*

Aremu, A. O. (2000). Impact of Home, School and Government on Primary School pupil's academic performance. *Journal of the exceptional child 5(1), 106-110*

Baghaei, P. & Amrahi, N. (2011). Validation of a Multiple Choice English Vocabulary Test with the Rasch Model. *Journal of Language Teaching and Research*, *2*(5), 1052-1060.

Baghaei, P. (2008). The Rasch model as a construct validation tool. *Rasch Measurement Transaction*, 22 (1), 1145-1146. Available:http://www.rasch.org/rmt/rmt221a.htm

Baker, F. B. (2001). The basics of item response theory ERIC clearinghouse on assessment and evaluation, *University of Maryland, College park, MD.*

Baker, F. B. & Kim, S.H. (2004). Item response theory: *Parameter estimation techniques 2nd ed. New York: Marcel Dekker.*

Barnard, J. J. (1996). In search of equity in education measurement: *Traditional versus modern equating methods. Paper presented at ASEESA's national conference in September 1996 at HSRC Conference Centre, Pretoria, South Africa.* http://www.aseesa-du.co.zaljee.htrn.

Barton, M. A. & Lord, F. M. (1981). An upper asymptote for the three parameter logistic item response model. *Princeton, N. J.: Educational Testing Services*

Beck, C. T., & Gable, R. K. (2001). Item response theory in affective Instrument development: An illustration. *Journal of Nursing Measurement, 9*, 5-22.

Beglar, D. (2010). A Rasch-based validation of the Vocabulary Size Test, *Language Testing*, *27*, 101-118.

Bejar, I.I. (1983). Achievement testing: recent advances. Beverly Hills, CA Sage.

Bessong, F. E. & Obo, U.B. (2005). Students' Performance in Science, Technology and Mathematics in the era of globalization. *Cross River University of Technology, Akpankpa.*

Bezruczko, N. (Ed.). (2005). *Rasch measurement in health sciences*. Maple Grove, MN: JAM Press.

Bielinski, J., Thurlow, M., Minnema, J. & Scott, J. (2000). How out-of-level testing affects the psychometric quality of test scores. (Out-of-level testing project report 2), Minneapolis MN: University of Minnesota, national centre on educational outcomes. *Retrieved October 1st, 2012 from http://www.education.unm.edu/ NCEOPubs/OOLT2.html.*

Binet, A., & Simon, T. (1905). Methodes nouvelles pour le diagnostic dunieveau intellectual anormoux {New methods for the diagnosis of levels of intellectual abnormality]. *Annee Psychologique, 11*, 191-244.

Birnbaum, A. (1968). Some latent trait models and their uses in inferring an examinee's ability. In F. M. Lord & M. R. Novick, Statistical theories of mental test scores. *Reading, M.A: Addison-wesley. 395-479*

Bock, R. D. (1997). A brief history of item response theory. *Educational Measurement: Issues and Practice, 16*, 21-33.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: application of an EM algorithm. *Psychometrika, 46*, 443-459.

Bode, R. K., & Wright, B. D. (1999). Rasch Measurement in Higher Education. In Smart, J. C., & Tierney, W.G. (Eds.), *Higher Education: Handbook of Theory and Research, Volume XIV*. NY: Agathon Press.

Bond, T. G., & Fox, C. M. (2001). *Apply the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.

Bond, T. G. (2003). Validity and assessment: a Rasch measurement perspective. *Metodologia de las Ciencias del Comportamiento* 5(2), 179-194.

Bond T. G., & Fox, C. M. (2007). (2nd ed.) *Applying the Rasch model: Fundamental measurement in the human sciences*. Lawrence Erlbaum.

Boone, W., & Scantlebury, K. (2006). The role of Rasch analysis in Science education utilizing multiple choice tests. *Science Education*, *90*, 253-269.

Brekke, S. E. (1994). Some factors affecting students' performance in physics, *Illinois, U.S.A.*

Bryc, W. (2001). Rasch model analysis of placement exams on a single difficulty scale. *Journal of the Royal statistical society 34(1), 42-54.*

Camminatiello, I.. Gallo, M. & Menini, T. (2011). The Rasch model for Evaluating Italian student performance. *Jounal of applied quanatitative methods 5(2), 331-349*

Casanova, F. P., Garcia-Linares, M.C., Torre, M.J., and Carpio, M.V., (2005). Influence of family and socio-demographic variables on students with low academic achievement. Educational Psychology. 25(4), 423-435.

Charleston, W. V., (2004). Clossing the Achievement Gap. *A focus on Rural schools*. Retrieved on May 18, 2005 from *http://www.educationalalliance.org/ruralschoolspdf*

Chen, W., & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics, 22*, 265-289.

Chuesathuchon, C. & Waugh, R. F. (2008). Item Banking with Rasch Measurement: An Example for Primary Mathematics in Thailand. *The proceedings of the EDU. COM international Conference Edith Cowan University Perth Western Australia 19-21, November 2008, 105- 117.*

Converse, J. M., & Presser, S. (1986). Survey questions. Newbury Park, CA: DeVellis, R. F. (1991). Scale development: *Theory and applications. Newbury Park, CA: Sage.*

Cooke, D. J., & Michie, C. (1997). An item response theory analysis of the Hare Psychopathy Checklist - Revised. *Psychological Assessment, 9*, 3-14.

Cox, D., Sproles, K. E. & Sproles, G. B. (1988). Learning style variations between Rural and Urban students. *Research in rural education 5(1): 27-31*

Cressic, N. (1983). Characterizing the manifest probabilities of latent trait models. *Psychometrika 48 (1). 129-141.*

Crocker, L. & Algina, J. (1986). Introduction to Classical & Modern Test Theory. Harcourt Brace Jovanovich, Inc: Orlando, FL.

Danesty, A. H. (2004). Psychosocial Determinants of Academic Performance and Vocational Learning of Students with Disabilities in Oyo State. Unpublished PhD Thesis, University of Ibadan.

Douglas, F. M., Khavari, K. A., and Farber, P.D., (1979). A comparison of Classical and latent trait item analysis procedures. Journal of Educational and Psychological Measurement, 36, 337-351

Eckes, T. (2011). Item banking for C-tests: A polytomouse Rasch modeling approach. *Psychological test and assessment modeling 53(4), 414-439.*

Embretson. S. E. & Reise. S. P. (2000). Item response theory for Psychologists. *Mahwah, New Jersey:Lawrence Erlbaum Associates.*

Ezebor, S. (1983). Tests, Evaluation and Performance in Nigeria. *Ile-Ife: University of Ife* Fabunmi, M. & Adeniji, I.A. (2007). Class factors as determinants of secondary school students' academic performance in Oyo State, Nigeria. *Journal of Social sciences 14(3), 243-247.*

Fan, X. (1998). Item Response Theory and Classical Test Theory: An empirical comparison of their item/person statistics.*Educational and Psychological Measurement 58, 357-381.*

Federal Ministry of Education, Science and Technology (1989). *Handbook of continuous Assessment Ibadan: Heinemann (Nig.) Ltd.*

Ferguson, G. A. (1943). Item selection by the constant process. *Psychometrika, 7*, 19-29.

Field, A. P. & Hole, G. J. (2012). How to Design and Report Experiments. *London: Sage publications*

Fischer, G. (1968). *Psychologische test theorie* [Psychological test theory]. Bern: Huber.

Fischer, G.H. & Molenaar, I.W. (1995). *Rasch Models: Foundations, Recent Developments, and Applications*. New York: Springer.

Fisher, W., (2010). IRT and confusion about Rasch measurement. *Rasch measurement Transaction.* 14(4), 771-774

Fisher, W. P., Jr., & Wright, B. D. (Eds.). (1994). Applications of probabilistic conjoint measurement. *International Journal of Educational Research*, 21(6), 557-664.

Flannery, W. P., Reise, S. P., & Widaman, K. F.(1995). An item response theory analysis of the general and academic scales of the self-description questionnaire II. *Journal of Research in Personality*, *29*, 168-188.

Fox, C. M., & Jones, J. A. (1998). Uses of Rasch modeling in counseling psychology research. *Journal of Counseling Psychology, 45*, 30-45.

Forde, B. (2004). Quality of education for Botswana. *A keynote address presented on the occasion of Botswana educational research association biennial general meeting, 27 Novembcr 2004 at University of Botswana. Garborne.*

Ford, D. Y., Harris, J. J. III., (1997). A Study of the racial identity and Achievement of Black males and females. Roeper Rev. 20, 10

Frantom, C. G. (2001). *Paternalism and the myth of perfection: Test and measurement of a theory underlying physicians' perceptions of patient autonomy*. Unpublished doctoral dissertation, University of Denver, Denver, CO.

Frantom, C. G., Green, K. E., & Lam, T. C. M. (2002). Item grouping Effects on invariance of attitude items. *Journal of Applied Measurement, 3*, 38-49.

Goh, C. & Aryadout, S. V. (2010). Investigating the Construct Validity of the MELABListening Test through the Rasch Analysisi and Correlated Uniquiness Modelling. *Spaan Fellow working papers in second or foreigh Language Assessment 8, 31-68.*

Golino, H. F, Gomes, C. M. A.,Commons, M. L., & Miller, P. M. (2012). The Construction and Validation of a Developmental Test for Stage Identification: Two Exploratory Studies. *A Journal of the laboratory for cognitive Architecture Mapping (Laico) Universidade Federal de Minas Gerais Brazil, 1- 43.*

Goldstein, H., & Blinkhorn, S. (1982). The Rasch model still does not fit. *British Educational Research Journal, 8(2), 167-170.*

Green, K. E. & Frantom, C. G. (2002). Survey Development and Validation with the Rasch model. *A paper presented at the international conference on questionnaire, development, evaluation and testing, Charleston, SC, November 14- 17, pp 3-8*

Grosse, M. E. & Wright, B. D. (1986). Setting, evaluating, and maintaining certification standards with the Rasch model. *Evaluation and the health professions, 9 (3), 267-285.*

Guiton, R. M. & lronson, G. H. (1983). Latent trait theory for organization research. *Organizational Behaviour and human Performance 31, 54-87.*

Gyll, S. P. (2013). Advantages of the Rasch Measurement model in Test Development Analysis. *Retrieved on the 24 March 2014 from www.rasch.org/demo.ng*

Hambleton, R. K. & Swaminathan, H. (1985). Item response theory: Principles and applications. *Kluwer Nijhoff Publishing: Boston, MA.*

Hambleton, R. K. Swaminathan, H, & Rogers, H.J. (1991). *Item response theory: Principles and applications. Boston: Kluwer-Nijhoff.*

Hambleton, R. (1991). The advantages of item response theory how it can be obtained only when the fit between the IRT model used and the test data of interest is satisfactory. *A poorly fitting IRT model will not yield invariant item and ability parameter.*

Hambleton, R. K. (1995). Meeting the measurement challenges of the 1990s and beyond. A new assessment models methods. In T. Oakland and R.K. Hambleton (Eds.) *International perspectives on academic assessment.. Boston: Kluwer Academic Publishers.*

Hambleton, R.K. (2000). Emergence of item response theory modeling in instrument development and data analysis. *Medical Care, 38 (9). 60-65.*

Hambleton, R. K., Swaminathan, H, & Rogers, H.J. (1991). *Fundamentals of Item Response Theory. Newbury Park, CA: Sage Press*

Hambleton, R.K., Swaminathan, H., Cook, L.L., Eignor, D.R. & Guilford, LA. (1978). Development in latent trait theory, models, technical issues and applications. *Review of Educational Research 48. 467-510.*

Hambleton, R. & Jones, R. (1995): Item bias review. Practical Assessment Research and Evaluation 4 (6). *Retrieved August 19, 2004 from http://www.PAREonline.net/getvn.asp?v-4andn-6.*

Hambleton, R .K & Jones, R. W. T. (1993) Comparison of classical test theory and item response theory and their applications to test development. *An NCME instructional module 16, 1993. Retrieved on 06/10/04 from http://www.ncme.org/pubs/itcms/24.pdf.*

Harris, D. (1989). Comparison of 1,2,3 parameter 1RT models: *America college testing program, NCME instructional model P. O Box 168, IOWA City IA 52243.*

Hashway, W. (1998). A Regional and local item response theory based test item bank system. *Portland public schools, department of research and evaluation, Portland, Oregon.*

Hays, R. D., Morales, L. S., & Reise, S. P. (2000). Item response theory and health outcomes measurement in the 21st century. *Medical Care*, *38* (9 Supplement), 28-42.

Henning, G., Hudson, T. & Turner, J. (1985). Item response theory and the assumption of unidimensionality for language test. *Language testing*, 2(2), 141-154.

Herman, J. (1990) . Item writing techniques. In Walberg, H. and Haertel G.D. (Eds.). *The International Encylopedia of Educational Evaluation. New York: Pergonion Press.*

Herrmann-Abel, C. F. & Deboer, G. E. (2011). Using distractor-driven standard based multiple choice assessments and Rasch modeling to investigate hierarchies of chemistry misconceptions and detect structural problems with individual items. *Journal of Chemistry education research and practices 12, 184-192.*

Hopkins, T. M,. (2004). Gender issues in Mathematics' Achievement in Tennesse: Does rural school location matter? *A Ph.D dissertation, University of Tennesse, Knoxville.*

Howes, E. V., (2002). Connecting girls and science constructivism, feminism and education reform. *New york: Teachers College Press.*

Hui, Y. T., Nyean, C. S., Jiun, Y. W., Rasak, N. & Ahmad, Z. K. (2012). Development of Mathematics question banking system for secondary school in Malaysia. *Journal of international conference on management and education innovation 37, 12-16.*

Idowu, C.B. (2002). Students' performance in the environmental aspect of integrated science: *A case study of the Oyo State College of Education, Oyo.*

lgbokwe, E. (2002). Development of Mathematics item bank for placement of students into Federal unity schools in Nigeria. *An unpublished Ph.D Thesis University of Ibadan*

International Centre for Educational Evaluation(ICEE),(2003/1982).Educational research in Nigeria: *Institute of Education, University of Ibadan.*

Ironson, G. H. (1982). Using item response theory to measure bias. *In R.K. Harnbleton (Ed.) applications of item response theory, Vancouver, BC. Educational institute of British Columbia.*

IRT from SSI: (2003). BILOG-MG, MULTILOG, PARSCALE, TESTFACT edited by Math; Ida du Toit SSI. *Scienfic software international.*

Jackson, T. R., Draugalis, J. R., Slack, M. K. & Zachry, W. M. (2002). Validation of authentic performance assessment. A process suited for Rasch modeling. *American journal of pharmaceutical education 66, 233-243*

Johnson, A. B., (2011). School Mapping and Resource supply as correlate of students Academic Achievement in Kwara state secondary schools. *Journal of language Teaching and Research. 2(5), 1052-1060*

Judith Runnels & Hiroshima Bunkyo. (2012). Using the Rasch model in validating multiple choice test items in English language. *Women's University, Japan International Journal of Language Studies* (IJLS), Vol. 6(4), 141-153

Karabatsos, G. (2000). A critique of Rasch residual fit statistics. *Journal of Applied Measurement,1*, 152-176.

Kerlinger, F.N. & Lee H.B. (2000). Foundations of behavioural research. Philadelphia. *Harcourt College Publishers.*

Kundiger-Schildkmp (1976). Advantages and possibilities of application of linear logistic test models in the area of social sciences. *Theory and decision, 7, 107-118.*

Lawley, D. N. (1943). On problems connected with item selection and test construction. *Proceedings of the Royal Society of Edinburgh*, 62(1),74-82.

Lazarsfeld, P. F. (1950). The logical and mathematical foundation of latent structure analysis. In S. A. Stouffer, L. Guttman, E. A. Suchman, P. F. Lazarsfeld, S. A. Star, & J. A. Clausen, *Measurement and Prediction*. 362-412. New York: Wiley.

Li-Grining, C. P., (2007). Effortful control among low-income preschoolers in three cities: Stability, change, and indifferences. *Developmental Psychology.43(1), 208-221.*

Linacre, J. M. (1996). The Rasch model cannot be “disproved”! *Rasch Measurement Transactions, 10 (3), 512-514.*

Linacre, J. M., & Wright, B. D. (1999-2001). Winsteps. *University of Chicago: MESA Press.*

Linacre, J. M. (2009). Understanding Rasch measurement: Optimizing rating scale category effectiveness. *Journal of Applied Measurement, 3*, 85-106.

Linacre, J. M. (2007). A user's guide to WINSTEPS-MINISTEP: Rasch model computer programs. Chicago, IL: winsteps.com.

Linacre, J. M. (2004). Test validity and Rasch measurement: Construct, content etc. *Rasch Measurement Transactions*, *18*(1), 970-971. Available: htpp://www.rasch.org

Linacre, J. M. (2004). Discrimination, Guessing and Carelessness Asymptotes: Estimating IRT parameters with Rasch. *Rasch measurement Transaction, 18(1), 959-960*

Linacre, J. M. (2012). *A user's guide to WINSTEPS-MINISTEP: Rasch-model computer programs*. Chicago, IL: winsteps.com.

Linacre, J. M. (2013). *A user's guide to WINSTEPS-MINISTEP: Rasch-model computer programs*. Chicago, IL: winsteps.com.

Linacre J. M.(2008). *A User's Guide to Winsteps/Ministeps Rasch Model Computer Programs.*

Linden, W. J., & Hambleton, R. K. (Eds.) (1997). *Handbook of modern item response theory*. New York, NY: Springer-Verlag.

Livingston, S.A. (1988). Reliability of test results. In Keeves J.P. (Ed.). Educational research, methodology and measurement. *An International handbook. New York: Pergainon Press.*

Lord, F.M. (1952). A theory of test scores, *Psychometric Monograph No. 7.*

Lord, F. M. (1977). Practical Applications of Item Characteristic Curve Theory; *Journal of Educational Measurement*, 14, 117-138.

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

Lord, F.M. & Novick, M.R. (1968). *Statistical theories of mental test scores. Reading, M.A. Addison-Wesley*

Louis, C., Lawrence, M., & Keith, M. (2010). Research Methods in Education (6th Ed.). *New York: Routledge*

Lunz, M. E., Stahl, J. A., and James, K. (1989). Content validity revisited: Transforming job analysis data into test specifications. *Evaluation and the Health Professions, 12(2), 192-206.*

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47(2), 149 174.*

Masters, G. N., & Wright, B. D. (1997). The partial credit model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of Modern Item Response Theory* (pp. 101-121). New York, NY: Springer-Verlag Inc.

Masters, G. N., & Keeves, J. P. (Eds.). (1999). *Advances in measurement in educational research and assessment*. New York: Pergamon.

McDonald, J. (2001*).* The susceptibility of item responses to instructions for completion. *Paper presented at the annual meeting of the American Educational Research Association, New Orleans, April.*

McNamara, T. F. (1996). Measuring second language performance. *New York: Longman.*

Mellenbergh, G. J. (1994). A unidimensional latent trait model for continuous item responses. *Multivariate Behavioral Research, 29*, 223-236.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (pp. 13-103). New York: Macmillan.

Messick, S. (1996). Validity and washback in language testing, *Language Testing*, 13(3), 241 256.

Messick, S. (1989). Validity. In R.L. Linn (ed.) *Educational measurement* (pp. 13-103). New York: Macmillan.

Mulligan, J., Mitchelmore, M., & Prescott, A. (2005). Case studies of children's development of structure in early mathematics: A two-year longitudinal study. In H.

L. Chick & J. L. Vincent (Eds.). *Proceedings of the 29th Conference of the International Group for the Psychology in Mathematics Education* 4, 1-8.

Musa, C. A., Mahmud, Z. & Ghani, M. A. (2012). Application of Rasch measurement model in validation and analysis of measurement instrument in statistical education research. *Journal of researches in education, centre for statistical studies. 3(2), 123-129.*

National Policy on Education (2004). Federal Ministry of Information, Abuja.

National Examinations Council (NECO)(2007-2012). Regulations and syllabuses for senior secondary certificate examination (SSCE) for candidates in Nigeria.

NCES. (2000). *America's kindergartners (NCES 2000070)*. Washington, DC: National Center for Education Statistics, U.S. Government Printing Office.

Ndukwe, J.O. (1990). The development and preliminary validation of an Integrated science achievement test (ISAT). *Unpublished Masters Dissertation, University of Nigeria Nsukka.*

Nenty, H.J. (2000). Some factors that influence students' pattern of responses to Mathematics examination items. *BOLESWA: Journal of Educational Research 17, 47-58.*

Nenty, H. J. (2001). Research in educational assessment: Challenges for the 21 century. In J.P.Lefoka and L.M. Matsoso (Eds.). *Educational research for development: challenges for 21st century 39-54. Maseru, Lesotho: Lesotho Educational Research Association and Institute of Education.*

Nenty, H. J. (2004). Designing measurement instruments for assessment and research in education. *A paper accepted for publication in a book series by Akwa Ibom State College of Education, Afaha Nsit, AKS, Nigeria.*

Nenty, H.J. (2004). From classical test theory to item response theory: an introduction to a desirable transition. A paper submitted for publication consideration as a chapter in issues in educational measurement and evaluation in Nigeria. *A book of reading Institute of Education, University of Ibadan, Nigeria*

Nenty, H. J. (2004). The application of item response theory in strengthening assessment role in the implementation of national education policy.

Nenty, H. J. (2005). Item response theory: *Quality-enhancing measurement technique in education.*

Nkpone, H. L. (2001). Application of latent trait models in the development and standardization of a physics achievement test for senior secondary students.

