INVARIANCE OF ITEM PARAMETERS OF POLYTOMOUS GEOGRAPHY ITEMS IN SENIOR SCHOOL CERTIFICATE EXAMINATION IN NIGERIA

BY

ABDULSALAM-NUHU, Rashidat Omotayo

01/25OC296

Educational Research, Measurement and Evaluation

A Ph.D. THESIS PRESENTED TO THE DEPARTMENT OF SOCIAL SCIENCES EDUCATION, FACULTY OF EDUCATION, UNIVERSITY OF ILORIN, ILORIN, NIGERIA IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF DOCTOR OF PHILOSOPHY(PH.D) IN EDUCATIONAL RESEARCH, MEASUREMENT AND EVALUATION

Supervisor: Dr. A.A. Ogunlade

CERTIFICATION

This is to certify that this research titled “Invariance of Item Parameters of Polytomous Geography Items in Senior School Certificate Examination in Nigeria” was carried out by **ABDULSALAM-NUHU**, Rashidat, Matric. No.01/25OC296, and has been read and approved as meeting part of the requirements of the Department of Social Sciences Education, Faculty of Education, University of Ilorin, Ilorin, Kwara State, Nigeria for the award of the Doctor of Philosophy (Ph.D.) Degree in Educational Research, Measurement and Evaluation.

____________________ ____________________

Dr. A.A Ogunlade Date
(Supervisor)

____________________ ____________________

Prof. Bolanle O. Olawuyi Date
(Head of Department)

____

____

Prof. Medinat F.Salman Date
(Dean, Faculty of Education)

____

____

(External Examiner) Date

DEDICATION

This research is dedicated to Almighty Allah (S.W.T) and my parents Alhaji and Alhaja Abdulkadir Nuhu.

ACKNOWLEDGEMENTS

All praises and Adorations are due to Almighty Allah (S.W.T), the Creator and the Sustainer of the universe, for making it possible for me to come this far in my career. I thank Him for making my academic pursuit a reality. May His peace and Blessing be upon the soul of Prophet Muhammad (S.A.W), his household, his companions and the generality of the Muslim Ummah till the day of accountability.

I am indebted to my supervisor, Dr. A.A. Ogunlade for his encouragement, intellectual contribution, and relentless effort which made the successful completion of this work a reality. May God Almighty grant him abundant blessings in all his endeavours. I also appreciate the intellectual contribution of the Head of Department of Social Sciences Education, University of Ilorin, Ilorin Nigeria, Prof. Bolanle O. Olawuyi. I am also indebted to Prof. H. O. Owolabi for his encouragement, relentless effort; intellectual input and his fatherly and professional advice from time to time which contributed immensely to the successful completion of this study. I pray God in His infinite mercy be with you always. The success of this work is also credited to the inspiration received from Professors C.O. Daramola, O.E. Abdullahi, F.A. Olasehinde-Williams and A.A. Jekayinfa. Drs. R.W. Okunloye (Postgraduate coordinator), I.O.O. Amali, H.T. Yusuf, M. B. Bello, D.S. Daramola, A. O. Balogun, A.A. AbdulSalam, and M.I. Jimoh, Messrs. I.N. Olokoba and R.K. Oniye, Mrs I.N. Balogun, Mrs J. Oladele, Miss. L.S. Akinsemoyin, other lecturers and non-academic staff of the Department of Social Science Education and lecturers and non-academic staff of the Department of Adult and Primary Education, University of Ilorin, Ilorin.

My gratitude also goes to my parents, Alhaji and Alhaja Nuhu. May Almighty Allah continue to be with them and reward them in this world and the hereafter for standing by me from childhood. I also express my gratitude to my siblings Nuhu AbdulAkeem, Nahimat Nuhu, Lukman Nuhu, Sukurat Nuhu, Baliqees Nuhu and my late sister Aminat Nuhu I pray Allah (swt) forgive and grant her jannah.