Nworgu, B. G. & Agha, J. J. (2012). Application of three-parameter logistics model in the calibration of a mathematics Achievement Test. Journal of Educational Assessment in Africa, 7. 161-171

Nworgu, B. G., (2011). Differential Item Function: A critical issue in Regional Quality Assurance. *A paper presented in NEARA conference.*

Nworgu, B.G. and Odili, J.N. (2005). Analysis of DIF in WAEC SSCE Biology Multiple Choice Test. *Review of Education 16,* (2) 140-152.

Nworgu, B.G. & Harbor-Peters, V.F. (1990). The development and validation of physics achievement test (PAT). *Journal of Science Institution and Curriculum Studies 15, 33-39.*

Obemeata, J.O. (2000). Principles of essay and multiple choice test construction. *Ibadan. Stirling-Horden Publisher (Nig) Limited.*

Obinne, A. D. E., Nworgu, B. G.. & Umobong, M. E. (2013). An investigation into Differential Item Functioning (DIF) of Tests Conducted by the two major Examinations Bodies in Nigeria. *Journal of Advances in Educational Research. 2(1): 001-008*

Oche, E. S. (2012). Issues in Test Items Bias in Public Examinations in Nigeria and Implications for Testing. International *Journal of Academic Research in Progressive Education and Development. 1(1), 175-187*

Odili, J. N. (2013). Writing Unidimensional Test Items in Achievement Testing. Delsu Journal of Educational Research and Development 12(1), 33-45

Odili, J. N. (2010). Modern Test Theory. *An unpublished lecture note, Delta state university.*

Odili, J. N. (2005). Effect of languaguage manipulation on differential item functioning in a multicultural setting. *A paper presented at the 27th annual conference of the association for education assessment in Africa, Yaounde Cameroon, August 24th – 28th, 2009.*

Odili, J.N. (2005a). DIF: A challenge to psychological measurement. *The Educational Psychologist 1*(2), 204-214.

Odili, J.N. (2005b). Quality assurance in credentialing: Analysis of DIF in 2001 WAEC SSCE Biology multiple choice test. *Studies in Education 8*, 102-109

Ogbebor, U. & Onuka, A. (2013). Differential Item Functioning Method as an Item Bias Indicator. *International Research Journal of Educational Research. 4(4), 367-373.*

Ogochukwu, R. N. (1990). The development and validation of an instrument (Shyness Scale) for the identilication of shyness. *Unpublished Master's Dissertation, University of Nigeria, Nsukka.*

Ogunniyi, M.A. (2004). Development, validation and use of university students evaluation of teaching effectiveness scale (USETES).

Ojerinde, D. Popoola, K. Ojo, F. & Onyeneha, P. (2013). Introduction to Item Response Theory (IRT) parameters models estimation and application. *Marvelouse Mike press Ltd*

Ojerinde, D. (2013). Classical test theory (CTT) versus Item Response Theory (IRT): An evaluation of the comparability of item analysis results. *Lecture presentation at the institute of education, University of Ibadan, May 23, 2013*

Okpala, P.N., Onocha, C.O. & Oyedeji,O.A.(1993). Measurement and Evaluation in Education.

Okorodudu, R. I. (2013). Research methods and statistics: A Practical Approach for undergraduate and postgraduate studies in Education, Social Sciences, Psychology and other Disciplines concerned with Empirical Research. *University printing press, DELSU.Abraka.*

Okorodudu, K. I. (2012). Understanding Educational and Psychological Measurement and Evaluation with CTT, GTT and IRT Theories. *University printing press, DELSU.Abraka.*

Oladokun, V.O. & Adebanjo, A.T. (2008). Predicting students' academic performance using artificial neural network: *A case study of an engineering. University of Ibadan,Ibadan. 9(1) http ://www.akamaiuniversity.us/PJST.htm*

Olaleye, O. O. & Aliyu, R. T. (2012). Relationship between parental educational level and students' Mathematics Achievement in Nigeria. *Oyo state journal of Mathematical Association of Nigeria (OSJMAN) 3(2), 251-271*

Olaleye, O. O. & Aliyu, R. T. (2013). Development and Validation of Mathematics Achievement test using Item Response Theory (IRT) Models in attaining quality education for national development. *Proceedings of September, 2013 Annual Conference of Mathematics Association of Nigeria (MAN) 82- 95.*

Olaleye, O. O. & Aliyu, R. T. (2012). Relationship between continuous Assessment and the Examination Scores of Students in JSCE for Attaining Nigerian Vision 20:2020. *Proceedings of September, 2012 Annual National Conference of Mathematics Association of Nigeria (MAN) 36-46.*

Olaleye, O. O.,(2004). Some Psychio-social determinant of secondary school female students Achievement in Mathematics in Osun and Oyo states Nigeria. *An unpublished Ph.D Thesis, university of Ibadan.*

Opasina, O. C. (2009). Development and validation of alternative to practical Physics test using item response theory model. *An unpublished Ph.D thesis, University of Ibadan.*

Opong, N. K. (2006). Construction and Validation of Economics Achievement Test. *Unpublished M.Ed Dissertation, Delta State University, Abraka, Nigeria.*

Orlando, M., Sherbourne, C. D., & Thissen, D. (2000). Summed-score linking using item response theory: application to depression measurement. *Psychological Assessment, 12*, 354-359.

Osadebe, P. U. & Aliyu, R. T. (2013). Construct validity of Mathematics Test Items using the Rasch model. *Paper submitted for publication in the University of Ibadan, Ibadan.*

Osadebe, P. U. (2010). Construction and Validation of Test Items. *An unpublished lecture note, Delta state university.*

Osadebe, P. U. (2001). Construction and Validation of Economics Achievement Test. *An unpublished Ph.d thesis, University of Porthartcourt*

Osiobe, G. A. (2012). Construction and Standardization of Geography Objective Test for Senior Secondary Schools in Delta State. *Unpublished Ph.D Dissertation, Delta State University, Abraka, Nigeria.*

Osadebe, P. U. (2014). Construction of Economics Achievement Test for Assessment of Students. *Journal of World Journal of Education. 4(2), 58-64*

Osman, S. A., Naam, S. I., & Jaafar, O. (2012). Application of Rasch Model in measuring students' performance in civil Engineering Design II course. *Centre for Engineering Education Research Department of Civil and Structure Engineering, University Kebangsaan Malysia.*

Oyekan, S.O. (1993). The effect of diagnostic remedial teaching strategies on students' exams in Biology. *An unpublished Ph.D. Thesis, Obafemi Awolowo University, Ile-Ife. Nigeria.*

Peter, A. & Stone, M. (2009). Progress in the Technology of Measurement: Applications of The Rasch models. *Psychological Assessment, 12*, 305-319.

Pishghadam, R., Baghaei, P. & Ahmadi, H. S. (2011). Development and Validation of an English Language Teacher Competency Test Using Item Response Theory. *International Journal of Educational and psychological Assessment. 8(2), 54-68*

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Chicago: MESA.

Rasch, G. (1966). An item analysis which takes individual differences into account. *British Journal of Mathematical and Statistical Psychology 19(1), 49-57.*

Rasch, G. (1980). Probabilistic models for some intelligence and attainment Tests (Copenhagen Danish institute for educational research), expanded edition 1980 with forward and afterward by Wright, B.D. Chicago: *The University of Chicago Press*

Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. Copenhagen: anish Institute for Educational Research.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press.

Rasch, G. (1960). *Probabilistic models for some intelligence and Attainment tests*. Copenhagen: Danish Institute for Educational Research.

Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests*. (Copenhagen, Danish Institute for Educational Research), expanded edition (1980) with foreword and afterword by B.D. Wright. Chicago: The University of Chicago Press.

Reeve, B. B. (2000). Item-and scale-level analysis of clinical and non-clinical sample responses to the MMPI-2 depression scales employing item response theory. *Unpublished doctoral Thesis. University of North Carolina at Chapel Hill*.

Reeves, B. B., (2011). The Effect of school location on Mathematics and science Achievement Trend. *A Prime on Growth modeling in Education policy Research. Occupational Research paper, No. 11*

Reise, S. P. (1999). Personality measurement issues viewed through the eyes of IRT. In S. B. Embretson & S. L. Hershberger (Eds.), the new rules of measurement: *What every Psychologist and educator should know. 219-242. Mahwah, New Jersey: Lawrence Erlbaum Associates.*

Reise, S. P., & Waller, N. G. (1990). Fitting the two-parameter model to personality data. *Applied Psychological Measurement, 14*, 45-58.

Richardson, M. W. (1936). The relationship between difficulty and the differential validity of a test. *Psychometrika, 1*, 33-49.

Reporting statistics in psychology (2012). General guidelines. *Extracted from the website www.amherst.edu November, 20th 2013.*

Saher, A. S., Mey, S. C. & Lan, O. S. (2010). The using of Rasch- model in validating the Arabic version of Multiple Intelligence Development Assessment Scale (MIDAS). International journal of Human & Social Science 5(9): 574-579.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometric monograph, No. 17, Psychometrika, 34 (2), 1-97.*

Samejima, F. (I972). A general model for free-response data. *Psychometric Monograph, No. 18.*

Santor, D. A., & Ramsay, J. O. (1998). Progress in the technology of measurement: applications of item response models. *Psychological Assessment, 10*, 345-359.

Sarama, J., Clements, D. H. & Liu, X. (2012). Development of a measure of early Mathematics Achievement using the Rasch model: The early Mathematics Assessment. *Journal for research in Mathematics education 38, 1-28.*

Smith, R.M. (1996). A comparison of the Rasch separate calibration and between fit methods of detecting item bias. *Educational and Psychological Measurement 55(3), 403-417.*

Smith, E. V., Lawless, K. A., Curda, L. & Curda, S. (1999). Measuring Change inefficacy. *Popular measurement 2, 31-33.*

Smith, E. V. (2001). Evidence for the reliability of measures and validity of measure interpretation: A Rasch measurement perspective. *Journal of Applied Measurement,* 2(3), 281-311.

Smith, R. M. (2004). Fit analysis in latent trait measurement models. In E. V. Smith & R. M. Smith (Eds.), *Introduction to Rasch measurement* (pp. 73-92). Maple Grove, MN: JAM Press.

Smith, R. M., Schneider, B. H., & Ruck, M. D. (2005). Thinking about making it: Blanck Canadian.

Smith, R. M. (1992). *Application of Rasch measurement*. Chicago: MESA

Smith, R. M., Schumacker, R. E., & Bush, J. M. (1995). *Using item mean squares to evaluate fit to the Rasch model*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Smith, R.M. (1990). Theory and practice of fit. Rasch Measurement Transactions, 3(4), 78. http://rasch.org/rmt/rmt34b.htm

Snyder, S., & Sheehan, R. (1992). The Rasch measurement model: An introduction. *Journal of Early Intervention, 16*, 87-95.

Spelke, E. S., (2005). Sex differences in intrinsic Aptitude for Mathematics and science: Attention and Performance. *20, 29-56. New York, Oxford University Press.*

Steinberg, L. & Thissen, D. (1995). Item response theory in personality Research. In P. E. Shrout and S. T. Fiske (Eds.), *Personality research, methods and theory: a festschrift honoring Donald W. Fiske, 161-181. Hilldale, NJ: Erlbaum.*

Steinberg, L,. Dornbusch, S.M.and Brown, B.B. (1992). Ethnic differences in adolescent achievement: An ecological perspective. American Psychology. (47), 723-729.

Steinberg, J. (2000). Frederic Lord, Who Devised Testing Yardstick, Dies at 87. New York Times, February 10, 2000.

Stelmack, J., Szlyk, J.P., Stelmack, T., Babcok-Parziale, J., Demers Turco, P., Williams, R.T., Massof, R.W. (2004). Use of Rasch person-item map in exploratory data analysis: A clinical perspective, *Journal of Rehabilitation Research & Development*, 41(2), 233-242.

Sudman, S., & Bradburn, N. M. (1982). Asking questions. SF: Jossey-Bass.

Tennant, A. (2000). Rasch perspectives on cross cultural validity. *Measurement of healthcare outcomes*. Symposium conducted at the Third International Outcomes Measure Conference, Chicago, IL.

Teresi, J. A., Kleinman, M., & Ocepek-Welikson, K. (2000). Modern psychometric methods for detection of differential item functioning: application to cognitive assessment measures. *Statistics in Medicine, 19*, 1651-1683.

Thissen, D. (1991). *Multilog User's Guide. Multiple, categorical item analysis and test scoring using Item Response Theory*, Version 6.3. Chicago, IL: Scientific Software International.

Thissen, D., Nelson, L., Billeaud, K. & McLeod, L. (2001). Chapter 4-Item response theory for items scored in more than two categories. In D. Thissen & H. Wainer (Eds.), *Test Scoring*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Thissen, D. & Orlando, M. (2001). Chapter 3-Item response theory for items scored in two categories. In D. Thissen & H. Wainer (Eds.), *Test Scoring(pp 73-140)*. Hillsdale, NJ: Erlbaum Associates, Inc.

Thissen, D. & Steinberg, L. (1988). Data analysis using item response theory. *Psychological Bulletin, 104*, 385-395.

Thissen, D., Steinberg, L., & Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 67-113). Hillsdale, NJ: Erlbaum.

Thompson, S. & Fleming, N. (2003). Summing it up: Mathematics Achievement in Australian Schools in TIMSS 2002 (TIMSS Australia Monograph no. 6). ACER: Camberwell, Vic.

Tracy, C. & Walter, B. (1998). In Asikhia O. (2010). A students and teachers' perception of the causes of poor academic performance in Ogun state secondary schools Nigeria: Implications for couselling for national development. European Journal of Social Sciences. (13) 2.

Thorndike, R.L. (1990). Reliability. In Keeves, J.P. (Ed.) Educational research, methodology and measurement. *An International handbook: New York: Pergamon Press*

Thorndike, R.M. (1997). Measurement and evaluation in educational psychology. *New Jersey. Prentice Hall.*

Thurstone, L. L. (1925). A method of scaling psychological and educational tests. *Journal of Educational Psychology* , *16*, 433-451.

Tucker, L. R. (1946). Maximum validity of a test with equivalent items. *Psychometrica, 11, 1 14.*

UNESCO (2003). Gender and Education for all: the leap for equality. *Global Monitoring Report*

Urry, V.M. (1974). Approximation to item parameters of mental test models and their uses. *Educational and Psychological Measurement 34, 253-269.*

U.S. Census Bureau. (2000). National household education survey. Washington, DC: National Center for Education Statistics.

Usoro H. & Ekpo K. (2002). Predicting students' performance in Mathematics from their performance in Physics and Chemistry in the Nigerian secondary schools Cross River University of Technology. Calabar.

Vanderlinden, W. I. (1981). A latent trait look at the pretest post test validation criterion referenced test items. *Review of Educational Research*

Verhelst, N.D. and Glas, C.A.W. (1995). *The one parameter logistic model.* In G.H. Fischer and I.W. Molenaar (Eds.), Rasch Models: Foundations, recent developments, and applications (pp. 215–238). New York: Springer Verlag.

Verhelst, N.D., Glas, C.A.W. and Verstralen, H.H.F.M. (1995). *One parameter logistic model (OPLM).* Arnhem: CITO.

Wainer, H., & Thissen, D. (2001). Chapter 2-True score theory. In D. Thissen & H. Wainer (Eds:). *Test Scoring. Hillsdale, NJ: Erlbaum*

Weiss, D.J. & Kingsbury, G.G. (1999). New horizons in testing: latent trait test theory and computerized adaptive testing. *New York. Academic Press.*

Weiss, D. J. (1995). Improving individual differences measurement with item response theory and computerized adaptive testing. In D. J. Lubinski, & R. V. Dawis (Eds.), *Assessing individual differences in human behavior: New concepts, methods, and findings* (pp. 49-79). Palo Alto, CA: Davies-Black Publishing.

West African Examination Council (WAEC) 2007- 2012.

West African Examinations Council (WAEC) 2012 Chief Examiners' Report in Mathematics, Lagos.

Williams, J. H., (2005). Cross-National variation in Rural Mathematics Achievement: A descriptive Overview. *Journal of Research in Rural Education 20, 1-8.*

Wistner, B. (2012). Applying the Rasch Model in the Validation of a Test of Metalinguishtics knowledge. *Language Teaching research. 2, 101-107*

Wolfe, E. W. & Smith, E. V. (2007). Instrument development tools and activities for measure validation using Rasch models: Part II-validation activities. *Journal of Applied Measurement, 8*(2), 204-234.

Wolfe, E. W., & Chiu, C. W. T. (1999). Measuring change across multiple occasions using the Rasch rating scale model. *Journal of Outcome Measurement, 3*, 360-381.

Wright, B. D. (1967). Sample-free test calibration and person measurement. Institute for Objective Measurement: www.rasch.org/memo1.htm.

Wright, B. D. (1977). Solving measurement problems with the Rasch model. Journal of Educational Measurement, 14 (2), 97 – 116.

Wright, B. D., Mead, R. and Draba, R. (1976). Detecting and correcting test item bias with a logistic response model. *MESA Research Memorandum Number 22,* MESA Psychometric Laboratory. www.rasch.org/memo22.htm.

Wright, B. D. & Stone, M. H (1977). Best Test Design: Rasch Measurement. Mesa Press: Chicago, IL

Wright, B.D. (1992). *IRT in the 1990s: Which Models Work Best?* Rasch measurement transactions, 6(1), 196-200.

Wright, B. D., & Masters, G. N. (2002). Number of person or item strata, *Rasch Measurement Transactions*, *16*, 888.

Wright, B. D. & Stone, M. H. (1979). Best test design. Chicago: MESA Press.

Wright, R. J. (2003). A mathematics recovery: Program of intervention in early number learning. *Australian Journal of Learning Disabilities, 8*(4), 6-11.

Wright, B. D., & Stone, M. H. (1998*). Best test design*. Chicago: MESA Press.

Wright, B. D., & Linacre, J. M. (1985). Microscale Manual. Westport, Conn.: Mediax Interactive Technologies, Inc.

Wright, B. D. (1998). Solving measurement problems with the Rasch model. *Journal of Educational Measurement, 14*, 97-116.

Wright, B. D.(1997).A history of social science measurement. *Educational Measurement: Issues and Practice, 16*, 21-33.

Wright, B. D. & Panchapakesan, W. (1969). A procedure for sample-free item analysis. *Educational and Psychological Measurement 29. 23-48.*

Wright, B. D. (1997). A history of social science measurement. Educational Measurement: *Issues and Practice, 16, 21-33.*

Wu, M. & Adams, R. (2007). *Applying the Rasch model to psychosocial measurement: A practical approach*. Melbourne: Educational Measurement Solutions.

Yoloye, T. W. (2004). Selecting a better option in assessing learning outcomes.

Zickar, M. J., & Drasgow, F. (1996). Detecting faking on a personality instrument using appropriateness measurement. *Applied Psychological Measurement, 20*, 71-87.

APPENDIX I

TABLE OF SPECIFICATION (TEST BLUE PRINT)

Table of specification is constructed in order to test the content validity of the items drawn. A table of specifications showing a one hundred and fifty test items in a Mathematics Achievement Test (MAT)

CONTENT AREA	INSTRUCTIONAL OBJECTIVES						Total
		Kno	Com	Appl	Ana	Synt	Eval	
48%Num., Numeratn.&Algebra	%	25%	10%	15%	20%	10%	20%	100%
Number Base	15	11		32	22	23	48	15
Indices	12	10	2		3		1	12
Algebraic Expression	9	12	31				8	9
Angles	9	17				13	20	9
Fractions	12	5		7	34		21	12
Sets	6	25	39					6
Ratio & Rates	6		14	15				6
Subject of the Formula	6	30			44			6
Percentages	6				16	42		6
36%Measuratn.,Trig.& Geometr								
Simplifications	6			4			9	6
Variations	6	49					24	6
Circles	6	29				33		6
Sequences	6			38	36			6
Geometry	9	18			27		28	9
Quadratics Expression	12	35		26	50		6	12
Inequalities	3				47			3
Angle of Elevation	3			41				3
Bearings	3			37				3
16%Stat.,Prob.&Further Mat								
Probability, Further Maths topics	6		19			40		6
Long & Lat/logarithms	3	45						3
Statistics	6				46		43	6
Total 100%	150	36	15	24	30	15	30	150

APPENDIX II

DELTA STATE UNIVERSITY (DELSU)

COUNSELLING PSYCHOLOGY DEPARTMENT

FACULTY OF EDUCATION

ABRAKA CAMPUS 1

Time: 3½ hrs

MATHEMATICS ACHIEVEMENT TEST-(MAT)

This instrument is purely for research purpose only. The information you give will be treated confidentially. Kindly shade the correct option on the answer sheet provided. Attempt all the questions.

SECTION A: Personal Information

i. Name of your School:

ii. Gender: Male ( ) Female ( )

iii. Age in years

iv. Place of residence - Rural ( ) Urban ( )

v. School type – Federal ( ) State ( ) Private ( )

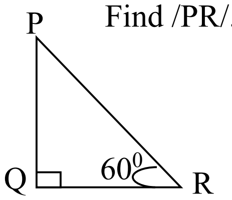
SECTION B: - MAT

Attempt all the questions.

1. If (x - 3) is a factor of $2x^3 + 3x^2 - 17x - 30$, find the remaining factors A. (2x – 5)(x – 2) B. (2x – 5)(x + 4) C. (2x + 5)(x – 2) D. (2x + 5)(x + 2)
2. Given that $\frac{1}{8^{2-3y}} = 2^{y+2}$. Find y A. $\frac{1}{5}$ B. $\frac{7}{8}$ C. 1 D. $1\frac{1}{5}$
3. A committee of 4 is to be selected from a group of 5 men and 3 women. In how many ways can this be done if the chairman of the committee must be a man? A. 15 B. 40 C. 70 D. 175
4. Simplify $\frac{{}^nP_4}{{}^nC_4}$ A. 24 B. 18 C. 12 D. 6
5. Find the coefficient of X^4 in the binomial expansion of $(1 - 2x)^6$ A. 320 B. 240 C. 250 D. 230
6. Evaluate $\log_{0.25} 8$ A. $\frac{3}{2}$ B. $\frac{2}{3}$ C. $-\frac{2}{3}$ D. $-\frac{3}{2}$
7. A particle is projected vertically upwards from a height 45metres above the ground with a velocity of 40m/s. How long does it take to hit the ground? [Take g = 10m/s] A. 1s B. 3s C. 7s D. 9s
8. A binary operation on the set of real numbers is defined by M * n = $\frac{Mn}{2}$ for all m, n ε R. If the identity element is 2, find the inverse of -5 A. $-\frac{4}{5}$ B. $-\frac{2}{5}$ C. 4 D. 5
9. Evaluate $\int_1^2 (x^2 - 4x)dx$ A. $\frac{11}{3}$ B. $\frac{3}{11}$ C. $-\frac{3}{11}$ D. $-\frac{11}{3}$
10. The distance between the point (4, 3) and the intersection of y = 2x + 4 and y = 7 – x is A. $\sqrt{13}$ B. $3\sqrt{2}$ C. $\sqrt{26}$ D. $10\sqrt{5}$
11. If $y = x^2 - \frac{1}{x}$, find $\frac{dy}{dx}$ A. $x^2 - \frac{1}{x^2}$ B. $2x + \frac{1}{x^2}$ C. $2x - x^2$ D. $x^2 + \frac{1}{x}$

12. The locus of a point equidistant from the intersection of lines 3x – 7y + 7 = 0 and 4x – 6y + 1 = 0 is a A. line parallel to 7x + 13y + 8 = 0 B. circle C. semi circle D. bisector of the line 7x + 13y + 8 = 0
13. A man stands on a tree 150cm high and sees a boat at an angle of depression of 74^0. Find the distance of the boat from the base of the tree A. 52cm B. 43cm C. 40cm D. 15cm
14. Find the remainder when $2x^3 - 11x^2 + 8x - 1$ is divided by x + 3 A. -871 B. -781 C. -187 D. -178
15. The probabilities that a man and his wife love for 80years are $\frac{2}{3}$ and $\frac{3}{15}$ respectively. Find the probability that at least one of them will live up to 80 years A. $\frac{2}{15}$ B. $\frac{3}{15}$ C. $\frac{7}{15}$ D. $\frac{13}{15}$
16. Two forces $F_1 = (7i + 8j)N$ and $F_2 = (3i + 4j)N$ act on a particle. Find the magnitude and direction of $(F_1 - F_2)$ A. $(4\sqrt{2}N, 00^0)$ B. $(4\sqrt{2}N, 45^0)$ C. $(4\sqrt{2}N, 90^0)$ D. $(4\sqrt{2}N, 180^0)$
17. A stone is thrown vertically upwards and its height at any time *t* seconds is $h = 45 - 9t^2$. Find the maximum height reached A. 45.25m B. 45.50m C. 56m D. 56.25m
18. The initial velocity of an object is u = $\binom{-5}{3} m/s$. If the acceleration of the object is a = $\binom{3}{4} m/s^2$ and it moved for 3 seconds. Find the final velocity A. $\binom{-14}{15} m/s$ B. $\binom{-2}{1} m/s$ C. $\binom{4}{-9} m/s$ D. $\binom{-14}{-9} m/s$
19. The sum and product of the roots of a quadratic equation are $\frac{4}{7}$ and $\frac{5}{7}$ respectively. Find its equation A. $7x^2 - 4x - 5 = 0$ B. $7x^2 - 4x - 5 = 0$ C. $7x^2 + 4x - 5 = 0$ D. $7x^2 - 4x + 5 = 0$
20. Simplify $(\sqrt{6} + 2)^2 - (\sqrt{6} - 2)^2$ A. $2\sqrt{6}$ B. $4\sqrt{6}$ C. $8\sqrt{6}$ D. $16\sqrt{6}$
21. In a class of 46 students 22 play Football and 26 play Volley ball. If 3 students play both games, how many play neither? A. 1 B. 2 C. 3 D. 4
22. The nth term of a sequence is $n^2 - 6n - 4$. Find the sum of the 3rd and 4th terms A. 24 B. 23 C. -24 D. -25
23. The value of y for $\frac{1}{5}y + \frac{1}{5} < \frac{1}{2}y + \frac{2}{5}$ is A. $y > \frac{2}{3}$ B. $y < \frac{2}{3}$ C. $y > -\frac{2}{3}$ D. $y < -\frac{2}{3}$
24. Two bodies of mass 8kg and 5kg traveling in the same direction with speed x m/s and 2 m/s respectively collide. If after collision, they move together with a speed of 3.85m/s, find correct to the nearest whole number, the value of x A. 2 B. 5 C. 8 D. 13
25. Calculate in surd form, the value of tan 15^0 A. $(2 + \sqrt{3})$ B. $(1 + \sqrt{3})$ C. $(\sqrt{3} - 1)$ D. $(2 - \sqrt{3})$
26. If P = $\begin{pmatrix} 1 & -2 \\ 3 & 4 \end{pmatrix}$ and Q = $\begin{pmatrix} -2 & 3 \\ 1 & 0 \end{pmatrix}$, find PQ A. $\begin{pmatrix} 4 & 1 \\ -2 & 9 \end{pmatrix}$ B. $\begin{pmatrix} -4 & 1 \\ 2 & 9 \end{pmatrix}$ C. $\begin{pmatrix} -4 & 3 \\ -2 & 13 \end{pmatrix}$ D. $\begin{pmatrix} -4 & 3 \\ -2 & 9 \end{pmatrix}$
27. Evaluate $\lim_{x \to 3} \left(\frac{x^2 - 2x - 3}{x - 3}\right)$ A. 4 B. 3 C. 2 D. 0
28. If $\begin{vmatrix} 4 & x \\ 5 & 3 \end{vmatrix}$ = 32, find the value of x A. 4 B. 2 C. -2 D. -4

29. The distance *s* metres of a particle from a fixed point at time t seconds is given by s = 7 + $pt^3 + t^2$, where p is a constant. If the acceleration at t = 3 seconds is $8m/s^2$, find the value of p. A. $\frac{1}{3}$ B. $\frac{4}{9}$ C. $\frac{5}{9}$ D. 1
30. Find the value of Cos $(60^0 + 45^0)$ leaving your answer in surd form A. $\frac{\sqrt{6}+\sqrt{2}}{4}$ B. $\frac{\sqrt{3}+\sqrt{6}}{4}$ C. $\frac{\sqrt{2}-\sqrt{6}}{4}$ D. $\frac{\sqrt{3}-\sqrt{6}}{4}$
31. Find the equation of tangent to the curve $y = 4x^2 - 12x + 7$ at point (2, -1) A. y + 4x – 9 = 0 B. y – 4x – 9 = 0 C. y – 4x + 9 = 0 D. y + 4x + 9 = 0
32. If $y = 2(2x + \sqrt{x})^2$, find $\frac{dy}{dx}$ A. $2\sqrt{x}(2x+\sqrt{x})$ B. $4(2x+\sqrt{x})(2+\frac{1}{2\sqrt{x}})$ C. $4(2x+\sqrt{x})(2+\sqrt{x})$ D. $8(2x+\sqrt{x})(2+\sqrt{x})$
33. The third term of geometric progression (G.P) is 10 and the sixth term is 80. Find the common ration A. 2 B. 3 C. 4 D. 8
34. Calculate, correct to one decimal place, the length of the line joining points x (3, 5) and y (5, 1) A. 4.0 B. 4.2 C. 4.5 D. 5.0
35. Simplify $\frac{\log_5 8}{\log_5 \sqrt{8}}$ A. -2 B. -$\frac{1}{2}$ C. $\frac{1}{2}$ D. 2
36. Evaluate ${}^{n+1}C_{n-2}$ if n = 15 A. 360 B. 3360 C. 1120 D. 560
37. In how many ways can the letters of the word TOTALITY be arranged? A. 6720 B. 6270 C. 6207 D. 6027
38. Evaluate $\int_0^{\frac{\pi}{4}} \sec^2 \theta d\theta$ A. 1 B. 2 C. 3 D. 4
39. Simplify $\frac{27^{\frac{1}{3}}}{16^{\frac{-1}{4}}}$ A. 6 B. 5 C. 4 D. 3
40. In $\triangle$PQR, <PQR is a right angle, /QR/ = 2cm and < PRQ = 60^0. Find /PR/. A. $4\sqrt{3}cm$ B. $4cm$ C. $2\sqrt{3}cm$ D. $5\sqrt{3}cm$



41. For what values of x is the expression $\frac{x-5}{x^2-2x-3}$ not defined? A. 3, 1 B. -1, -3 C. -1, 3 D. 3, -2
42. The sides of a right –angled triangle in ascending order of magnitude are 8cm, (x – 2)cm and x cm. Find x A. 16 B. 17 C. 34 D. 90
43. If y = $\sqrt{ax-b}$ express x in terms of y, a and b A. x = $\frac{y^{2-b}}{a}$ B. x = $\frac{y+b}{a}$ C. x = $\frac{y-b}{a}$ D. x = $\frac{y^2+b}{a}$
44. If $k\sqrt{28} + \sqrt{63} - \sqrt{7} = 0$, find k A. -2 B. -1 C. 1 D. 2
45. Solve the inequality 2x + 3 < 5x A. x > 1 B. x < $\frac{3}{7}$ C. x > $\frac{3}{7}$ D. x <1
46. Solve the equation $3y^2 = 27y$ A. y = 0 or 3 B. y = 0 or 9 C. y = -3 or 3 D. y = 3 or 9
47. Given that $\log_4^x$ = -3, find the value of x A. $\frac{1}{81}$ B. $\frac{1}{64}$ C. 64 D. 81

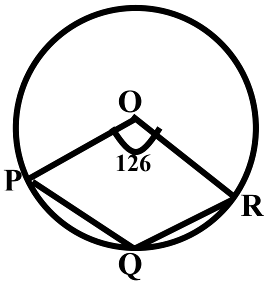
48. The sum to infinity of a G.P is A. $\frac{a}{1-r}$ B. r^2 C. r - $\frac{1}{2}$ D. r – ½
49. The bearing $S40^0E$ is the same as A. 040^0 B. 050^0 C. 130^0 D. 140^0
50. Simplify $\frac{2x-1}{3} - \frac{x+3}{2}$ A. $\frac{x+7}{3}$ B. $\frac{x+8}{6}$ C. $\frac{x-11}{6}$ D. $\frac{x-4}{6}$
51. Simplify 4/x +1 – 3/x-1
(A. X + 7 /X^2–1 B. X – 1 /X^2+ 1 C. X – 7/X^2–1 D. X – 11 /X^2+ 1)
52. Evaluate without using tables $(0.2)^3$ x 30/$(0.4)^2$
(A. 11/2 B. 21/5 C. 22/5 D. 2)
53. Factorize the expression : x(a-c) + y(c-a)
(A. (a-c)(y-x) B. (a-c)(x-y) C. (a-c)(x+y) D. (a+c)(x+y)
54. Given that m = 3 and n = 2, find the value of $3n^2 – 3m^3$/m
(A. –22 B. –15 C. 14 D. -31)
55. What is the value of x1 if the expression x^2+15x+50/x–5 is not defined.
(A. 0. B. 1 C. 2 D. 5)
56. Simplify log √8/log8
(A. 1/3 B. ½ C. 1/3log√2 D. ½ log √2)
57. Given that 27(1+x) = 9. find x.
(A. –3 B. –1/3 C. 5/3 D. 2)
58. Simplify 11011_{two} - 1101_{two}
A. 101000_{two} B. 1100_{two} C. 1110_{two} D. 1011_{two}
59. Factorize $3p^2$ + 2p-1
A. (3p-2)(p-1) B. (3p-2)(p+1) C. (3p+1)(p-1) D. (3p-1)(p+1)
60. The values of three angles at a point are $3y-45^0$; $y+25^0$ and y^0. find the value of y.
(A. 40 B. 58 C. 68 D. 76)
61. Ngozi sold an article for N 1755.00 and made a profit of 35%. Find the cost price.
(A. N 2,370 B. N 1300 C. N 614.25 D. N 614.00)
62. Find the compound interest on 450 in 2yrs at 5%.
(A) # 497) (B) 496.13 (C) 45.00 (D) 46.13}
63. The monthly salary of a man increase from #2,700 to 3,200. Find the percentage increase.
(A) 10% (B) 15% (C) 15.6% (D) 18.5%)
64. The sum of the interior angles of a polygon is 1260^0. Find the number of its sides.
(A) 7 (B) 8 (C) 9 (D) 10.)
65. What is the total surface area of a cube of side 4cm?
(A) $36cm^2$ (B) $64cm^2$ (C) $96cm^2$ (D) $144cm^2$)
66. A boy threw two dice at once what is the probability of having a total of six? (A) 1/9 (B) 3/8 (C) 4/35 (D) 2/7)
67. The angles of a triangle are in the ratio 5:3:2 what is the size of the smallest angle?
(A) 90^0 (B) 72^0 (C) 54^0 (D) 36^0)
68. Simplify the fraction ¼ ÷½ / ¾
(A) 1 (B) 2/3 (C) 8/3 (D) 3/8)
69. Convert 1001101 to a number in base ten
(A) 61 (B) 46 (C) 45 (D) 44.)
70. What is the value of x if $(101)_x=172_8$
(A) 5 (B) 4 (C) 11 (D) 2)
71. If x-3 is directly proportional to the square of y and x=5 when y=2, find x when y=6.
(A) 30 (B) 21 (C) 16 (D) 12)
72. Let u- {a,b,c,d}, z=[c,d] and y(a,b,c) what is (zny)′
(A). (a,b,c) (B). (a,b,d) (C). (b,c) (D). (a,d))
73. What must be added to the expression x^2-18x to make it a perfect square?

(A. 3 B. 9 C. 39 D. 72)

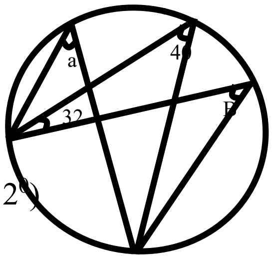
74. If x + 3 varies directly as y and x= 3 when y = 12, what is the value of x when y = 8?

A. 1 B. $\frac{1}{2}$ C.- $\frac{1}{2}$ D. -1 E. 5

75. The area of a parallelogram is 573cm^2 and the height is 19cm. Calculate the base.
(A. 13.5cm . 25cm C. 27cm D. 54cm)
76. In the diagram, O is the center of the circle and
<POR = 126^0. Find <PQR.
(A. 117^0 B. 72^0 C. 63^0 D. 54^0)



77. If P = 3/5√qr, express q in terms of p and r.
((A). 9/25pr^2 (B). 9/25p^2 r (C). 25/9p^{2}r (D). 25/9pr^2)
78. Find (x – y), if 4x – 3y = 7 and 3x – 2y = 5.
(A. 4 B. 3 C. 2 D. –2)
79. If $512_x - 354_x = 125_x$. Find the number base x
(A. 6 B. 7 C. 8 D. 9)
80. Find the lettered angles in the figure ii below.
0 is the center of the circle.
(A. a=b=32^0 B. a=32^0, b=40^0 C. a=b=40^0 D. a=40^0, b=32^0)



81. Given that 6x – y/x + 2y =2, find the value of x/y
(A. 3/8 B. 5/8 C. 4/5 D. 5/4)
82. Find the quadratic equation whose roots are x = -2 or x=7
(A. X^2 + 2X–7 =0 B. X^2–5X – 14 =0 C. X^2 + 5X + 14 = 0
D. X^2 + 5X – 14)
83. The fifth term of an A.P is 24 and the eleventh term is 96, find the first term.
(A. 12 B. 4 C. -12 D. -24)
84. The bearing of a point X from a point Y is 074^0. What is the bearing of Y from X?
(A. 106^0 B. 148^0 C. 164^0 D. 254^0.)
85. Find the 4th term of an A.P whose first term is 2 and the common difference is 0.5.
(A. 0.5 B. 25 C. 3.5 D. 4)
86. Let U = (1,2,3,4), P=(2,3) and Q = (2,4) what is (PnQ)' ?
(A. (1,2,3) B. (1,3,4) C. (2,3) D. (1,4)
87. Three balls are drawn one after the other with replacement from a bag containing 5 red, 9 while and 4 blue identical balls. What is the probability that they are one red, one white and one blue?
(A. 5/102 B. 5/136 C. 5/204 D. 5/162)
88. The angle of elevation of the top of a tower from a point on the horizontal ground, 40m away from the foot of the tower is 30^0 find the height of the tower.
(A. 20m B. 40 √ 3m/3 C. 20√3m D. 60m)
89. A trader bought 100 tubers of yam at 5 for #350.00. She sold them in sets of 4 for #290.00. Find her gain percent.
(A. 3.6% B. 3.5% C. 3.4% D. 2.5%)
90. What is the mode of the numbers 8,10,9,9,10,8,11,8,10,9,8 and 14?
(A. 8 B. 9 C. 10 D. 11)
91. If h(m + n) = m (h + r) find h in terms of m, n and r.
(A. h= mr/2m + n B. h = mr/n-m C. h= m+r/n D. h= mr/n
92. The radius of a Geographical globe is 60cm. Find the length of the parallel of latitude 600N.

(A. 66πcm B. 60πcm C. 30πcm D. 15πcm)

93. Find the mean deviation of 2, 4, 5 and 9.
(A. 1 B. 2 C. 5 D. 7)

94. Solve the inequality 2x + 3 < 5x.
(A. π>1 B. x < 3/7 C. x > 3/7 D. x > -1)

95. Express as a single fraction $\frac{x}{x-2} - \frac{x+2}{x+3}$. A. $\frac{2x^2-3x-4}{(x-2)(x+3)}$. B. $\frac{2x^2+3x-4}{(x-2)(x+3)}$. C. $\frac{2}{(x-2)(x+3)}$. D. $\frac{3x+4}{(x-2)(x+3)}$

96. Express 12.625ten in base two. A. 101.1.0. B. 101.110. C. 1100.011. D. 1100.101

97. If X = {all prime factors of 44} and Y = {all prime factors of 60}. The elements of X U Y and XnY. A. {2, 4, 3, 5, 11} and {4}. B. {4, 3, 5, 1} and {3, 4}. C. {2, 5, 11} and {2}. D. {2, 3, 5, 11} and {2}.

98. Factorize $X^2 + 4x - 192$. A. (x – 4) (x + 48). B. (x – 48)(x + 4). C. (x – 12) (x + 16). D. (x – 12)(x – 16)

99. The curved surface of a cylindrical tin is 704cm^2. If the radius of its base is 8cm. Find the height. (Take ∏ = 22/7). A. 14cm. B. 9cm. C. 8cm. D. 7cm

100. Make P the subject of the relation: q = $\frac{3p}{r} + \frac{s}{2}$. A. $\frac{2q-rs}{6}$. B. p = 2qr –sr-3. C. p = $\frac{2qr-s}{6}$. D. p = $\frac{2qr-rs}{6}$

101. If P = {Prime factors of 21}, and Q = {prime numbers less than 10}, find P n Q. a. {1, 2, 3}. B. {2, 3, 5}. C. {1, 3, 5, 7}. D. {2, 3, 5, 7}

102. The sum of the interior angles of a regular polygon is 1800°. How many sides has the polygon? A. 16. B. 12. C. 10. D. 8

103. If N2, 500 amounted to N3, 500 in 4 years at simple interest, find the rate at which the interest was charged. A. 35%. B. 7 ½ %. C. 8%. D. 10%

104. Given that tan x = 1 where 0° ≤y≤90°. evaluate $\frac{1-Sin^2x}{Cos\ x}$. A. $2\sqrt{2}$. B. $\sqrt{2}$. C. $\frac{\sqrt{2}}{2}$. D. ½.

105. A rectangle has length X cm and width (x – 1) cm if the perimeter is 16cm. Find the value of x. a. 3 ½ cm. b. 4cm. c. 4 ½ cm. d. 5cm.

106. A cylindrical container has a base radius of 14cm and height 18cm. how many litres correct to the nearest litre of liquid can it hold? (Take ∏ = 22/7). A. 11. B. 14. C. 16. D. 18

107. An arc of a circle of radius 14cm subtends angle 300° at the centre. Find the perimeter of the sector formed by the arc (Take ∏= 22/7). A. 14. 67cm. B. 42.67cm. C. 101.33cm. d. 513.33cm.

108. A regular polygon has 9 sides, what is the size of one of its exterior angles? A. 20°. B. 40 °. C. 90 °. D. 140 °

109. The area of a square field is 110.25m ^{2}. Find the cost of fencing it round at N75 per metre. A. N1, 575. B. N3, 150. C. N4, 734.30. d. N8, 268.75.