I sincerely acknowledge the contributions of Professors L.A Yahaya, R.A. Lawal and K. Etsey, Dr M. O. Ogunjimi, Mallam, Abubakar Toyin, the Principal I.G.S Ilorin the entire staff and students of I.G.S. Ilorin, my friends and colleagues at work Mrs Nurudeen, Mrs Sakibu, Imam Osere, Mallam Saheed, and Mallam Abdulfatah for their support. Also, Mr Abubakar Sidiq Imam (unilorin radio), Mr Muritala Eleja, Alfa Kamil, Owonikoko business centre and my colleagues S. Adewoye, O.A. Jacob (Postgraduate Representative), T.O. Akano, J.K. Akande, M. O. Aderinto, K.E. Paul, E. Peter, A. Cobina, A. Yekeen, R. Suleiman, A. Osho, S.T. Bamidele, Mr. Mahmud, Jamiu AbdulQodir and Dr. Metibemu. My

profound gratitude also goes to Majeobaje of Ibokun (Former Director Osun State Ministry of Education), Alhaji Shittu Suraju, my brother Mr Musa Alatare (Imo state College of Education) and the Ademolas in Ibadan (Water Water)

Finally, I must sincerely appreciate the support and prayers of my husband Mr. AbdulSalam Mujeeb and children Muhammad Mustapha, Mansurah, AbdulBasit, and Hammad for the sacrifice, patience and understanding they displayed during the period of my studies.

May Almighty Allah (Subuhanah Watahala) bless them all.

AbdulSalam-Nuhu,
Rashidat

TABLE OF CONTENTS

Content

Page

APPENDICES

LIST OF TABLES

LIST OF FIGURES

Page

ABSTRACT

The hallmark of a test instrument is its quality. Test validity is key in ascertaining the quality of tests in all school subjects and it is critical at the School Certificate Level. Geography is examined at this level and a lot of studies have been conducted to determine the item parameters of objective tests using traditional Classical Test Theory. Therefore, it has become necessary to utilize the Item Response Theory procedure for analyzing the invariance of Item Parameters of Polytomous Items in Geography at the Senior School Certificate Examination Level in Nigeria. The objectives of the study were to examine: (i) examinees' ability estimates; (ii) item difficulty and discrimination estimates; (iii) the extent of invariance of item difficulty estimates across male and female students; and (iv) the extent of invariance of item difficulty estimates across Geography students in rural and urban areas.

This study employed the descriptive survey design. The population comprised all senior secondary school students in Nigeria. The target population was all senior secondary school three (SSS3) students that offered Geography in their final examinations in 2016/2017 academic session. A sample of 1,546 drawn out of 32,499 students from 876 senior secondary schools in Nigeria was selected. A multi-stage sampling procedure involving stratified random sampling, simple random sampling and proportionate sampling technique were used to select the respondents. The research instruments were questions on Element of Practical and Physical Geography (WASSCE 2016 Geography Paper 3). The data collected were analyzed using descriptive statistics, Robust Maximum Likelihood Estimation and Bayesian Ability Estimation methods and to answer the research questions, while t-test statistics was employed for hypotheses testing.

The findings of the study were that:

1. the examinees' ability estimates in Physical Geography showed that 49.35% of the examinees had performance below average, while 50.65% of the examinees had performance above average;
2. item 2B was the most difficult with difficulty index of 62.366 (mean = 3.719) ;
3. item 1B discriminated most with discrimination index of 4.364 (mean = 0.5076);
4. five out of 23 items representing 22% of the test items showed obvious variation in the difficulty estimates of the test items across gender;
5. 14 out of 23 items were more difficult for students of urban schools (Mean = -0.11, SD = 3.754) than the students of rural schools (Mean = -1.98, SD = 4.420);
6. there was a significant invariance of item parameter estimates across male and female geography students; and
7. there was a significant invariance of item parameter estimates across rural and urban geography students.

The study concluded that the items in the 2016 WASSCE Paper 3 had a large variation of item parameters based on school location (rural and urban) which implies that the test was not good for group comparison. Based on the findings, the study recommended among others that examination bodies should adopt Item Response Theory to reduce measurement problems encountered in score comparison and test equating.

Word Count: 486

CHAPTER ONE

INTRODUCTION

Background to the Study

Test is an important instrument in educational sector that is used to measure examinees' attainment. It is used to elicit information on a particular trait or behaviour of the examinees. Test is expected to treat every examinee equally without bias. When test is given to examinees, it is marked to make vital decisions such as placement, promotion and certification of the learners. Test is used to estimate examinees' ability in a given skill or task. When ability is estimated, one would be able to determine whether the examinee has a high or low ability. The obtained scores are used to locate and estimate the individual on a graph called the scale of ability. When the ability measured is obtained then, the student can be assessed in based on the amount of knowledge he or she has and comparison among students can also be made for the purpose of assigning grades, positioning and awarding scholarships.