110. From the diagram below, find the value of x. A. 40. B. 50. C. 25. D. 20

111. If P $\alpha \frac{1}{q}$ which of the following is true? A. q α p^2 . B. q α $\frac{1}{P^2}$. C. q α $\sqrt{P}$. D. q $\alpha \frac{1}{p}$

112. If y = 23_{five} + 101_{three} find y leaving your answer in base two. A. 1110. B. 10111. C. 11101. D. 111100

113. Evaluate $202_{three} - 1122_{three}$. A. 21120. B. 21121. C. 21112. D. 21011.

114. Expand (2x – 3y)(x-5y). A. $2x^2 - 13xy - 15y^2$. B. $2x^2 + 13xy - 15y^2$. C. $2x^2 + 13xy + 15y^2$. D. $2x^2 - 13xy + 15y^2$

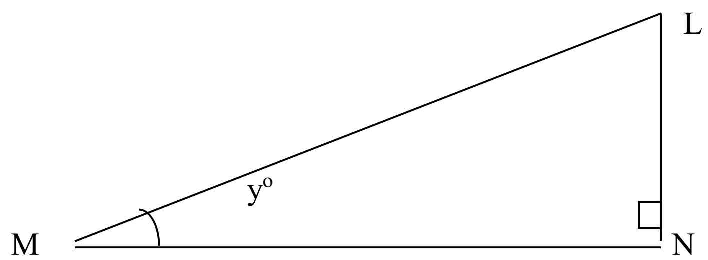
115. A machine valued at N20, 000 depreciates by 10% every year, what will be the value of the machine at the end of two years? A. N16, 200. B. N8000. C. N14, 200. D. N12, 000

116. For what value of x is the expression $\frac{2x-1}{x+3}$ not defined? A. 3. B. 2. C. ½. D. -3

117. If x varies inversely as y and x = 2/3 when y = 9, find the value of y when x = ¾ (A. 1/18 B. 8/81 C. 9/2 D. 8)

118 Given that (2x + 7) is a factor of $2x^2 + 3x - 14$, find the other factor.
(A. x+2 B. 2 – x C. x – 2 D. x + 1)

119. In the diagram below |LN| = 4cm $L\hat{N}M$ = 90° and tan y = 2/3. What is the area of the LMN? A. 24cm^2. B. 12cm^2. C. 10cm^2. D. 6cm^2.



120. Find the volume of a solid cylinder with base radius 10cm and length 14cm (Take ∏ = 22/7). A. 220cm^3. B. 880cm^3. C. 1400cm^3. D. 440cm^3

121. Find the volume of a cone of radius 3.5cm and vertical height 12cm (Take ∏ = 22/7). A. 3.4cm^3. B. 15.5cm^3. C. 21.0cm^3. D. 154cm^3

122. If sin X = 12/13 where 0º < x < 90º, find the value of 1-cos^{2}x. A. 25/169. B. 64/169. C. 105/169. D. 144/169

123. Given that tan x = 5/12, what is the value of sin X + Cos X?. a. 5/13. B. 7/13. C. 12/13. D. 17/13.

124. A house bought for N100, 000 was later auctioned for N80,000. Find the loss percent. A. 20%. B. 30%. C. 40%. D. 60%

125. If N varies directly as M and N = 8 when M = 20. Find M when N = 7. A. 13. B. 15. C. 7 ½. D. 18 ½.

126. A solid cylinder of radius 7cm is 10cm long. Find its total surface area. A. 70∏cm^2. B. 189∏cm^2. C. 210∏cm^2. D. 238∏cm^2

127. The height of a right circular cone is 4cm. The radius of its base is 3cm. Find its curved surface area. A. 9∏cm^2. B. 15∏cm^2. C. 16∏cm^2. D. 20∏cm^2

128. If (x+3) varies directly as y and x = 3 when y = 12, what is the value of x when y = 8? A. 1. B. ½. C. – ½ . d. -1

129. The lengths of the parallel sides of a trapezium m are 9cm and 12cm. If the area of the trapezium is 105cm^2, find the perpendicular distance between the parallel sides. A. 5cm. B. 7cm. C. 10cm. D. 15cm

130. The base diameter of a cone is 14cm and its volume is 462 cm^3. Find its height. Take ∏=22/7 (a. 0.75cm b. 2.25cm c. 0.25cm d. 2.05cm)

131. A student bought 3 notebooks and 1 pen for N35. After misplacing these items, she again bought 2 note books and 2 pens all of the same type for N30. What is the cost of a pen? A. N5. B. N7.50. c. N10. D. N15.

132. A cooperative society charges an interest of 5 ½ % per annum on any amount borrowed by its members. If a member borrows N125, 000. How much does he pay back after one year? A. N136, 875. B. N131, 875. C. N128, 750. D. N126, 250

133. The sides of two cubes are in the ratio 2:5 what is the ratio of their volume? A. 4:5. B. 8:15. C. 6:125. D. 8:125

134. What is the volume of a solid cylinder of diameter 7cm and height 7cm? (Take ∏=22/7). A. 38.5cm^3. B. 77cm^3. C. 269.5cm^3. d. 1078cm^3.

135. Ladi sold a car for N84, 000 at a loss of 4%. How much did Ladi buy the car?

A. N80,500. B. N80, 640. C. N87, 360. D. N87, 500

136. The monthly salary of a man increased from N2,700 to N3, 200. Find the percentage increase. A. 10%. B. 15%. C. 15.6%. d. 18.5%.

137. If y $\alpha \frac{1}{x^2}$ and y = 1 ¼ when x = 4 find the value of y when x = ½ . a. 2 ½. B. 5. C. 10. D. 80

138. A rectangular packet has inner dimension 16cm by 12cm by 6cm. How many cubes of sugar of sides 2cm can be neatly packed into the packet? A. 90. B. 144. C. 150. D. 288

139. A baker used 40% of a 50kg bag of four. If 1/8 of the amount was used for cake, how many kilograms of flour was used for cake? A. 2 ½. B. 6 ¼. C.15 $^5/_8$. d. 17 ½.

140. If 30% of y is equal to x, what in terms of x is 30% of 3y? a. x/9. B. x/3. C. x. D. 3x.

141. What is the difference in longitude between P(lat. 50^0N, long. 50^0W) and Q(Lat. 50^0N, long 150^0W)? A. 300^0. B. 200^0. C. 130^0. D. 100^0.

142. If Cos 60° = ½, which of the following angles has a cosine of ½? A. 30o. B. 120^0. C. 150^0. D. 140^0.

143. Find the roots of the equation $2x^2 - 3x - 2 = 0$. A. X= -2 or 1 ½. B. X = -2 or 1. C. X = 1 or 2. D. X = - ½ or 2.

144. Evaluate $\log_{10}6 \ \ \log_{10}45 - \log_{10}27$ without using logarithm tables. A. 0. B. 1. C. 1.3802. D. 10

145. Find the sum of the first five terms of the G.P. 2, 6, 18. A. 484. B. 243. C. 242. D. 130.

146. In a class of 80 students, every student had to study Economics or Geography, or both Economics -and Geography. If 65 students studied Economics and 50 studied Geography, how many studied both subjects? A 15. B. 30. C. 35. D. 45.

147. ***Use the following information to answer questions 7 and 8. (Take*** π = 22/7). A cylindrical container, closed at both ends, has a radius of 7cm and height 5cm

148. Find the total surface area of the container. A. 35cm^2. B. 220cm^2. C. 528cm^2. D. 770cm^2

149. What is the volume of the container? A. 35cm^3. B. 220cm^3. C. 528cm^3. D. 770cm^3.

150. A section of a circle of radius 7cm has an area of 44cm^2. Calculate the angle of the sector, correct to the nearest degree. A. 6°. B. 26°. C. 52°. D. 103°.

APPENDIX III

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	27.264	18.176	18.176	27.264	18.176	18.176	24.411	16.274	16.274
2	23.150	15.433	33.609	23.150	15.433	33.609	23.090	15.393	31.667
3	20.492	13.661	47.270	20.492	13.661	47.270	16.695	11.130	42.797
4	18.310	12.207	59.477	18.310	12.207	59.477	16.444	10.962	53.759
5	15.038	10.025	69.503	15.038	10.025	69.503	15.592	10.395	64.154
6	12.128	8.085	77.588	12.128	8.085	77.588	14.795	9.863	74.017
7	9.270	6.180	83.768	9.270	6.180	83.768	12.911	8.607	82.625
8	6.633	4.422	88.190	6.633	4.422	88.190	7.388	4.926	87.550
9	4.495	2.997	91.187	4.495	2.997	91.187	5.016	3.344	90.894
10	2.080	1.387	92.573	2.080	1.387	92.573	2.000	1.334	92.228
11	1.732	1.155	93.728	1.732	1.155	93.728	1.907	1.271	93.499
12	1.450	.967	94.695	1.450	.967	94.695	1.404	.936	94.435
13	1.275	.850	95.545	1.275	.850	95.545	1.345	.897	95.332
14	1.229	.819	96.364	1.229	.819	96.364	1.298	.865	96.197
15	1.111	.740	97.104	1.111	.740	97.104	1.223	.815	97.012
16	1.079	.720	97.824	1.079	.720	97.824	1.218	.812	97.824
17	.876	.584	98.408						
18	.643	.429	98.836						
19	.542	.362	99.198						
20	.522	.348	99.546						
21	.346	.231	99.776						
22	.186	.124	99.900						
23	.103	.069	99.969						
24	.046	.031	100.000						
25	6.629E-15	4.419E-15	100.000						
26	2.875E-15	1.917E-15	100.000						
27	2.560E-15	1.707E-15	100.000						
28	2.320E-15	1.546E-15	100.000						
29	1.807E-15	1.205E-15	100.000						
30	1.731E-15	1.154E-15	100.000						

31	1.464E-15	9.758E-16	100.000						
32	1.343E-15	8.952E-16	100.000						
33	1.131E-15	7.538E-16	100.000						
34	1.037E-15	6.913E-16	100.000						
35	8.878E-16	5.919E-16	100.000						
36	7.993E-16	5.329E-16	100.000						
37	7.606E-16	5.071E-16	100.000						
38	7.063E-16	4.708E-16	100.000						
39	6.588E-16	4.392E-16	100.000						
40	5.566E-16	3.710E-16	100.000						
41	5.447E-16	3.631E-16	100.000						
42	4.980E-16	3.320E-16	100.000						
43	4.719E-16	3.146E-16	100.000						
44	3.987E-16	2.658E-16	100.000						
45	3.678E-16	2.452E-16	100.000						
46	3.562E-16	2.375E-16	100.000						
47	3.350E-16	2.234E-16	100.000						
48	2.904E-16	1.936E-16	100.000						
49	2.715E-16	1.810E-16	100.000						
50	2.355E-16	1.570E-16	100.000						
51	2.248E-16	1.499E-16	100.000						
52	2.187E-16	1.458E-16	100.000						
53	2.033E-16	1.356E-16	100.000						
54	1.837E-16	1.224E-16	100.000						
55	1.669E-16	1.112E-16	100.000						
56	1.537E-16	1.025E-16	100.000						
57	1.484E-16	9.896E-17	100.000						
58	1.469E-16	9.793E-17	100.000						
59	1.336E-16	8.906E-17	100.000						
60	1.283E-16	8.552E-17	100.000						
61	1.256E-16	8.372E-17	100.000						
62	1.228E-16	8.188E-17	100.000						
63	1.205E-16	8.033E-17	100.000						
64	1.190E-16	7.936E-17	100.000						
65	1.181E-16	7.875E-17	100.000						
66	1.163E-16	7.755E-17	100.000						

67	1.160E-16	7.731E-17	100.000						
68	1.154E-16	7.697E-17	100.000						
69	1.150E-16	7.670E-17	100.000						
70	1.146E-16	7.637E-17	100.000						
71	1.145E-16	7.632E-17	100.000						
72	1.144E-16	7.626E-17	100.000						
73	1.124E-16	7.494E-17	100.000						
74	1.119E-16	7.458E-17	100.000						
75	1.117E-16	7.445E-17	100.000						
76	1.114E-16	7.426E-17	100.000						
77	1.111E-16	7.407E-17	100.000						
78	1.111E-16	7.403E-17	100.000						
79	1.110E-16	7.403E-17	100.000						
80	1.109E-16	7.394E-17	100.000						
81	1.101E-16	7.339E-17	100.000						
82	1.085E-16	7.231E-17	100.000						
83	1.081E-16	7.208E-17	100.000						
84	1.077E-16	7.181E-17	100.000						
85	1.073E-16	7.155E-17	100.000						
86	1.072E-16	7.148E-17	100.000						
87	1.068E-16	7.119E-17	100.000						
88	1.064E-16	7.095E-17	100.000						
89	1.059E-16	7.057E-17	100.000						
90	1.053E-16	7.021E-17	100.000						
91	1.051E-16	7.007E-17	100.000						
92	1.025E-16	6.830E-17	100.000						
93	9.936E-17	6.624E-17	100.000						
94	9.791E-17	6.528E-17	100.000						
95	9.335E-17	6.223E-17	100.000						
96	8.265E-17	5.510E-17	100.000						
97	7.759E-17	5.173E-17	100.000						
98	6.322E-17	4.215E-17	100.000						
99	4.733E-17	3.155E-17	100.000						
100	4.409E-17	2.940E-17	100.000						
101	3.414E-17	2.276E-17	100.000						
102	9.334E-18	6.223E-18	100.000						

103	5.545E-18	3.696E-18	100.000					
104	2.963E-18	1.976E-18	100.000					
105	2.561E-18	1.707E-18	100.000					
106	9.507E-19	6.338E-19	100.000					
107	2.041E-19	1.361E-19	100.000					
108	9.202E-20	6.135E-20	100.000					
109	-7.186E-19	-4.790E-19	100.000					
110	-1.534E-18	-1.023E-18	100.000					
111	-4.002E-18	-2.668E-18	100.000					
112	-5.032E-18	-3.355E-18	100.000					
113	-6.613E-18	-4.408E-18	100.000					
114	-1.149E-17	-7.657E-18	100.000					
115	-1.350E-17	-9.003E-18	100.000					
116	-1.558E-17	-1.039E-17	100.000					
117	-2.156E-17	-1.437E-17	100.000					
118	-3.062E-17	-2.041E-17	100.000					
119	-3.326E-17	-2.218E-17	100.000					
120	-4.031E-17	-2.687E-17	100.000					
121	-4.935E-17	-3.290E-17	100.000					
122	-5.685E-17	-3.790E-17	100.000					
123	-6.544E-17	-4.362E-17	100.000					
124	-8.230E-17	-5.486E-17	100.000					
125	-9.176E-17	-6.117E-17	100.000					
126	-1.015E-16	-6.765E-17	100.000					
127	-1.372E-16	-9.144E-17	100.000					
128	-1.794E-16	-1.196E-16	100.000					
129	-1.864E-16	-1.242E-16	100.000					
130	-2.210E-16	-1.473E-16	100.000					
131	-2.452E-16	-1.635E-16	100.000					
132	-2.980E-16	-1.987E-16	100.000					
133	-3.260E-16	-2.174E-16	100.000					
134	-3.562E-16	-2.375E-16	100.000					
135	-3.736E-16	-2.490E-16	100.000					
136	-4.582E-16	-3.055E-16	100.000					
137	-6.048E-16	-4.032E-16	100.000					
138	-6.201E-16	-4.134E-16	100.000					

139	-6.821E-16	-4.548E-16	100.000						
140	-7.713E-16	-5.142E-16	100.000						
141	-8.860E-16	-5.907E-16	100.000						
142	-9.341E-16	-6.227E-16	100.000						
143	-1.313E-15	-8.754E-16	100.000						
144	-1.372E-15	-9.145E-16	100.000						
145	-2.017E-15	-1.345E-15	100.000						
146	-2.166E-15	-1.444E-15	100.000						
147	-3.520E-15	-2.347E-15	100.000						
148	-4.226E-15	-2.817E-15	100.000						
149	-4.680E-15	-3.120E-15	100.000						
150	-6.018E-15	-4.012E-15	100.000						

Extraction Method: Principal Component Analysis.

APPENDIX IV
ROTATED COMPONENT MATRIX

Rotated Component Matrix[a]

RCM	Component C															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Q68	.994															
Q90	.994															
Q66	.994															
Q64	.994															
Q95	.994															
Q146	.994															
Q62	.994															
Q82	.994															
Q87	.994															
Q115	.994															
Q6	.994															
Q70	.994															
Q75	.994															
Q80	.994															
Q97	.994															
Q113	.994															
Q119	.994															
Q144	.994															
Q85	.994															
Q149	.994															
Q103	.970													.100		
Q101	.907													.314		
Q110	.884				-.108									.393		
Q106	.884				-.108									.393		
Q108	.850				-.132									.457		
Q150		.995														
Q84		.995														
Q116		.995														

Q120		.995														
Q63		.995														
Q73		.995														
Q76		.995														
Q81		.995														
Q89		.995														
Q91		.995														
Q96		.995														
Q118		.995														
Q142		.995														
Q4		.995														
Q61		.995														
Q67		.995														
Q78		.995														
Q86		.995														
Q93		.995														
Q99		.995														
Q71		.995														
Q111		.995														
Q107		.888				-.121	.118							.130		
Q22			.968			.147									-.138	
Q133			.968			.147									-.138	
Q47			.959			.157							.103			
Q31			.959			.157							.103			
Q29			.955			.132							-.103			
Q140			.955			.132							-.103			
Q33			.955			.132							-.103			
Q15			.951			.144							-.157			.116
Q126			.951			.144							-.157			.116
Q51			.944		-.110	.141										-.209
Q28			.942			.127										.260
Q139			.942			.127										.260
Q36			.932		-.107	.157							.200			
Q114			.928			.116										
Q117			.928			.116										
Q25			.921		-.129	.140							.261			

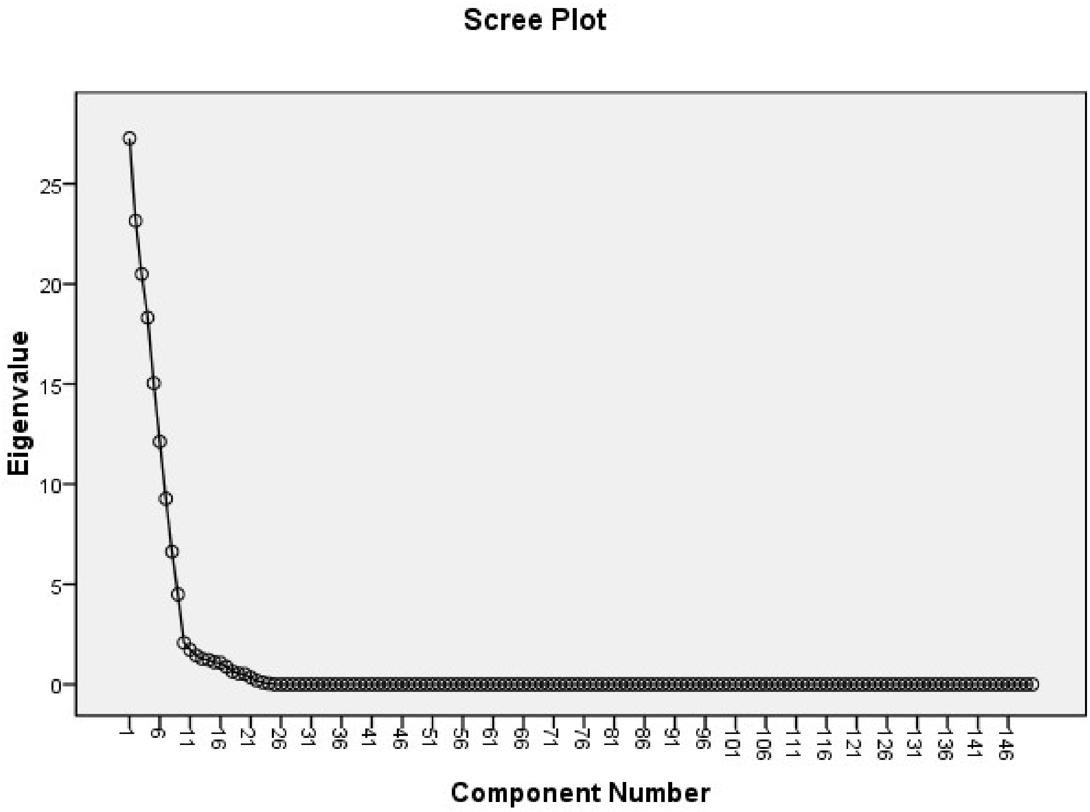
Q136			.921		-.129	.140							.261			
Q39			.908		-.112	.115							-.225		.132	-.191
Q18				.962									-.120		.115	
Q54				.962									-.120		.115	
Q129				.962									-.120		.115	
Q138				.958												.194
Q11				.958												.194
Q27				.958												.194
Q122				.958												.194
Q125				.958												
Q14				.958												
Q35				.947								-.133			-.119	
Q23				.940							.108				-.224	
Q134				.940							.108				-.224	
Q30				.935	-.154					.102					.203	-.115
Q37				.925	-.133										.104	-.266
Q10	.121			.918	-.125							.174				.119
Q127				.916	-.152								.227		.115	-.133
Q16				.916	-.152								.227		.115	-.133
Q57				.907					.103							-.317
Q83					.990											
Q88					.990											
Q148					.990											
Q65					.990											
Q112					.990											
Q145					.990											
Q100					.990											
Q1					.990											
Q69					.990											
Q94					.990											
Q74					.990											
Q79					.990											
Q143					.990											
Q109					.961									.118		
Q105	-.142			-.172	.824									.482		
Q102	-.142			-.172	.824									.482		

Q45			.166			.955			.113				.107			
Q131			.147			.952							.202			
Q20			.147			.952							.202			
Q13			.155			.948						.146				
Q124			.155			.948						.146				
Q42			.183		.104	.942									-.150	
Q40	-.108		.164			.940									-.102	
Q48			.159			.936			.164							-.173
Q53			.161			.934							.169			.110
Q55	-.104		.149			.930			.136			.121	-.131		.113	
Q26	-.136		.158			.923			.101							
Q137	-.136		.158			.923			.101							
Q58			.150			.918										.257
Q49	-.151		.148		.102	.917							-.232			
Q52	-.184		.129			.914				.138			-.135			
Q34	-.179		.141			.898							-.323			
Q56							.962									
Q9		.118					.953			.113						
Q50		.122					.951					.182				
Q60		.122					.951					.182				
Q135							.949						-.251			
Q24							.949						-.251			
Q130							.949									.214
Q19							.949									.214
Q41	-.101						.943						-.153			-.103
Q43							.940									.269
Q21		.148					.940						.202			-.148
Q132		.148					.940						.202			-.148
Q59		.152					.938			-.102			.188			
Q46		.152					.938			-.102			.188			
Q12								.973							-.111	
Q123								.973							-.111	
Q8								.952								.216
Q38								.938					.199			.106
Q32								.934		.118			.149			-.136
Q128					-.124			.919				.187	-.112		.162	-.111

Q17					-.124			.919				.187	-.112		.162	-.111
Q72						.204			.970							
Q92						.204			.970							
Q98						.204			.970							
Q141						.204			.970							
Q77						.204			.970							
Q3	.222			.139						.924						
Q44	.253			.170	.116					.903						
Q104		.185	-.115					.139			.924			.124		
Q147		.248		.126			-.160	.110	.113		.905					
Q2		.179	.154			.163	.236	.204		-.139	.180	.693	-.115	.144		
Q5	-.153				-.231		.168			.138		.669	.117	-.153	-.205	
Q7				.185		-.114	.156		.117			-.129			.785	

EXTRACTION METHOD: *PRINCIPLE COMPONENT ANALYSIS*
ROTATION METHOD: *VARIMAX WITH KAISER NORMALIZATION*
ROTATION CONVERGED IN 7 ITERATION

APPENDIX V



APPENDIX VI

DELTA STATE UNIVERSITY (DELSU)
COUNSELLING PSYCHOLOGY DEPARTMENT
FACULTY OF EDUCATION
ABRAKA CAMPUS 1

Time: 2 ½ hrs

FINAL DRAFT OF MATHEMATICS ACHIEVEMENT TEST-(FDOMAT)

This instrument is purely for research purpose only. The information you give will be treated confidentially. Kindly shade the correct option on the answer sheet provided. Attempt all the questions in the three sections.

SECTION A: Personal Information

i. Name of your School:

ii. Gender:	Male ( )	Female ( )
iii. Age in years	--------------	
iv. Place of residence -	Rural ( )	Urban ( )
v. School type –	Public ( )	Private ( )

SECTION B: - FDOMAT

Attempt all the questions.

1. If (x - 3) is a factor of $2x^3 + 3x^2 - 17x - 30$, find the remaining factors A. $(2x - 5)(x - 2)$ B. $(2x - 5)(x + 4)$ C. $(2x + 5)(x - 2)$ D. $(2x + 5)(x + 2)$
2. Simplify $\frac{^nP_4}{^nC_4}$ A. 24 B. 18 C. 12 D. 6
3. Evaluate $\log_{0.25} 8$ A. $\frac{3}{2}$ B. $\frac{2}{3}$ C. $-\frac{2}{3}$ D. $-\frac{3}{2}$
4. Evaluate $\int_1^2 (x^2 - 4x)dx$ A. $\frac{11}{3}$ B. $\frac{3}{11}$ C. $-\frac{3}{11}$ D. $-\frac{11}{3}$
5. If $y = x^2 - \frac{1}{x}$, find $\frac{dy}{dx}$ A. $x^2 - \frac{1}{x^2}$ B. $2x + \frac{1}{x^2}$ C. $2x - x^2$ D. $x^2 + \frac{1}{x}$
6. The locus of a point equidistant from the intersection of lines $3x - 7y + 7 = 0$ and $4x - 6y + 1 = 0$ is a A. line parallel to $7x + 13y + 8 = 0$ B. circle C. semi circle D. bisector of the line $7x + 13y + 8 = 0$
7. A man stands on a tree 150cm high and sees a boat at an angle of depression of 74^0. Find the distance of the boat from the base of the tree A. 52cm B. 43cm C. 40cm D. 15cm
8. Find the remainder when $2x^3 - 11x^2 + 8x - 1$ is divided by x + 3 A. -871 B. -781 C. -187 D. -178
9. The probabilities that a man and his wife love for 80years are $\frac{2}{3}$ and $\frac{3}{15}$ respectively. Find the probability that at least one of them will live up to 80 years A. $\frac{2}{15}$ B. $\frac{3}{15}$ C. $\frac{7}{15}$ D. $\frac{13}{15}$
10. The initial velocity of an object is u = $\binom{-5}{3} m/s$. If the acceleration of the object is a = $\binom{3}{4} m/s^2$ and it moved for 3 seconds. Find the final velocity A. $\binom{-14}{15} m/s$ B. $\binom{-2}{1} m/s$ C. $\binom{4}{-9} m/s$ D. $\binom{-14}{-9} m/s$
11. The sum and product of the roots of a quadratic equation are $\frac{4}{7}$ and $\frac{5}{7}$ respectively. Find its equation A. $7x^2 - 4x - 5 = 0$ B. $7x^2 - 4x - 5 = 0$ C. $7x^2 + 4x - 5 = 0$ D. $7x^2 - 4x + 5 = 0$
12. Simplify $(\sqrt{6} + 2)^2 - (\sqrt{6} - 2)^2$ A. $2\sqrt{6}$ B. $4\sqrt{6}$ C. $8\sqrt{6}$ D. $16\sqrt{6}$
13. The nth term of a sequence is $n^2 - 6n - 4$. Find the sum of the 3rd and 4th terms A. 24 B. 23 C. -24 D. -25

14. Two bodies of mass 8kg and 5kg traveling in the same direction with speed x m/s and 2 m/s respectively collide. If after collision, they move together with a speed of 3.85m/s, find correct to the nearest whole number, the value of x A. 2 B. 5 C. 8 D. 13
15. Evaluate $\lim_{x \mapsto 3} \left(\frac{x^2-2x-3}{x-3}\right)$ A. 4 B. 3 C. 2 D. 0
16. If $\begin{vmatrix} 4 & x \\ 5 & 3 \end{vmatrix}$ = 32, find the value of x A. 4 B. 2 C. -2 D. -4
17. The distance *s* metres of a particle from a fixed point at time t seconds is given by s = 7 + $pt^3 + t^2$, where p is a constant. If the acceleration at t = 3 seconds is $8m/s^2$, find the value of p. A. $\frac{1}{3}$ B. $\frac{4}{9}$ C. $\frac{5}{9}$ D. 1
18. Find the equation of tangent to the curve $y = 4x^2 – 12x + 7$ at point (2, -1) A. y + 4x – 9 = 0 B. y – 4x – 9 = 0 C. y – 4x + 9 = 0 D. y + 4x + 9 = 0
19. The third term of geometric progression (G.P) is 10 and the sixth term is 80. Find the common ration A. 2 B. 3 C. 4 D. 8
20. Simplify $\frac{\log_5{}^8}{\log_5{}^{\sqrt{8}}}$ A. -2 B. -$\frac{1}{2}$ C. $\frac{1}{2}$ D. 2
21. For what values of x is the expression $\frac{x-5}{x^2-2x-3}$ not defined? A. 3, 1 B. -1, -3 C. -1, 3 D. 3, -2
22. The sides of a right –angled triangle in ascending order of magnitude are 8cm, (x – 2)cm and x cm. Find x A. 16 B. 17 C. 34 D. 90
23. Solve the inequality 2x + 3 < 5x A. x > 1 B. x < $\frac{3}{7}$ C. x > $\frac{3}{7}$ D. x <1
24. Given that $\log_4{}^x$ = -3, find the value of x A. $\frac{1}{81}$ B. $\frac{1}{64}$ C. 64 D. 81
25. Simplify $\frac{2x-1}{3} - \frac{x+3}{2}$ A. $\frac{x+7}{3}$ B. $\frac{x+8}{6}$ C. $\frac{x-11}{6}$ D. $\frac{x-4}{6}$
26. Simplify 4/x +1 – 3/x-1
(A. X + 7 /X^2–1 B. X – 1 /X^2+ 1 C. X – 7/X^2–1 D. X – 11 /X^2+ 1)
27. Given that m = 3 and n = 2, find the value of $3n^2 – 3m^3/m$
(A. –22 B. –15 C. 14 D. -31)
28. Simplify log √8/log8
(A. 1/3 B. ½ C. 1/3log√2 D. ½ log √2)
29. The values of three angles at a point are $3y-45^0$; $y+25^0$ and y^0. find the value of y.
(A. 40 B. 58 C. 68 D. 76)
30. Ngozi sold an article for N 1755.00 and made a profit of 35%. Find the cost price.
(A. N 2,370 B. N 1300 C. N 614.25 D. N 614.00)
31. Find the compound interest on #450 in 2yrs at 5%.
(A) # 497 (B) #496.13 (C) #45.00 (D) #46.13
32. The monthly salary of a man increase from #2,700 to #3,200. Find the percentage increase.
(A) 10% (B) 15% (C) 15.6% (D) 18.5%)
33. The sum of the interior angles of a polygon is 1260^0. Find the number of its sides.
(A) 7 (B) 8 (C) 9 (D) 10.)
34. What is the total surface area of a cube of side 4cm?
(A) $36cm^2$ (B) $64cm^2$ (C) $96cm^2$ (D) $144cm^2$)
35. A boy threw two dice at once what is the probability of having a total of six? (A) 1/9 (B) 3/8 (C) 4/35 (D) 2/7)
36. The angles of a triangle are in the ratio 5:3:2 what is the size of the smallest angle?

(A) 90^0 (B) 72^0 (C) 54^0 (D) 36^0)

37. Simplify the fraction ¼ ÷½ / ¾
(A) 1 (B) 2/3 (C) 8/3 (D) 3/8)

38. Convert 1001101 to a number in base ten
(A) 61 (B) 46 (C) 45 (D) 44.)

39. What is the value of x if $(101)_x=172_8$
(A) 5 (B) 4 (C) 11 (D) 2)

40. If x-3 is directly proportional to the square of y and x=5 when y=2, find x when y=6.
(A) 30 (B) 21 (C) 16 (D) 12)

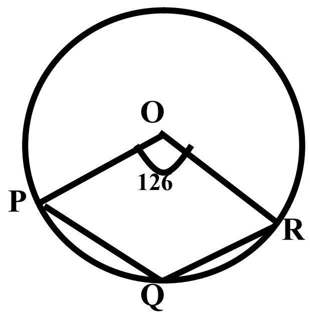
41. Let u- {a,b,c,d}, z=[c,d] and y(a,b,c) what is (zny)′
(A). (a,b,c) (B). (a,b,d) (C). (b,c) (D). (a,d))

42. What must be added to the expression x^2-18x to make it a perfect square?
(A. 3 B. 9 C. 39 D. 72)

43. If x + 3 varies directly as y and x= 3 when y = 12, what is the value of x when y = 8?
A. 1 B. $\frac{1}{2}$ C.- $\frac{1}{2}$ D. -1

44. The area of a parallelogram is 573cm^2 and the height is 19cm. Calculate the base.
(A. 13.5cm . 25cm C. 27cm D. 54cm)

45. In the diagram, O is the center of the circle and
<POR = 126^0. Find <PQR.
(A. 117^0 B. 72^0 C. 63^0 D. 54^0)

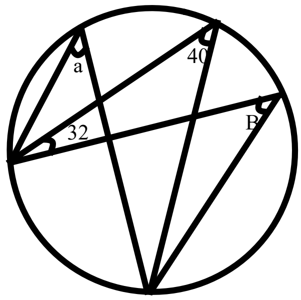


46. If P = 3/5√qr, express q in terms of p and r.
((A). 9/25pr^2 (B). 9/25p^2 r (C). 25/9p^{2}r (D). 25/9pr^2)

47. Find (x – y), if 4x – 3y = 7 and 3x – 2y = 5.
(A. 4 B. 3 C. 2 D. –2)

48. If $512_x - 354_x = 125_x$. Find the number base x
(A. 6 B. 7 C. 8 D. 9)

49. Find the lettered angles in the figure ii below.
0 is the center of the circle.
(A. a=b=32^0 B. a=32^0, b=40^0 C. a=b=40^0 D. a=40^0, b=32^0)



50. Given that 6x – y/x + 2y =2, find the value of x/y
(A. 3/8 B. 5/8 C. 4/5 D. 5/4)

51. Find the quadratic equation whose roots are x = -2 or x=7
(A. X^2 + 2X–7 =0 B. X^2–5X – 14 =0 C. X^2 + 5X + 14 = 0
D. X^2 + 5X – 14=0)

52. The fifth term of an A.P is 24 and the eleventh term is 96, find the first term.
(A. 12 B. 4 C. -12 D. -24)

53. The bearing of a point X from a point Y is 074^0. What is the bearing of Y from X?
(A. 106^0 B. 148^0 C. 164^0 D. 254^0.)

54. Find the 4th term of an A.P whose first term is 2 and the common difference is 0.5.
(A. 0.5 B. 25 C. 3.5 D. 4)

55. Let U = (1,2,3,4), P=(2,3) and Q = (2,4) what is (PnQ)' ?
(A. (1,2,3) B. (1,3,4) C. (2,3) D. (1,4)

56. Three balls are drawn one after the other with replacement from a bag containing 5 red, 9 while and 4 blue identical balls. What is the probability that they are one red, one white and one blue?
(A. 5/102 B. 5/136 C. 5/204 D. 5/162)

57. The angle of elevation of the top of a tower from a point on the horizontal ground, 40m

away from the foot of the tower is 30^0 find the height of the tower.

(A. 20m B. $40\sqrt{3}$m/3 C. $20\sqrt{3}$m D. 60m)

58. A trader bought 100 tubers of yam at 5 for #350.00. She sold them in sets of 4 for #290.00. Find her gain percent.

(A. 3.6% B. 3.5% C. 3.4% D. 2.5%)

59. What is the mode of the numbers 8,10,9,9,10,8,11,8,10,9,8 and 14?

(A. 8 B. 9 C. 10 D. 11)

60. If h(m + n) = m (h + r) find h in terms of m, n and r.

(A. h= mr/2m + n B. h = mr/n-m C. h= m+r/n D. h= mr/n

61. The radius of a Geographical globe is 60cm. Find the length of the parallel of latitude 600N.(A. 66πcm B. 60πcm C. 30πcm D. 15πcm)

62. Find the mean deviation of 2, 4, 5 and 9.

(A. 1 B. 2 C. 5 D. 7)

63. Solve the inequality 2x + 3 < 5x.

(A. π>1 B. x < 3/7 C. x > 3/7 D. x > -1)

64. Express as a single fraction $\frac{x}{x-2} - \frac{x+2}{x+3}$. A. $\frac{2x^2-3x-4}{(x-2)(x+3)}$. B. $\frac{2x^2+3x-4}{(x-2)(x+3)}$. C. $\frac{2}{(x-2)(x+3)}$. D. $\frac{3x+4}{(x-2)(x+3)}$

65. Express 12.625ten in base two. A. 101.1.0. B. 101.110. C. 1100.011. D. 1100.101

66. If X = {all prime factors of 44} and Y = {all prime factors of 60}. The elements of X U Y and XnY. A. {2, 4, 3, 5, 11} and {4}. B. {4, 3, 5, 1} and {3, 4}. C. {2, 5, 11} and {2}. D. {2, 3, 5, 11} and {2}.

67. Factorize $X^2 + 4x - 192$. A. (x – 4) (x + 48). B. (x – 48)(x + 4). C. (x – 12) (x + 16). D. (x – 12)(x – 16)

68. The curved surface of a cylindrical tin is 704cm^2. If the radius of its base is 8cm. Find the height. (Take ∏ = 22/7). A. 14cm. B. 9cm. C. 8cm. D. 7cm

69. Make P the subject of the relation: q = $\frac{3p}{r} + \frac{s}{2}$. A. $\frac{2q-rs}{6}$. B. p = 2qr –sr-3. C. p = $\frac{2qr-s}{6}$. D. p = $\frac{2qr-rs}{6}$

70. If N2, 500 amounted to N3, 500 in 4 years at simple interest, find the rate at which the interest was charged. A. 35%. B. 7 ½ %. C. 8%. D. 10%

71. If P α $\frac{1}{q}$ which of the following is true? A. q α p^2 . B. q α $\frac{1}{P^2}$. C. q α $\sqrt{P}$. D. q α$\frac{1}{p}$

72. If y = 23_{five} + 101_{three} find y leaving your answer in base two. A. 1110. B. 10111. C. 11101. D. 111100

73. Evaluate $2021_{three} - 1122_{three}$. A. 122. B. 112. C. 211. D. 210.

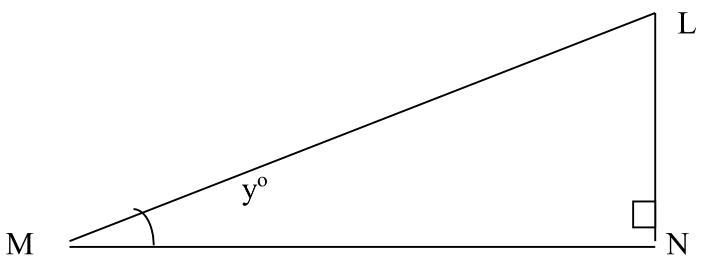
74. A machine valued at N20, 000 depreciates by 10% every year, what will be the value

of the machine at the end of two years? A. N16, 200. B. N8000. C. N14, 200. D. N12, 000

75. For what value of x is the expression $\frac{2x-1}{x+3}$ not defined? A. 3. B. 2. C. ½. D. -3

76. Given that (2x + 7) is a factor of $2x^2 + 3x - 14$, find the other factor.
A. x+2 B. 2 – x C. x – 2 D. x + 1

77. In the diagram below |LN| = 4cm $L\hat{N}M = 90^o$ and tan y = 2/3. What is the area of the LMN? A. $24cm^2$. B. $12cm^2$. C. $10cm^2$. D. $6cm^2$.



78. Find the volume of a solid cylinder with base radius 10cm and length 14cm (Take ∏ = 22/7). A. $220cm^3$. B. $880cm^3$. C. $1400cm^3$. D. $440cm^3$

79. If sin X = 12/13 where $0^o < x < 90^o$, find the value of $1-\cos^2 x$. A. 25/169. B. 64/169. C. 105/169. D. 144/169

80. Given that tan x = 5/12, what is the value of sin X + Cos X?. a. 5/13. B. 7/13. C. 12/13. D. 17/13.

81. A house bought for N100, 000 was later auctioned for N80,000. Find the loss percent. A. 20%. B. 30%. C. 40%. D. 60%

82. If N varies directly as M and N = 8 when M = 20. Find M when N = 7. A. 13. B. 15. C. 7 ½. D. 18 ½.

83. A solid cylinder of radius 7cm is 10cm long. Find its total surface area. A. $70∏cm^2$. B. $189∏cm^2$. C. $210∏cm^2$. D. $238∏cm^2$

84. The lengths of the parallel sides of a trapezium m are 9cm and 12cm. If the area of the trapezium is $105cm^2$, find the perpendicular distance between the parallel sides. A. 5cm. B. 7cm. C. 10cm. D. 15cm

85. The base diameter of a cone is 14cm and its volume is 462 cm3. Find its height. Take ∏=22/7 A. 0.75cm B. 2.25cm C. 0.25cm D. 2.05cm

86. A student bought 3 notebooks and 1 pen for N35. After misplacing these items, she again bought 2 note books and 2 pens all of the same type for N30. What is the cost of a pen? A. N5. B. N7.50. c. N10. D. N15.

87. The sides of two cubes are in the ratio 2:5 what is the ratio of their volume? A. 4:5. B. 8:15. C. 6:125. D. 8:125
88. Ladi sold a car for N84, 000 at a loss of 4%. How much did Ladi buy the car? A. N80,500. B. N80, 640. C. N87, 360. D. N87, 500
89. A rectangular packet has inner dimension 16cm by 12cm by 6cm. How many cubes of sugar of sides 2cm can be neatly packed into the packet? A. 90. B. 144. C. 150. D. 288
90. A baker used 40% of a 50kg bag of flour. If 1/8 of the amount was used for cake, how many kilograms of flour was used for cake? A. 2 ½. B. 6 ¼. C.15 $^{5}/_{8}$. D. 17 ½.
91. If 30% of y is equal to x, what in terms of x is 30% of 3y? A. x/9. B. x/3. C. x. D. 3x.
92. What is the difference in longitude between P(lat. 50^0N, long. 50^0W) and Q(Lat. 50^0N, long 150^0W)? A. 300^0. B. 200^0. C. 130^0. D. 100^0.
93. If Cos 60° = ½, which of the following angles has a cosine of ½? A. 30°. B. 120^0. C. 150^0. D. 140^0.
94. Find the roots of the equation $2x^2 - 3x - 2 = 0$. A. X= -2 or 1 ½. B. X = -2 or 1. C. X = 1 or 2. D. X = - ½ or 2.
95. Evaluate $\log_{10}6 + \log_{10}45 - \log_{10}27$ without using logarithm tables. A. 0. B. 1. C. 1.3802. D. 10
96. Find the sum of the first five terms of the G.P. 2, 6, 18. A. 484. B. 243. C. 242. D. 130.
97. In a class of 80 students, every student had to study Economics or Geography, or both Economics and Geography. If 65 students studied Economics and 50 studied Geography, how many studied both subjects? A 15. B. 30. C. 35. D. 45.

Use the following information to answer questions 98 and 99. (Take $\pi = 22/7$).