Abiri (2006) described a test as a task, treatment or situation designed to elicit the behaviour or performance of persons or things with a view to drawing inferences about specific abilities or attributes of those persons or things. Most of the tests could either be performance or achievement tests. Achievement tests are therefore, tests administered in order to ascertain how much behavioural objectives in the curriculum have been achieved. According to Odili (2013), achievement testing requires that test items should call for knowledge gained in the subject matter in the school and nothing less.

Onwuka (1981), explained educational objectives in terms of knowledge (cognitive domain), attitude (affective domain) and practical (psychomotor domain) which was in line with Bloom's (1956) observation. With this, assessment is said to be balanced in education when it satisfies the demands of the three major domains of educational objectives. Hence, assessment of teaching and learning process is usually directed towards these three domains. Bandele (2002) noted that the three domains should be assessed analytically to have individual in totality and make the receiver of education to live a fulfilled life and contribute meaningfully to the society in which he lives.

The cognitive objectives refer to the intellectual results of schooling, the improvement in the child's cerebral structure, his increase in knowledge and his ability to reason rather than just to remember. The affective domain is the emotional education and the learners' attainment of certain desirable attitudes, interest and appreciation; while psychomotor objectives refer to physical and practical manipulative skills learnt at school (Nwana, 1979). The three domains of educational objectives according to Oyesola, (1986) are inter-related. In general, cognitive domain deals with writing long statements, ability to recall what has being taught. To measure the cognitive domain, then the use of essay and objective test is important as tests are given to learners so as to measure their ability levels. FRN (2014) sees the acquisition of appropriate skills, abilities and the development of mental, physical and social abilities and competences as equipment for the individual to live and contribute to the development of the society.

Also, FRN (2014) pointed out that secondary education is expected to raise a generation of people who can think for themselves, respect the views and feelings of others, respect the dignity of labour, appreciate those values as specified by the broad national goals and live as good citizens. For these objectives to be achieved, the use of cognitive domain cannot be wiped out, thus the use of essay and objective tests. Okafor (2015) asserted that test procedure employed in the evaluation are intended to produce educated individuals who may or not take any particular subject in their professional pursuits but who could find themselves in any profession that will be useful to them in their future lives. He explained further that practical work promotes problem solving and self-reliance in real life situation. Getting involved in tests can also enable students to learn the inter-relationship between Geography and other science subjects.

Geography for a very long time is a subject being offered by students at the Secondary School Level. Geography as a subject deals with the study of man and his environment. The component of Geography includes: Physical Geography, Human and Regional Geography and Map reading at the West African Senior School Certificate Examination (WASSCE). Geography is divided into examination Paper 1: this consists of objective questions which are selected from all the components of Geography. Paper 2 consists of nine essay-type questions selected from Human and Regional Geography while Paper 3 consists of eight essay questions selected from Physical Geography and Map Reading.

Most of the Geography topics taught to the students apply to what is happening in day to day activities. In the course of assessing students' ability, essay examination is compulsory to the students of geography and this can only be possible by giving out tests to assess the cognitive and psychomotive aspect of the subject. Nenty (2004) noted that for measurement in education to be meaningful, the objectives to be measured, the number to be assigned and the rules of the assignment of the number must be made clear. West African Examination Council (WAEC), National Examination Council (NECO) and National Business and Technical Education Board (NABTEB) are the three examining bodies in Nigeria that assess the cognitive aspects of knowledge. They are saddled with the responsibility of awarding Ordinary Level Certificate. All these examination bodies measure students' ability with the use of conventional Classical Test Theory (CTT) which is (X= T + E) with 'T' representing the hypothetical indicator, 'X' the observed indicator, and 'E' the amount of disagreement between 'T' and 'X' i.e. error. Classical test theory is equivalently stated as observed indicator.