A cylindrical container, closed at both ends, has a radius of 7cm and height 5cm

98. Find the total surface area of the container. A. $35cm^2$. B. $220cm^2$. C. $528cm^2$. D. $770cm^2$
99. What is the volume of the container? A. $35cm^2$. B. $220cm^2$. C. $528cm^2$. D. $770cm^2$
100. A section of a circle of radius 7cm has an area of $44cm^2$. Calculate the angle of the sector, correct to the nearest degree. A. 6°. B. 26°. C. 52°. D. 103°.

SECTION C

QUESTIONNAIRE FOR SOCIAL ECONOMIC STATUS (QSES)

Attempt all questions in this section. Kindly thick your response

1. Your parents have □ not more than 4 children, □ 5-6 children, □ more than 6

2. You are your father's □ 1st or 2nd child, □ 3rd or 4th child, □ 5th, 6th or --….. child
3. My family lives in □ a whole house, □ a two room apartment or less, □ a flat of 2/3 bedroom
4. At home you □ own a room to yourself □ share a room with one or two other, □ share a room with more than 2 other
5. Apart from your parents and their children, how many other lives in your home □ Nobody, □ one or two other, □ 3 or more other
6. At home you speak English □ At all time, □ sometimes, □ never or rarely
7. In your home there is □ Plasma TV and home theater, □ a TV and Radio set, □ a radio set or none of these things
8. Your parents own □ more than one car, □ a car, □ do not own a car
9. You attended □ an expensive fee paying private school, □ a fee paying private primary school, □ a free UBE primary school
10. Your parents buy you books to read □ often, □ sometime, □ rarely/never
11. Apart from school textbooks □ more than 50 books are in my home, □ about 10 books are in my home, □ about 30 books are in my home,
12. At home you speak your native language □ all the time, □ sometimes, □ rarely/never
13. At home your parents buy □ more than 1 daily newspaper □ no daily newspaper, □ 1 daily newspaper
14. Do your parents encourage you to speak with them □ often,□ sometimes, □ rarely/never
15. When you speak at home, do your parents or guardian insist that you speak in English? □ often, □ sometime, □ rarely/never
16. In your free time, do your parents □ encourage you to read as much as possible, □ sometime ask you to read, □ never mind if you ever read
17. Does your father /mother help you with your home work? □ often, □ sometime, □ rarely/never
18. Your father/guidance attended □ no school at all, □ primary/ secondary school, □ college/polytechnic/university
19. Your mother/guidance attended □ no school at all, □ primary/ secondary school, □ college/polytechnic/university

20. Your father/guidance pay your school fees and buy school books promptly?
 every time, sometime, rarely/never
21. In your father's house you have deep well and pumping machine sure, barely
 not sure/no
22. Your parents own (have) another house besides the one you are living. sure,
 barely not sure/no
23. You have graduates brother or sister in your family sure, barely, not sure/no
24. In your home, there are children in higher institutions, children in secondary schools children in primary schools two or more of the school types
25. At home you have an indoor game sure, barely, not sure/no

APPENDIX VII

MATHEMATICS ACHIEVEMENT TEST (MAT) ANSWER SHEET

EXAMINEES PERSONAL DATA ---

SEX: MALE: [ ] FEMALE: [ ] HSES: [ ] LSES: [ ]

SCHOOL LOCATION: URBAN: [ ] RURAL: [ ]

SCHOOL TYPE: PUBLIC: [ ] PRIVATE: [ ]

1		21		41		61		81	
2		22		42		62		82	
3		23		43		63		83	
4		24		44		64		84	
5		25		45		65		85	
6		26		46		66		86	
7		27		47		67		87	
8		28		48		68		88	
9		29		49		69		89	
10		30		50		70		90	
11		31		51		71		91	
12		32		52		72		92	
13		33		53		73		93	
14		34		54		74		94	
15		35		55		75		95	
16		36		56		76		96	
17		37		57		77		97	
18		38		58		78		98	
19		39		59		79		99	
20		40		60		80		100	

APPENDIX VIII

ANSWER SHEET FOR QUESTIONNAIRE FOR SOCIAL ECONOMIC STATUS (QSES)

EXAMINEES PERSONAL DATA--

SEX: MALE: [ ] FEMALE: [ ] SES: H [ ] L [ ]

SCHOOL LOCATION: URBAN: [ ] RURAL: [ ]

SCHOOL TYPE: PUBLIC: [ ] PRIVATE: [ ]

1		6		11		16		21	
2		7		12		17		22	
3		8		13		18		23	
4		9		14		19		24	
5		10		15		20		25	

APPENDIX 1X
MATHEMATICS ACHIEVEMENT TEST (MAT)
PROVISIONAL ANSWERS

1	D	21	B	41	C	61	A	81	A
2	A	22	A	42	C	62	D	82	C
3	D	23	B	43	D	63	D	83	D
4	A	24	C	44	A	64	D	84	C
5	D	25	C	45	C	65	C	85	B
6	B	26	A	46	C	66	A	86	A
7	B	27	D	47	C	67	D	87	D
8	D	28	D	48	D	68	D	88	B
9	D	29	B	49	B	69	D	89	A
10	A	30	D	50	D	70	B	90	D
11	D	31	D	51	D	71	B	91	B
12	C	32	C	52	C	72	D	92	C
13	B	33	C	53	B	73	A	93	D
14	A	34	B	54	D	74	A	94	D
15	D	35	D	55	B	75	D	95	D
16	A	36	D	56	A	76	C	96	D
17	C	37	D	57	A	77	B	97	A
18	A	38	C	58	D	78	D	98	D
19	D	39	B	59	B	79	D	99	C
20	C	40	B	60	B	80	D	100	A

APPENDIX X

Rasch Dichotomous Model vs. One-parameter Logistic Model (1PL 1-PL)

For most practical purposes these models are the same, despite their conceptual differences.

Aspect	Rasch Dichotomous Model	Item Response Theory: One-Parameter Logistic Model
Abbreviation	Rasch	1-PL IRT, also 1PL
For practical purposes	When each individual in the person sample is parameterized for item estimation, it is Rasch.	When the person sample is parameterized by a mean and standard deviation for item estimation, it is 1PL IRT.
Motivation	Prescriptive: Distribution-free person ability estimates and distribution-free item difficulty estimates on an additive latent variable	Descriptive: Computationally simpler approximation to the Normal Ogive Model of L.L. Thurstone, D.N. Lawley, F.M. Lord
Persons, objects, subjects, cases, etc.	Person n of ability B_n, or Person ν (Greek nu) of ability β_n in logits	Normally-distributed person sample of ability distribution θ, conceptualized as N(0,1), in probits: incidental parameters
Items, agents, prompts, probes, multiple-choice questions, etc.: structural parameters	Item i of difficulty D_i, or Item ι (Greek iota) of difficulty δ_i in logits	Itemi of difficulty b_i (the "one parameter") in probits
Nature of binary data	1 = "success" – presence of property 0 = "failure" - absence of property	1 = "success" - presence of property 0 = "failure" - absence of property
Probability of binary data	P_{ni} = probability that person n is observed to have the requisite property, "succeeds", when encountering item i	$P_i(\theta)$ = overall probability of "success" by person distribution θ on item i
Formulation: exponential form e = 2.71828	$P_{ni} = \frac{e^{B_n - D_i}}{1 + e^{B_n - D_i}}$	$P_i(\theta) = \frac{e^{1.7(\theta - b_i)}}{1 + e^{1.7(\theta - b_i)}}$

Formulation: logit-linear form $\log_e$ = natural logarithm	$\log_e\left(\frac{P_{ni}}{1-P_{ni}}\right)=B_n-D_i$	$\log_e\left(\frac{P_i(\theta)}{1-P_i(\theta)}\right)=1.7(\theta-b_i)$
Local origin of scale: zero of parameter estimates	Average item difficulty, or difficulty of specified item. (Criterion-referenced)	Average person ability. (Norm-referenced)
Item discrimination	Item characteristic curves (ICCs) modeled to be parallel with a slope of 1 (the natural logistic ogive)	ICCs modeled to be parallel with a slope of 1.7 (approximating the slope of the cumulative normal ogive)
Missing data allowed	Yes, depending on estimation method	Yes, depending on estimation method
Fixed (anchored) parameter values for persons and items	Yes, depending on software	Items: depending on software. Persons: only for distributional form.
Fit evaluation	Fit of the data to the model Local, one parameter at a time	Fit of the model to the data Global, accept or reject the model
Data-model mismatch	Defective data do not support parameter separability in an additive framework. Consider editing the data.	Defective model does not adequately describe the data. Consider adding discrimination (2-PL), lower asymptote (guessability, 3-PL) parameters.
Differential item functioning (DIF) detection	Yes, in secondary analysis	Yes, in secondary analysis
First conspicuous appearance	Rasch, Georg. (1960) Probabilistic models for some intelligence and attainment tests. Copenhagen: Danish Institute for Educational Research.	Birnbaum, Allan. (1968). Some latent trait models. In F.M. Lord & M.R. Novick, (Eds.), Statistical theories of mental test scores. Reading, MA: Addison-Wesley.
First conspicuous advocate	Benjamin D. Wright, University of Chicago	Frederic M. Lord, Educational Testing Service
Widely-	David Andrich,Univ. of	Ronald Hambleton, University of

authoritative currently-active proponent	Western Australia, Perth, Australia	Massachusetts
Introductory textbook	Applying The Rasch Model.T.G. Bond and C.M. Fox	Fundamentals of Item Response Theory.R.K. Hambleton, H. Swaminathan, and H.J. Rogers.
Widely used software	Winsteps, RUMM, ConQuest	Logist, BILOG
Minimum reasonable sample size	30	200 (Downing 2003)

Linacre J.M. (2005). Rasch dichotomous model vs. One-parameter Logistic Model. Rasch Measurement Transactions, 19:3, 1032

APPENDIX X1

Letter to Principles of Senior Secondary Schools in Some Selected Schools in Oyo and Delta States

Department of Guidance and
Counselling,
(Measurement and Evaluation Unit)
Faculty of Education,
Delta State University,
Abraka.

Dear Sir/Madam,

I am presently conducting a research on the Development and Validation of an Achievement Test in Mathematics for Senior Secondary three students in Oyo and Delta states and your schools fall within the selected Local Government Areas in your state.

I therefore solicit for your support for administering the developed test items to your students. The administration of the test will involve a total of thirty students as the case may be. The developed test will cover all topics treated from SS1 to SS3 based on WAEC and NECO syllabi. Names of the students and school will not be used for the final analysis of the data

I am looking forward to your cooperation

Yours faithfully,

ALIYU, R. TAIWO.

APPENDIX XII
THE SIXTY-FIVE CALIBRATED MAT ITEM FOR ITEM BANKING

Attempt all the questions.

1. If (x - 3) is a factor of $2x^3 + 3x^2 - 17x - 30$, find the remaining factors
   A. (2x – 5)(x – 2)
   B. (2x – 5)(x + 4)
   C. (2x + 5)(x – 2)
   D. (2x + 5)(x + 2)
2. Simplify $\frac{{}^nP_4}{{}^nC_4}$
   A. 24
   B. 18
   C. 12
   D. 6
3. Evaluate $\log_{0.25} 8$
   A. $\frac{3}{2}$
   B. $\frac{2}{3}$
   C. $-\frac{2}{3}$
   D. $-\frac{3}{2}$
4. If $y = x^2 - \frac{1}{x}$, find $\frac{dy}{dx}$
   A. $x^2 - \frac{1}{x^2}$
   B. $2x + \frac{1}{x^2}$
   C. $2x - x^2$
   D. $x^2 + \frac{1}{x}$
5. A man stands on a tree 150cm high and sees a boat at an angle of depression of 74^0. Find the distance of the boat from the base of the tree
   A. 52cm
   B. 43cm
   C. 40cm
   D. 15cm
6. Find the remainder when $2x^3 - 11x^2 + 8x - 1$ is divided by x + 3
   A. -871
   B. -781
   C. -187
   D. -178
7. The probabilities that a man and his wife love for 80years are $\frac{2}{3}$ and $\frac{3}{15}$ respectively. Find the probability that at least one of them will live up to 80 years
   A. $\frac{2}{15}$
   B. $\frac{3}{15}$
   C. $\frac{7}{15}$

D. $\frac{13}{15}$

8. The initial velocity of an object is u = $\binom{-5}{3} m/s$. If the acceleration of the object is a = $\binom{3}{4} m/s^2$ and it moved for 3 seconds. Find the final velocity

A. $\binom{-14}{15} m/s$

B. $\binom{-2}{1} m/s$

C. $\binom{4}{-9} m/s$

D. $\binom{-14}{-9} m/s$

9. The sum and product of the roots of a quadratic equation are $\frac{4}{7}$ and $\frac{5}{7}$ respectively. Find its equation

A. $7x^2 - 4x - 5 = 0$
B. $7x^2 - 4x - 5 = 0$
C. $7x^2 + 4x - 5 = 0$
D. $7x^2 - 4x + 5 = 0$

10. Simplify $(\sqrt{6} + 2)^2 - (\sqrt{6} - 2)^2$

A. $2\sqrt{6}$

B. $4\sqrt{6}$

C. $8\sqrt{6}$

D. $16\sqrt{6}$

11. The nth term of a sequence is $n^2 - 6n - 4$. Find the sum of the 3rd and 4th terms
A. 24
B. 23
C. -24
D. -25

12. Two bodies of mass 8kg and 5kg traveling in the same direction with speed x m/s and 2 m/s respectively collide. If after collision, they move together with a speed of 3.85m/s, find correct to the nearest whole number, the value of x
A. 2
B. 5
C. 8
D. 13

13. Simplify log √8/log8
A. 1/3
B. ½
C. 1/3log√2
D. ½ log √2

14. What must be added to the expression x^2-18x to make it a perfect square?
A. 3
B. 9
C. 39
D. 72

15. The area of a parallelogram is 573cm^2 and the height is 19cm. Calculate the base.
A. 13.5cm
B. 25cm
C. 27cm
D. 54cm

16. If $512_x - 354_x = 125_x$. Find the number base x
    A. 6
    B. 7
    C. 8
    D. 9
17. Find the quadratic equation whose roots are x = -2 or x=7
    A. $x^2 + 2x - 7 = 0$
    B. $x^2 - 5x - 14 = 0$
    C. $x^2 + 5x + 14 = 0$
    D. $x^2 + 5x - 14 = 0$
18. The fifth term of an A.P is 24 and the eleventh term is 96, find the first term.
    A. 12
    B. 4
    C. -12
    D. -24
19. The bearing of a point X from a point Y is 074^0. What is the bearing of Y from X?
    A. 106^0
    B. 148^0
    C. 164^0
    D. 254^0.
20. Find the 4th term of an A.P whose first term is 2 and the common difference is 0.5.
    A. 0.5
    B. 25
    C. 3.5
    D. 4
21. Let U = (1,2,3,4), P=(2,3) and Q = (2,4) what is $(PnQ)^1$?
    A. (1,2,3)
    B. (1,3,4)
    C. (2,3)
    D. (1,4)
22. Three balls are drawn one after the other with replacement from a bag containing 5 red, 9 while and 4 blue identical balls. What is the probability that they are one red, one white and one blue?
    A. 5/102
    B. 5/136
    C. 5/204
    D. 5/162
23. The angle of elevation of the top of a tower from a point on the horizontal ground, 40m away from the foot of the tower is 30^0 find the height of the tower.
    A. 20m
    B. $40\sqrt{3}$m/3
    C. $20\sqrt{3}$m
    D. 60m
24. A trader bought 100 tubers of yam at 5 for #350.00. She sold them in sets of 4 for #290.00. Find her gain percent.
    A. 3.6%
    B. 3.5%
    C. 3.4%
    D. 2.5%

25. What is the mode of the numbers 8,10,9,9,10,8,11,8,10,9,8 and 14?
   A. 8
   B. 9
   C. 10
   D. 11
26. If h(m + n) = m (h + r) find h in terms of m, n and r.
   A. h= mr/2m + n
   B. h = mr/n-m
   C. h= m+r/n
   D. h= mr/n
27. The radius of a Geographical globe is 60cm. Find the length of the parallel of latitude 600N.
   A. 66πcm
   B. 60πcm
   C. 30πcm
   D. 15πcm
28. Find the mean deviation of 2, 4, 5 and 9.
   A. 1
   B. 2
   C. 5
   D. 7

29. Solve the inequality 2x + 3 < 5x.
   A. π>1
   B. x < 3/7
   C. x > 3/7
   D. x > -1

30. Express as a single fraction $\frac{x}{x-2} - \frac{x+2}{x+3}$. A. $\frac{2x^2-3x-4}{(x-2)(x+3)}$. B. $\frac{2x^2+3x-4}{(x-2)(x+3)}$. C. $\frac{2}{(x-2)(x+3)}$. D. $\frac{3x+4}{(x-2)(x+3)}$

31. Express 12.625ten in base two.

   A. 101.1.0.

   B. 101.110.

   C. 1100.011.

   D. 1100.101

32. If X = {all prime factors of 44} and Y = {all prime factors of 60}. The elements of X U Y and XnY.

   A. {2, 4, 3, 5, 11} and {4}.

   B. {4, 3, 5, 1} and {3, 4}.

   C. {2, 5, 11} and {2}.

   D. {2, 3, 5, 11} and {2}.

33. Factorize $X^2 + 4x - 192$.

A. $(x - 4)(x + 48)$.

B. $(x - 48)(x + 4)$.

C. $(x - 12)(x + 16)$.

D. $(x - 12)(x - 16)$

34. The curved surface of a cylindrical tin is $704cm^2$. If the radius of its base is 8cm. Find the height. (Take $\prod = 22/7$).

A. 14cm.

B. 9cm.

C. 8cm.

D. 7cm

35. Make P the subject of the relation: $q = \frac{3p}{r} + \frac{s}{2}$. A. $\frac{2q - rs}{6}$. B. $p = 2qr - sr-3$. C. $p = \frac{2qr - s}{6}$. D. $p = \frac{2qr - rs}{6}$

36. If N2, 500 amounted to N3, 500 in 4 years at simple interest, find the rate at which the interest was charged.

A. 35%.

B. 7 ½ %.

C. 8%.

D. 10%

37. If $P \alpha \frac{1}{q}$ which of the following is true?

A. $q \alpha p^2$.

B. $q \alpha \frac{1}{p^2}$.

C. $q \alpha \sqrt{P}$.

D. $q \alpha \frac{1}{p}$

38. If $y = 23_{five} + 101_{three}$ find y leaving your answer in base two.

A. 1110.

B. 10111.

C. 11101.

D. 111100

39. Evaluate $2021_{three} - 1122_{three}$.

A. 122.

B. 112.

C. 211.

D. 210.

40. A machine valued at N20, 000 depreciates by 10% every year, what will be the value of the machine at the end of two years?

A. N16, 200.

B. N8000.

C. N14, 200.

D. N12, 000

41. For what value of x is the expression $\frac{2x-1}{x+3}$ not defined?

A. 3.

B. 2.

C. ½.

D. -3

42. Given that (2x + 7) is a factor of $2x^2 + 3x - 14$, find the other factor.

A. x+2

B. 2 – x

C. x – 2

D. x + 1

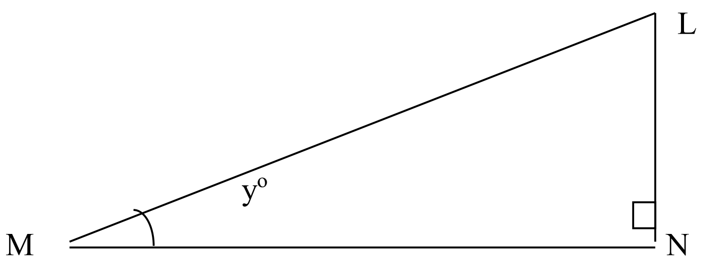
43. In the diagram below |LN| = 4cm $L\hat{N}M = 90^o$ and tan y = 2/3. What is the area of the LMN?

A. $24cm^2$.

B. $12cm^2$.

C. $10cm^2$.

D. $6cm^2$.



44. Find the volume of a solid cylinder with base radius 10cm and length 14cm (Take ∏ = 22/7).

A. $220cm^3$.

B. $880cm^3$.

C. $1400cm^3$.

D. $440cm^3$

45. If sin X = 12/13 where $0^o < x < 90^o$, find the value of $1\text{-}cos^2x$.

A. 25/169.

B. 64/169.

C. 105/169.

D. 144/169

46. Given that tan x = 5/12, what is the value of sin X + Cos X?.

A. 5/13.

B. 7/13.

C. 12/13.

D. 17/13.

47. A house bought for N100, 000 was later auctioned for N80,000. Find the loss percent.

A. 20%.

B. 30%.

C. 40%.

D. 60%

48. If N varies directly as M and N = 8 when M = 20. Find M when N = 7.

A. 13.

B. 15.

C. 7 ½.

D. 18 ½.

49. A solid cylinder of radius 7cm is 10cm long. Find its total surface area.

A. $70∏cm^2$.

B. $189∏cm^2$.

C. $210∏cm^2$.

D. $238∏cm^2$

50. The lengths of the parallel sides of a trapezium m are 9cm and 12cm. If the area of the trapezium is $105cm^2$, find the perpendicular distance between the parallel sides.

A. 5cm.

B. 7cm.

C. 10cm.

D. 15cm

51. The base diameter of a cone is 14cm and its volume is 462 cm^3. Find its height. Take ∏=22/7

A. 0.75cm

B. 2.25cm

C. 0.25cm

D. 2.05cm

52. A student bought 3 notebooks and 1 pen for N35. After misplacing these items, she again bought 2 note books and 2 pens all of the same type for N30. What is the cost of a pen?

A. N5.

B. N7.50.

C. N10.

D. N15.

53. The sides of two cubes are in the ratio 2:5 what is the ratio of their volume?

A. 4:5.

B. 8:15.

C. 6:125.

D. 8:125

54. Ladi sold a car for N84, 000 at a loss of 4%. How much did Ladi buy the car?

A. N80,500.

B. N80, 640.

C. N87, 360.

D. N87, 500

55. A rectangular packet has inner dimension 16cm by 12cm by 6cm. How many cubes of sugar of sides 2cm can be neatly packed into the packet?

A. 90.

B. 144.

C. 150.

D. 288

56. A baker used 40% of a 50kg bag of flour. If 1/8 of the amount was used for cake, how many kilograms of flour was used for cake?

A. 2 ½.

B. 6 ¼.

C.15 $^5/_8$.

D. 17 ½.

57. If 30% of y is equal to x, what in terms of x is 30% of 3y? a. x/9. B. x/3. C. x. D. 3x.
58. What is the difference in longitude between P(lat. 50^0N, long. 50^0W) and Q(Lat. 50^0N, long 150^0W)? A. 300^0. B. 200^0. C. 130^0. D. 100^0.
59. If Cos 60° = ½, which of the following angles has a cosine of ½?

A. 30°.

B. 120^0.

C. 150^0.

D. 140^0.

60. Find the roots of the equation $2x^2 – 3x – 2 = 0$.

A. X= -2 or 1 ½.

B. X = -2 or 1.

C. X = 1 or 2.

D. X = - ½ or 2.

61. Evaluate $\log_{10}6 + \log_{10}45 – \log_{10}27$ without using logarithm tables.

A. 0.

B. 1.

C. 1.3802.

D. 10

62. Find the sum of the first five terms of the G.P. 2, 6, 18.

A. 484.

B. 243.

C. 242.

D. 130.

63. In a class of 80 students, every student had to study Economics or Geography, or both Economics and Geography. If 65 students studied Economics and 50 studied Geography, how many studied both subjects?

A. 15.

B. 30.

C. 35.

D. 45.

Use the following information to answer questions 98 and 99. (Take $\pi = 22/7$).

A cylindrical container, closed at both ends, has a radius of 7cm and height 5cm

64. Find the total surface area of the container.

A. $35cm^2$.

B. $220cm^2$.

C. $528cm^2$.

D. $770cm^2$

65. What is the volume of the container?

A. $35cm^3$.

B. $220cm^3$.

C. $528cm^3$.

D. $770cm^3$.

APPENDIX XIII

SIXTY-FIVE CALIBRATED MATHEMATICS ACHIEVEMENT TEST (MAT) ITEMS IN LOGIT UNITS PRIOR TO AND AFTER REMOVAL OF MISFIT RESPONSES

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	EXACT MATCH MAT TOPICS	Item
1	746	1499	.52	.05	1	746	1499	.83	.05	ALGEBRAIC	I0001
2	1002	1499	-.20	.06	2	1002	1499	.10	.06	ALGEBRAIC	I0002
3	555	1499	1.05	.05	3	555	1499	1.37	.05	NUMERATION	I0003
5	738	1499	.54	.05	4	738	1499	.85	.05	ALGEBRAIC	I0004
8	579	1499	.98	.05	5	427	1499	1.77	.06	TRIGNOMETRY	I0005
9	674	1499	.72	.05	6	579	1499	1.30	.05	ALGEBRAIC	I0006
10	785	1499	.41	.05	7	785	1499	.72	.05	PROBABILITY	I0007
11	986	1499	-.16	.05	8	986	1499	.15	.06	ALGEBRAE	I0008
12	702	1499	.64	.05	9	702	1499	.95	.05	ALGEBRAE	I0009
13	804	1499	.36	.05	10	804	1499	.67	.05	TRIGONOMETR	I0010
14	698	1499	.65	.05	11	698	1499	.96	.05	TRIGONOMETR	I0011
22	566	1499	1.02	.05	12	566	1499	1.34	.05	ALGEBRAIC	I0012
28	578	1499	.99	.05	13	578	1499	1.30	.05	NUMERATION	I0013
42	426	1499	1.45	.06	14	1217	1499	-.68	.07	ALGEBRAIC	I0014
44	792	1499	.39	.05	15	426	1499	1.77	.06	MENSURATION	I0015
48	958	1499	-.07	.05	16	792	1499	.71	.05	NUMERATION	I0016
51	1377	1499	-1.95	.09	17	958	1499	.23	.05	ALGEBRAE	I0017
52	1351	1499	-1.73	.09	18	1377	1499	-1.66	.09	TRIGONOMETR	I0018
53	1319	1499	-1.51	.08	19	1351	1499	-1.44	.09	TRIGONOMETR	I0019
54	1280	1499	-1.28	.07	20	1319	1499	-1.22	.08	TRIGONOMETR	I0020
55	1200	1499	-.90	.07	21	1280	1499	-.99	.07	NUMERATION	I0021
56	1143	1499	-.68	.06	22	1200	1499	-.61	.07	PROBABILITY	I0022
57	1095	1499	-.51	.06	23	1143	1499	-.38	.06	TRIGONOMETR	I0023
58	1072	1499	-.43	.06	24	1095	1499	-.21	.06	NUMERATION	I0024
59	1055	1499	-.37	.06	25	1072	1499	-.13	.06	STATISTICS	I0025
60	1068	1499	-.41	.06	26	1055	1499	-.07	.06	ALGEBRAE	I0026
61	1082	1499	-.46	.06	27	1068	1499	-.11	.06	GEOMETRY	I0027
62	1094	1499	-.50	.06	28	1094	1499	-.20	.06	STATISTICS	I0028
63	1103	1499	-.53	.06	29	1103	1499	-.23	.06	ALGEBRAE	I0029
64	1163	1499	-.75	.06	30	1163	1499	-.46	.06	ALGEBRAE	I0030
65	1150	1499	-.70	.06	31	1150	1499	-.41	.06	NUMERATION	I0031
66	1205	1499	-.92	.07	32	1205	1499	-.63	.07	NUMERATION	I0032
67	1160	1499	-.74	.06	33	1160	1499	-.44	.06	ALGEBRAE	I0033
68	1135	1499	-.65	.06	34	1135	1499	-.35	.06	MENSURATION	I0034
69	1131	1499	-.63	.06	35	1131	1499	-.33	.06	ALGEBRAE	I0035
70	1122	1499	-.60	.06	36	1122	1499	-.30	.06	NUMERATION	I0036
71	1165	1499	-.76	.06	37	1165	1499	-.46	.06	TRIGONOMETR	I0037
72	1145	1499	-.68	.06	38	1145	1499	-.39	.06	NUMERATION	I0038
73	1057	1499	-.38	.06	39	1057	1499	-.08	.06	NUMERATION	I0039
74	1040	1499	-.32	.06	40	1040	1499	-.02	.06	NUMERATION	I0040
75	1081	1499	-.46	.06	41	1081	1499	-.16	.06	ALGEBRAE	I0041
76	1027	1499	-.28	.06	42	1027	1499	.02	.06	ALGEBRAE	I0042
77	1099	1499	-.52	.06	43	1099	1499	-.22	.06	GEOMETRY	I0043
78	1090	1499	-.49	.06	44	1090	1499	-.19	.06	MENSURATION	I0044
79	1002	1499	-.20	.06	45	1002	1499	.10	.06	TRIGONOMETR	I0045
80	995	1499	-.18	.06	46	995	1499	.12	.06	TRIGONOMETR	I0046
81	1043	1499	-.33	.06	47	1043	1499	-.03	.06	NUMERATION	I0047
82	1081	1499	-.46	.06	48	1081	1499	-.16	.06	ALGEBRAE	I0048
83	1048	1499	-.35	.06	49	1048	1499	-.05	.06	MENSURATION	I0049
84	1116	1499	-.58	.06	50	1116	1499	-.28	.06	TRIGONOMETR	I0050
85	1090	1499	-.49	.06	51	1090	1499	-.19	.06	MENSURATION	I0051
86	1070	1499	-.42	.06	52	1070	1499	-.12	.06	NUMERATION	I0052
87	1073	1499	-.43	.06	53	1073	1499	-.13	.06	MENSURATIO	I0053
88	1049	1499	-.35	.06	54	1049	1499	-.05	.06	NUMERATION	I0054
89	1137	1499	-.66	.06	55	1137	1499	-.36	.06	MENSURATION	I0055
90	1105	1499	-.54	.06	56	1105	1499	-.24	.06	NUMERATION	I0056
91	1043	1499	-.33	.06	57	1043	1499	-.03	.06	ALGEBRAE	I0057
92	1056	1499	-.37	.06	58	1056	1499	-.07	.06	MENSURATION	I0058
93	1033	1499	-.30	.06	59	1033	1499	.00	.06	TRIGONOMETR	I0059
94	1042	1499	-.33	.06	60	1042	1499	-.03	.06	ALGEBRAE	I0060
95	1065	1499	-.40	.06	61	1065	1499	-.10	.06	NUMERATION	I0061
96	1133	1499	-.64	.06	62	1133	1499	-.34	.06	TRIGONOMETR	I0062
97	1110	1499	-.56	.06	63	1110	1499	-.26	.06	TRIGONOMETR	I0063
98	1053	1499	-.37	.06	64	1053	1499	-.06	.06	MENSURATION	I0064
99	1155	1499	-.72	.06	65	1155	1499	-.42	.06	MENSURATION	I0065
MEAN	915.1	1499.0	.00	.06	MEAN	1015.2	1499.0	.00	.06	71.1 71.4	
S.D.	233.7	.0	.72	.01	S.D.	209.1	.0	.67	.01	8.1 7.5	

APPENDIX XIV

ITEMS LOGIT PRIOR TO REMOVAL OF MISFIT RESPONSES

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH MAT TOPICS	Item
1	746	1499	.52	.05	.99	-.9	.99	-1.1	.17	.14	ALGEBRAIC	I0001
2	1002	1499	-.20	.06	1.03	1.8	1.04	1.8	.01	.13	ALGEBRAIC	I0002
3	555	1499	1.05	.05	.98	-1.2	.98	-1.2	.20	.14	NUMERATION	I0003
5	738	1499	.54	.05	.99	-1.1	.99	-1.0	.17	.14	ALGEBRAIC	I0004
8	579	1499	.98	.05	1.00	.3	1.01	.4	.12	.14	TRIGNOMETRY	I0005
9	674	1499	.72	.05	.99	-1.0	.99	-1.0	.17	.14	ALGEBRAIC	I0006
10	785	1499	.41	.05	1.01	1.2	1.01	1.0	.11	.14	PROBABILITY	I0007
11	986	1499	-.16	.05	.98	-1.3	.98	-1.2	.21	.13	ALGEBRAE	I0008
12	702	1499	.64	.05	.99	-1.7	.99	-1.6	.19	.14	ALGEBRAE	I0009
13	804	1499	.36	.05	1.00	-.4	1.00	-.5	.15	.14	TRIGONOMETRY	I0010
14	698	1499	.65	.05	1.00	.5	1.00	.6	.13	.14	TRIGONOMETRY	I0011
22	566	1499	1.02	.05	1.00	.1	1.00	.2	.13	.14	ALGEBRAIC	I0012
28	578	1499	.99	.05	1.03	2.0	1.03	2.0	.04	.14	NUMERATION	I0013
42	426	1499	1.45	.06	.98	-.9	.98	-.9	.21	.13	ALGEBRAIC	I0014
44	792	1499	.39	.05	.99	-1.6	.99	-1.6	.19	.14	MENSURATION	I0015
48	958	1499	-.07	.05	.99	-.3	.99	-.4	.16	.14	NUMERATION	I0016
51	1377	1499	-1.95	.09	.99	-.1	.95	-.5	.12	.08	ALGEBRAE	I0017
52	1351	1499	-1.73	.09	1.00	.0	1.00	.0	.09	.08	TRIGONOMETRY	I0018
53	1319	1499	-1.51	.08	.99	-.2	.96	-.7	.15	.09	TRIGONOMETRY	I0019
54	1280	1499	-1.28	.07	.98	-.3	.96	-.8	.18	.10	TRIGONOMETRY	I0020
55	1200	1499	-.90	.07	.98	-.6	.95	-1.2	.21	.11	NUMERATION	I0021
56	1143	1499	-.68	.06	.98	-.7	.96	-1.1	.21	.12	PROBABILITY	I0022
57	1095	1499	-.51	.06	.99	-.3	.98	-.6	.16	.13	TRIGONOMETRY	I0023
58	1072	1499	-.43	.06	.99	-.5	.98	-.7	.18	.13	NUMERATION	I0024
59	1055	1499	-.37	.06	.98	-.9	.97	-1.4	.21	.13	STATISTICS	I0025
60	1068	1499	-.41	.06	.98	-.7	.97	-1.0	.20	.13	ALGEBRAE	I0026
61	1082	1499	-.46	.06	.99	-.6	.97	-1.0	.19	.13	GEOMETRY	I0027
62	1094	1499	-.50	.06	.97	-.9	.96	-1.3	.22	.13	STATISTICS	I0028
63	1103	1499	-.53	.06	.97	-1.0	.96	-1.3	.22	.12	ALGEBRAE	I0029
64	1163	1499	-.75	.06	.97	-.8	.94	-1.6	.23	.12	ALGEBRAE	I0030
65	1150	1499	-.70	.06	.98	-.6	.97	-.9	.19	.12	NUMERATION	I0031
66	1205	1499	-.92	.07	.98	-.4	.96	-.9	.18	.11	NUMERATION	I0032
67	1160	1499	-.74	.06	.98	-.5	.97	-.9	.19	.12	ALGEBRAE	I0033
68	1135	1499	-.65	.06	.97	-.8	.95	-1.5	.23	.12	MENSURATION	I0034
69	1131	1499	-.63	.06	.97	-.9	.96	-1.2	.22	.12	ALGEBRAE	I0035
70	1122	1499	-.60	.06	.98	-.7	.97	-1.0	.20	.12	NUMERATION	I0036
71	1165	1499	-.76	.06	.99	-.2	.99	-.2	.14	.12	TRIGONOMETRY	I0037
72	1145	1499	-.68	.06	.98	-.6	.96	-1.0	.20	.12	NUMERATION	I0038
73	1057	1499	-.38	.06	1.00	.2	1.01	.2	.11	.13	NUMERATION	I0039
74	1040	1499	-.32	.06	1.02	.9	1.03	1.4	.05	.13	NUMERATION	I0040
75	1081	1499	-.46	.06	.99	-.4	.99	-.4	.16	.13	ALGEBRAE	I0041
76	1027	1499	-.28	.06	1.00	.0	1.00	-.1	.14	.13	ALGEBRAE	I0042
77	1099	1499	-.52	.06	1.01	.4	1.02	.6	.08	.12	GEOMETRY	I0043
78	1090	1499	-.49	.06	1.00	-.1	.99	-.2	.14	.13	MENSURATION	I0044
79	1002	1499	-.20	.06	.99	-.7	.98	-.9	.19	.13	TRIGONOMETRY	I0045
80	995	1499	-.18	.06	1.00	.1	1.01	.3	.12	.13	TRIGONOMETRY	I0046
81	1043	1499	-.33	.06	1.00	-.2	1.00	-.2	.14	.13	NUMERATION	I0047
82	1081	1499	-.46	.06	.99	-.3	.99	-.3	.16	.13	ALGEBRAE	I0048
83	1048	1499	-.35	.06	.97	-1.1	.97	-1.4	.22	.13	MENSURATION	I0049
84	1116	1499	-.58	.06	.99	-.2	.99	-.4	.15	.12	TRIGONOMETRY	I0050
85	1090	1499	-.49	.06	.98	-.6	.98	-.8	.18	.13	MENSURATION	I0051
86	1070	1499	-.42	.06	.99	-.2	.99	-.4	.15	.13	NUMERATION	I0052
87	1073	1499	-.43	.06	.99	-.5	.98	-.6	.17	.13	MENSURATION	I0053
88	1049	1499	-.35	.06	.99	-.3	.99	-.4	.15	.13	NUMERATION	I0054
89	1137	1499	-.66	.06	1.00	.1	1.01	.2	.11	.12	MENSURATION	I0055
90	1105	1499	-.54	.06	1.01	.2	1.01	.3	.10	.12	NUMERATION	I0056
91	1043	1499	-.33	.06	1.02	.7	1.03	1.1	.07	.13	ALGEBRAE	I0057
92	1056	1499	-.37	.06	1.03	1.1	1.04	1.5	.03	.13	MENSURATION	I0058
93	1033	1499	-.30	.06	1.03	1.4	1.05	2.0	.01	.13	TRIGONOMETRY	I0059
94	1042	1499	-.33	.06	1.02	.8	1.02	1.0	.06	.13	ALGEBRAE	I0060
95	1065	1499	-.40	.06	1.00	.0	1.00	.0	.13	.13	NUMERATION	I0061
96	1133	1499	-.64	.06	1.01	.5	1.02	.6	.07	.12	TRIGONOMETRY	I0062
97	1110	1499	-.56	.06	1.01	.4	1.01	.4	.09	.12	TRIGONOMETRY	I0063
98	1053	1499	-.37	.06	1.02	.7	1.02	.9	.07	.13	MENSURATION	I0064
99	1155	1499	-.72	.06	1.03	.8	1.05	1.5	.00	.12	MENSURATION	I0065
MEAN	915.1	1499.0	.00	.06	1.00	.1	1.00	.0			67.3 67.2	
S.D.	233.7	.0	.72	.01	.03	2.5	.04	2.6			9.1 8.8	

APPENDIX XV

ITEMS LOGIT AFTER THE REMOVAL OF MISFIT RESPONSES

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH MAT TOPICS	Item
1	746	1499	.83	.05	.99	-.7	.99	-1.0	.20	.18	ALGEBRAIC	I0001
2	1002	1499	.10	.06	1.06	1.8	1.06	1.7	.02	.17	ALGEBRAIC	I0002
3	555	1499	1.37	.05	1.02	1.2	1.02	1.3	.12	.18	NUMERATION	I0003
4	738	1499	.85	.05	1.01	.6	1.01	.6	.17	.18	ALGEBRAIC	I0004
5	427	1499	1.77	.06	1.04	1.5	1.05	1.6	.06	.17	TRIGNOMETRY	I0005
6	579	1499	1.30	.05	1.01	.8	1.02	1.2	.14	.18	ALGEBRAIC	I0006
7	785	1499	.72	.05	1.05	1.5	1.06	1.3	.04	.18	PROBABILITY	I0007
8	986	1499	.15	.06	.98	-.9	.98	-.8	.21	.17	ALGEBRAE	I0008
9	702	1499	.95	.05	1.02	2.0	1.03	2.0	.12	.18	ALGEBRAE	I0009
10	804	1499	.67	.05	1.02	2.0	1.02	2.0	.12	.18	TRIGONOMETR	I0010
11	698	1499	.96	.05	1.03	2.0	1.03	2.0	.11	.18	TRIGONOMETR	I0011
12	566	1499	1.34	.05	1.03	1.7	1.03	1.6	.11	.18	ALGEBRAIC	I0012
13	578	1499	1.30	.05	1.01	.5	1.01	.3	.16	.18	NUMERATION	I0013
14	1217	1499	-.68	.07	1.05	1.3	1.10	2.0	.02	.14	ALGEBRAIC	I0014
15	426	1499	1.77	.06	.99	-.3	.99	-.2	.18	.17	MENSURATION	I0015
16	792	1499	.71	.05	.99	-.6	1.00	-.3	.20	.18	NUMERATION	I0016
17	958	1499	.23	.05	1.03	1.9	1.05	2.0	.08	.17	ALGEBRAE	I0017
18	1377	1499	-1.66	.09	.99	-.1	.93	-.9	.16	.10	TRIGONOMETR	I0018
19	1351	1499	-1.44	.09	1.00	.0	1.02	.2	.11	.11	TRIGONOMETR	I0019
20	1319	1499	-1.22	.08	.98	-.4	.92	-1.2	.20	.12	TRIGONOMETR	I0020
21	1280	1499	-.99	.07	.97	-.5	.93	-1.2	.22	.13	NUMERATION	I0021
22	1200	1499	-.61	.07	.96	-1.1	.91	-2.0	.28	.15	PROBABILITY	I0022
23	1143	1499	-.38	.06	.96	-1.2	.94	-1.8	.27	.16	TRIGONOMETR	I0023
24	1095	1499	-.21	.06	.98	-.7	.96	-1.3	.22	.16	NUMERATION	I0024
25	1072	1499	-.13	.06	.98	-.8	.96	-1.3	.22	.16	STATISTICS	I0025
26	1055	1499	-.07	.06	.98	-1.0	.96	-1.6	.24	.17	ALGEBRAE	I0026
27	1068	1499	-.11	.06	.97	-1.2	.95	-1.7	.25	.16	GEOMETRY	I0027
28	1094	1499	-.20	.06	.95	-1.8	.93	-2.0	.30	.16	STATISTICS	I0028
29	1103	1499	-.23	.06	.97	-1.1	.95	-1.5	.25	.16	ALGEBRAE	I0029
30	1163	1499	-.46	.06	.96	-1.3	.91	-2.4	.29	.15	ALGEBRAE	I0030
31	1150	1499	-.41	.06	.97	-.9	.95	-1.4	.24	.15	NUMERATION	I0031
32	1205	1499	-.63	.07	.98	-.5	.95	-1.1	.21	.15	NUMERATION	I0032
33	1160	1499	-.44	.06	.97	-.9	.94	-1.6	.24	.15	ALGEBRAE	I0033
34	1135	1499	-.35	.06	.97	-1.0	.93	-1.9	.26	.16	MENSURATION	I0034
35	1131	1499	-.33	.06	.96	-1.4	.93	-2.0	.28	.16	ALGEBRAE	I0035
36	1122	1499	-.30	.06	.96	-1.3	.94	-2.0	.27	.16	NUMERATION	I0036
37	1165	1499	-.46	.06	.98	-.5	.97	-.9	.21	.15	TRIGONOMETR	I0037
38	1145	1499	-.39	.06	.98	-.8	.95	-1.4	.23	.16	NUMERATION	I0038
39	1057	1499	-.08	.06	1.01	.3	1.00	.1	.15	.17	NUMERATION	I0039
40	1040	1499	-.02	.06	1.04	1.6	1.06	2.2	.06	.17	NUMERATION	I0040
41	1081	1499	-.16	.06	.98	-.7	.97	-1.0	.22	.16	ALGEBRAE	I0041
42	1027	1499	.02	.06	1.01	.3	1.01	.3	.15	.17	ALGEBRAE	I0042
43	1099	1499	-.22	.06	1.01	.5	1.01	.5	.13	.16	GEOMETRY	I0043
44	1090	1499	-.19	.06	1.00	.0	.99	-.2	.17	.16	MENSURATION	I0044
45	1002	1499	.10	.06	.98	-.9	.97	-1.2	.22	.17	TRIGONOMETR	I0045
46	995	1499	.12	.06	1.00	.1	1.01	.3	.16	.17	TRIGONOMETR	I0046
47	1043	1499	-.03	.06	1.00	-.1	1.00	.0	.17	.17	NUMERATION	I0047
48	1081	1499	-.16	.06	.99	-.2	1.00	-.1	.18	.16	ALGEBRAE	I0048
49	1048	1499	-.05	.06	.96	-1.7	.95	-1.8	.27	.17	MENSURATION	I0049
50	1116	1499	-.28	.06	.99	-.4	.97	-.8	.20	.16	TRIGONOMETR	I0050
51	1090	1499	-.19	.06	.98	-.8	.97	-1.0	.22	.16	MENSURATION	I0051
52	1070	1499	-.12	.06	1.00	.0	1.00	-.1	.16	.16	NUMERATION	I0052
53	1073	1499	-.13	.06	.98	-.7	.98	-.7	.21	.16	MENSURATIO	I0053
54	1049	1499	-.05	.06	1.00	-.1	1.00	-.2	.17	.17	NUMERATION	I0054
55	1137	1499	-.36	.06	1.00	.0	1.00	.0	.16	.16	MENSURATION	I0055
56	1105	1499	-.24	.06	1.02	.6	1.02	.6	.12	.16	NUMERATION	I0056
57	1043	1499	-.03	.06	1.04	1.7	1.06	2.0	.06	.17	ALGEBRAE	I0057
58	1056	1499	-.07	.06	1.03	1.3	1.04	1.6	.08	.17	MENSURATION	I0058
59	1033	1499	.00	.06	1.06	2.0	1.08	1.1	.01	.17	TRIGONOMETR	I0059
60	1042	1499	-.03	.06	1.03	1.5	1.04	1.7	.08	.17	ALGEBRAE	I0060
61	1065	1499	-.10	.06	1.00	.1	1.00	.0	.16	.17	NUMERATION	I0061
62	1133	1499	-.34	.06	1.02	.8	1.04	1.2	.08	.16	TRIGONOMETR	I0062
63	1110	1499	-.26	.06	1.01	.5	1.02	.6	.12	.16	TRIGONOMETR	I0063
64	1053	1499	-.06	.06	1.05	2.0	1.06	2.0	.04	.17	MENSURATION	I0064
65	1155	1499	-.42	.06	1.05	1.4	1.09	2.0	.01	.15	MENSURATION	I0065
MEAN	1015.2	1499.0	.00	.06	1.00	.0	1.00	.0			71.1 71.4	
S.D.	209.1	.0	.67	.01	.03	1.3	.05	1.6			8.1 7.5	

APPENDIX XVI

INDEPENDENT T-TEST ANALYSIS OF SPSS

Research Hypotheses

The Research Hypotheses are used to test the significance of the Research Questions in this regards. The purpose is to test the invariance of the MAT items across the subgroups in the selected states of study. The sixty-five items that were selected was used in testing the invariance of MAT items across the subgroup. In resolving the stated hypotheses in this study, the SPSS statistics software was used. The t-test statistics of SPSS was used to test the hypotheses in order to determine the invariance of the items across the subgroups.

TABLE 17 Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Gender * Social Economic Status	1499	100.0%	0	.0%	1499	100.0%
Gender * School Type	1499	100.0%	0	.0%	1499	100.0%
Gender * Location	1499	100.0%	0	.0%	1499	100.0%

These are the cross tabulation for Gender, SES, School types, and location.

TABLE 18 Gender* social Economic Status Crosstabulation

Count		Social Economic Status		Total
		High Social Economic Status	Low Social Economic Status	
Gender	Female	279	463	742
	Male	350	407	757
Total		629	870	1499

TABLE 19 Gender * School Type Crosstabulation

Count		School Type		Total
		Public School	Private School	
Gender	Female	552	190	742
	Male	488	269	757
Total		1040	459	1499

TABLE 20 Gender * Location Crosstabulation

Count		Location Rural	Urban	Total
Gender	Female	188	554	742
	Male	121	636	757
Total		309	1190	1499

Research Hypothesis One

There is no significant difference between the mean achievement scores of the male and female testees in Mathematics Achievement Test (MAT)?

In resolving the above stated hypothesis, the SPSS t-test statistics software was used.

TABLE 21 Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
TEST	FEMALE	742	.6742	.07697	.00283
	MALE	757	.6763	.07135	.00259

An independent-sample t-test indicated that scores were insignificantly higher for male (N=757, M= .6763, SD= .07135) than for female (N=742, M=.6742, SD= .07697), t(1497)= -.561, p> .05.

TABLE 22 Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
TEST	Equal variances assumed	3.239	.072	-.561	1497	.575	-.00215	.00383	-.00967	.00537
	Equal variances not assumed			-.561	1483.414	.575	-.00215	.00384	-.00967	.00537

Since Levene's test for equality of variance is insignificant, i.e. p> .05, the statistics for the row equal variance assumed was reported with the degree of freedom rounded to the nearest whole number (APA, 2005).

Scores on the gender subscale were higher for male (N=757, M= .6763, SD= .07135) than for female (N=742, M=.6742, SD= .07697), t(1497)= -.561, p> .05, d=.43. Levene's test indicated equal variance (F= 3.239, p= .072), so degree of freedom were 1497. This indicates no significant since t(1497)= -.561 at 5% significant level has p= .575. This means p> .05 therefore, the stated hypothesis was accepted, that is, there was no significant difference between the mean achievement score of male and female testees in MAT.

Research Hypotheses Two

There is no significant difference between the mean achievement score of Low Social Economic Status and High Social Economic Status testees' in Mathematics Achievement Test (MAT).

TABLE 23 T-TEST Group Statistics

	Social Economic Status	N	Mean	Std. Deviation	Std. Error Mean
TEST	Low Social Economic Status	870	.6735	.07253	.00246
	High Social Economic Status	629	.6777	.07637	.00304

In the group statistics above, an independent-sample t-test indicated that scores were slightly insignificantly higher for HSES (N=629, M= .6777, SD= .07637) than for LSES (N=870, M=.6735, SD= .07253), t(1497)= -1.071,
p> .05,

TABLE 24 Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
TEST	Equal variances assumed	.980	.322	-1.071	1497	.284	-.00416	.00388	-.01177	.00346
	Equal variances not assumed			-1.062	1311.211	.288	-.00416	.00391	-.01184	.00352

Scores on the SES subscales were higher for HSES (N=629, M= .6777, SD= .07637) than for LSES (N=870, M=.6735, SD= .07253), t(1497)= -1.071, p> .05. Levene's test indicated equal variances assumed (F= .980, p= .322) This simply means t(1497)= -1.071 at .05 level of significance has P= .284. The indication was that P> 0.05 which means no significant. Therefore, the stated hypothesis of no significant difference between the mean achievement score of **LSES** and **HSES** of testees in MAT was accepted.

Research Hypothesis 3

There is no significant difference between the mean achievement scores of the public school testees and the private school testees in MAT.

In the group t-test statistics below, an independent-sample t-test of the group statistics indicated that scores were higher for private schools (N=459, M= .6848, SD= .07354) than for public schools (N=1040, M=.6871, SD= .07409), t(1497)= -3.328, p> .05,. This means that the M and SD of the private schools are slightly higher but it was insignificant.

TABLE 25 T- TEST Group Statistics

	SCHL	N	Mean	Std. Deviation	Std. Error Mean
TEST	PUBLIC	1040	.6871	.07409	.00230
	PRIVATE	459	.6848	.07354	.00343

Scores on the school types subscale below were higher for private schools insignificantly (N=459, M= .6848, SD= .07354) than for public schools (N=1040, M=.6871, SD= .07409), t(1497)= -3.328, p> .05 Levene's test indicated equal variances assumed (F= .028, p= .866). This indicated that t(1497)= -3.328 at .05 level of significance had p= .062 which means that p> .05. This implies non- significance.

TABLE 26 Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	T	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
TEST	Equal variances assumed	.203	.866	-3.328	1497	.062	-.01379	.00414	-.02191	-.00566
	Equal variances not assumed			-3.338	882.240	.062	-.01379	.00413	-.02189	-.00568

This indication of non-significance therefore shows that, the stated hypothesis of no significant different was accepted. This implies that there was no significant difference between the mean achievement scores of the **public** school testees and the **private** school testees in MAT.

Research Hypothesis 4

There is no significant difference between the mean achievement scores of the urban testees and the rural testees in Mathematics Achievement Test (MAT).

In the group t-test statistics below, an independent-sample t-test of the group statistics indicated that scores were higher insignificantly for Rural (N=309, M= .6782, SD= .07066) than for Urban (N=1190, M=.6745, SD= .07506), t(1497)= .781, p> .05.

TABLE 27 T-TEST Group Statistics

	LOCATION	N	Mean	Std. Deviation	Std. Error Mean
TEST	RURAL	309	.6782	.07066	.00402
	URBAN	1190	.6745	.07506	.00218

Scores on Location subscale above were higher for Rural (M= .6782, SD= .07066) than for Urban (M=.6745, SD= .07506), t(1497)= .781, p> .05, d=0.96. The independent-samples t-test table shows that the P-value of the Levene's test for equality of variances was assumed (F= 1.071, p= .301). Therefore t(1497)= .781 at 95% confidence has p= .435 which was greater than .05 i.e. p> .05. This means that there was no significance.

TABLE 28 Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
TEST	Equal variances assumed	1.071	.301	.781	1497	.435	.00370	.00474	-.00559	.01299
	Equal variances not assumed			.810	503.715	.419	.00370	.00457	-.00528	.01268

This indicates that the stated hypothesis was hereby accepted. There was no significant difference between the mean achievement scores of the **urban** testees and the **rural** testees in MAT.

APPENDIX XVII

POPULATION, BASIC EDUCATION AND SECONDARY SCHOOL ANALYSIS OF DELTA STATE GOVERNMENT BETWEEN 2007 AND 2013

ADMINISTRATION

Education, Basic and Secondary in Delta State

Published on Wednesday, 13 June 2012 15:15 Hits: 2562

GOVERNMENT EXPENDITURE ON EXAMINATION 2009-2013

S/No	Name of Examination	Year	No. of Candidates	No. of Pass	% Pass	Government Expenditure (N)	
1.	WASSCE	2009 2010 2011 2012 2013	29,761 38,705 51,042 32769 50033	12,529 14,065 19,762	42.1% 36.34% 38.72%	218,916,840.00 218,916,840.00 176,447,400.00	614,281,080.00
2.	Cognitive/Placement Exam for Primary six Pupils	2009 2010 2011 2012	74,892 80,469 84,945 56,647	68,630 64,094 76,339	91.64% 79.65% 89.87%	37,500,000.00 37,500,000.00 37,500,000.00 37,500,000.00	150,000,000.00
3.	Basic Education Certificate Examination (formally Junior School Certificate Examination	2009 2010 2011 2012	57,646 58,272 66,800 66,551	47,641 53,739 56,007	82,64% 92.22% 83.84%	65,000,000.00 84,000,000.00 82,000,000.00 82,000,000.00	313,000,000.00
4.	Promotion Examination in public post primary Schools	2009 2010 2011 2012	158,000 164,995 179,700 159,946			62,159,146.00 68,940,160.00 85,344,720.00 86,566,022.00	303,010,048.00
5.	Promotion Examination in public primary schools	2009 2010 2011 2012	288,778 330,452 380,721 282,090			28,877,854.00 33,045,200.00 38,072,100.00 40,331,500.00	140,326,654.00
6.	1st & 2nd Terms Exams in public primary &post primary Schools	2009 2010 2011 2012	158,000 164,995 179,700 112,022			150,000,000,00 150,000,000.00 168,145,950.00	468,145,950.00
7.	Enrolment	2012	27,934			218,961,840.00	218,916,840.00

	fees and Incidental Costs Payable by SS III Students in public Senior Secondary and Technical Schools towards 2012 May/June and June/July Senior School Certificate (WAEC & NABTEB) Examination						
				Total			2,207,680,572.00

Website: info@deltastate.gov.ng

POPULATION, BASIC EDUCATION AND SECONDARY SCHOOL ANALYSIS OF OYO STATE GOVERNMENT BETWEEN 2007 AND 2013 ADMINISTRATION
GOVERNMENT EXPENDITURE ON EXAMINATION 2009-2013

S/No	Name of Examination	Year	No. of Candidates	No. of Pass	% Pass	Government Expenditure (N)	
1.	WASSCE	2009 2010 2011 2012 2013	37,761 48,705 54,042 36,321 50,437	11,429 14,095 29,762	39.1% 37.34% 36.72%	328,716,440.00 28,816,540.00 276,567,600.00	634,100,580.00
3.	Basic Education Certificate Examination (formally Junior School Certificate Examination	2009 2010 2011 2012 2013	57,646 58,272 66,800 66,551	48,641 56,739 66,007	82,64% 92.22% 83.84%	76,000,000.00 85,000,000.00 72,000,000.00 96,000,000.00	329,000,000.00

Website:info@oyostate.gov.ng

APPENDIX XVIII

Glossary of Rasch Measurement Terminology	
Glosario Español www.rasch.org/rmt/glosario.htm	
Ability	the level of successful performance of the objects of measurement on the variable.
Agent of Measurement	the tool (items, questions, etc.) used to define a variable and position objects or persons along that variable.
Anchor	the process of using anchor values to insure that different analyses produce directly comparable results.
Anchor Value	a pre-set logit value assigned to a particular object, agent or step to be used as a reference value for determining the measurements or calibrations of other objects, agents or steps.
Anchor Table	the table of Anchor Values used during Rasch analysis of an Input Grid and so included in the Results Table produced. The Anchor Table has the same format as the Results Table.
Bias	A change in logit values based on the particular agents or objects measured.
Calibration	a difficulty measure in logits used to position the agents of measurement along the variable.
Categories	levels of performance on an observational or response format.
Cell	Location of data in the spreadsheet, given by a column letter designation and row number designation e.g. B7
Common Scale	a scale of measurement on which all agents and objects can be represented.
Column	Vertical line of data in the Spreadsheet data, usually representing in an Input Grid all responses to a particular item, or in a Results Table, all statisitics measuring the same attribute of agents or objects.
Content	The subject area evoked and defined by an agent.
Dichotomous Response	a response format of two categories such as correct-incorrect, yes-no, agree-disagree.

Difficulty	the level of resistance to successful performance of the agents of measurement on the variable.
Discrepancy	one or more unexpected responses.
Disturbance	one or more unexpected responses.
Expected Response	the predicted response by an object to an agent, according to the Rasch model analysis.
Fit Statistic	a summary of the discrepancies between what is observed and what we expect to observe.
Heading	An identifier or title for use on tables, maps and plots.
Independent	Not dependent on which particular agents and objects are included in the analysis. Rasch analysis is independent of agent or object population as long as the measures are used to compare objects or agents which are of a reasonably similar nature.
Infit	an information weighted fit statistic that focuses on the overall performance of an item or person, i.e, the information-weighted average of the squared standardized deviation of observed performance from expected performance. The statistic plotted and tabled by Rasch is this mean square normalized.
Interval scale	Scale of measurement on which equal intervals represent equal amounts of the vairable being measured.
Item	agent of measurement, not necessarily a test question, e.g., a product rating.
Iteration	one run through the data by the Rasch calculation program, done to improve estimates by minimizing residuals.
Link	Relating the measures derived from one test with those from another test, so that the measures can be directly compared.
Logit/ Wit	the unit of measure used by Rasch for calibrating items and measuring persons. A $\log_e$ odds transformation of the probability of a correct response.
Map	a bar chart showing the frequency and spread of agents and objects along the variable.
Matrix	a rectangle of responses with rows (or columns) defined by objects and columns (or rows) defined by agents.

Measure	the location (usually in logits) on the latent variable. The Rasch measure for persons is the person ability. The Rasch measure for items is the item difficulty.
Normal	a random distribution, graphically represented as a "bell" curve which has a mean value of 0 and a standard deviation of 1.
Normalized	the transformation of the actual statistics obtained so that they are theoretically part of a normal distribution.
Object of Measurement	people, products, sites, to be measured or positioned along the variable.
Observed Response	The actual response by an object to an agent.
Outfit	an outlier sensitive fit statistic that picks up rare events that have occurred in an unexpected way. It is the average of the squared standardized deviations of the observed performance from the expected performance. Rasch plots and tables use the normalized unweighted mean squares so that the graphs are symmetrically centered on zero.
Outliers	unexpected responses usually produced by agents and objects far from one another in location along the variable.
Person	The object of measurement, not necessarily human, e.g., a product.
Plot	an x-y graph used by Rasch to show the fit statistics for agents and objects.
Point Labels	the placing on plots of the identifier for each point next to the point as it is displayed.
Poisson Counting	a method of scoring tests based on the number of occurences or non-occurences of an event, e.g. spelling mistakes in a pice of dictation.
Process	the psychological quality, i.e.,the ability, skill, attitude, etc., being measured by an item.
PROX	the normal approximation estimation formula, used by some Rasch programs for the first part of the iteration process.
Rasch, Georg	Danish Mathematician (1906-1980), who first propounded the application of the statistical approach used by Rasch.

Rasch Model	a mathematical formula for the relationship between the probability of success (P) and the difference between an individual's ability (B) and an item's difficulty (D). P=exp(B-D)/(1+exp(B-D)) or $\log_e$ [P/(1-P)] = B – D
Rating Scale	A format for observing responses wherein the categories increase in the level of the variable they define, and this increase is uniform for all agents of measurement.
Reliability	the ratio of sample or test variance, corrected for estimation error, to the total variance observed.
Residuals	The difference between data observed and values expected.
Response	The value indicating degree of success by an object on an agent, and entered into the appropriate cell of an Input Grid.
Results Table	a report of Rasch calculations.
Rigidity	when agents, objects and steps are all anchored, this is the logit inconsistency between the anchoring values, and is reported on the Iteration Screen and Results Table. 0 represents no inconsistency.
Row	a horizontal line of data on a Spreadsheet, usually used, in the Input Grid, to represent all responses by a particular object. The top row of each spreadsheet is reserved for Rasch control information.
Scale	The quantitive representation of a variable.
Score points	the numerical values assigned to responses when summed to produce a score for an agent or object.
Separation	the ratio of sample or test standard deviation, corrected for estimation error, to the average estimation error.
Standard Deviation	the root mean square of the differences between the calculated logits and their mean.
Standard Error	an estimated quantity which, when added to and subtracted from a logit measure or calibration, gives the least distance required before a difference becomes meaningful.
Steps	the transitions between adjacent categories ordered by the definition of the variable.

TOP	The value shown in the Results Table for an agent on which no objects were successful, (so it was of top difficulty), or for an object which succeeded on every agent (so it was of top ability)
Top Category	The response category at which maximum performance is manifested.
UCON	the unconditional (or "joint" JMLE) maximum likelihood estimation formula, used by some Rasch programs for the second part of the iteration process.
UNSURE	Rasch was unable to calibrate this data and treated it as missing.
Unweighted	the situation in which all residuals are given equal significance in fit analysis, regardless of the amount of the information contained in them.
Variable	the idea of what we want to measure A variable is defined by the items or agents of measurement used to elicit its manifestations or responses.
Weighted	the adjustment of a residual for fit analysis, according to the amount of information contained in it.

Based on: Wright, B.D. & Linacre J.M. (1985) Microscale Manual. Westport, Conn.: Mediax Interactive Technologies, Inc.

APPENDIX XIX

Steps in Test Construction

The steps below based on the item response theory (IRT) using the Rasch model shows the dependent variable (academic achievement) which is dependent on gender, school type, SES and the school location factors of the testees. The independent variables as indicated in the frame work below are gender (male or female), school type (public (A) or private (B)), school location (urban or rural) and Social Economic Status (Low and High). The chart summaries the various steps taken in the development and validation of the achievement test. The development of the test is based on the item response theory using the Rasch Model.

Since the steps are based on Item Response Theory (IRT) using the Rasch Model, it is therefore made up of three essential components for the purpose of this study, they are:

Steps in Test Development and Validation: These are the various steps involved in test development and validation for an achievement test using the Item Response Theory (IRT) approach.

Independent Variable: These are variables when manipulated causes some differences or changes in the dependent variable.

Dependent Variables: are the changes or differences that occur as a result of the independent variables. It may also be regarded as the outcome or output variable in a study. It helps to ascertain the effect of the independent variable. The independent variables of gender (male or female), school type (public or private), school location (urban or rural) and social economy status (High and Low) were studied as they affect the dependent variable (academic achievement).

Chart showing Steps in Test Construction