In practice, we are interested in ‘X’ which is the observed score and is the performance of a task (test) given to an examinee which can be derived by counting. The error cannot be observed but it occurs without being planned for and this increases or decreases the observed score. In multiple choice tests, for instance, the error score might be through guessing. This error score makes estimates of examinees performance to be different even if the same test is assigned at different time. This CTT model yields scale that produces different results across different population that is it has weak theories and the assumption that error scores in high and low ability students are equal (Embretson & Reise, 2000). Also, CTT has been criticized for lacking the invariant principle thus, making researchers to shift from CTT to IRT. Hambleton and Jones (1993) therefore, explained that CTT has being the mostly used measurement theory but because of its lack of the principle of invariance and other criticisms by researchers, there is a paradigm shift from CTT to Item Response Theory (IRT).

IRT is a theory that builds around hidden trait ability and the observed ability of the examinee to correctly respond to any particular test item. IRT models are therefore mathematical functions which relate the probability of success on a task to the underlying proficiency measured by the task. IRT gives us the opportunity of attaining invariant of item parameters such as difficulty index (b-parameter), discrimination index (a-parameter), and guessing index, (c-parameter) in the case of dichotomously scored responses. IRT on the whole is a statistical framework for addressing measurement problems such as test development, test equating and identification of biased test items (Hambleton & Jones, 1991). With IRT, it is possible to construct trait line for exact measurement of a particular trait possessed by an individual. The above-mentioned merits of IRT made possible by invariant principle made a great difference between CTT and IRT. Most item response theory researchers and applications make use of unidimensional IRT models.

IRT has the principle of invariance which makes the examinee's ability to be stable. The invariance principle shows that the item parameters are independent of the distribution of examinees over the ability scale. When a test is used across subgroups of a population, measurement invariance is assumed. This is because test is expected to treat every examinee equally irrespective of the different ability proficiencies. If test does not treat every examinee equally then there is test unfairness. Odili (2013) explained test unfairness as a psychometric condition, in which a test item gives differential difficulty to testees of the same subject matter ability, due to the fact that they are from different sub-population of test-takers. However, Millsap (2000) noted that measurement invariance refers to stability in the psychometric properties of a measure across populations or occasions. So, it is assumed that the test measures the traits of interest in the same way through all groups. When this assumption is violated for a variable, the variable could be biased against one or more groups in a test. Glockner-Rist and Hoijtink (2003) explained that the establishment of measurement invariance is particularly essential in assessing group difference on a measure. Invariance in assessment is the ability of a test to treat the examinees' equally without bias. It should be noted that invariance includes stability in the ability of the examinee.

Rupp and Zumbo (2006) defined invariance as a characteristic that shows the values of the parameters that are the same in different populations or across different situations of interest, which is assessed when they are estimated repeatedly with different samples. In measurement, one need to consider two different groups of population parameters because test data are the result of the intersection of item and examinee sample spaces and the model bind the examinees and items together. In essence, invariance implies that, when an instrument like test is given to different groups of examinee say two, one should be able to say that the item properties measure ability of the examinees equally. It is essential to make comment about a specific assessment instrument for groups of examinees that share characteristics with another group that have being scored with the instrument to show different groups are comparable.

In IRT, when an examinee takes an easy or hard test it is possible to have the same performance which is not possible in CTT (Hambleton & Jones 1993). The correct responses to items are counted and it is shown on a metric scale which will show the ability estimate of an examinee based on low, medium and high scale. This is telling us the level of possession of the latent trait in an individual that is how much of a trait or characteristic an examinee possesses. IRT models are also categorized depending on the type of scoring pattern.

The assessment of cognitive domain in any area will be better studied if the examinees responses are polytomously scored rather than dichotomously scored. Cognitive outcomes may be simply studied using dichotomous scoring but are better done with polytomous scoring. This makes generalized partial credit (GPCM) scoring essential in many assessment situations this is because it believed that it will lead to a more precise estimate of person's ability than a simple pass/fail score.

Ability of examinees may be determined by gender, school location, ability level and ethnicity. Literatures have shown that gender have influence on students' academic performance. Ogunkola and Fayombo (2009) said that urban students performed better than their rural counterparts while Mberekpe (2013) said rural students performed better than their urban counterparts. However, Agbaje and Awodun (2013) were of the opinion that neither rural nor urban students can claim to be superior to one another as there was no significant difference in their performance. Also, school location had been found to determine students' performance. Filgona and Sababa (2017) and Erguven (2014) revealed that generally, females do better than males because the females get higher grades than the males. While researchers like, Emeka and Adegoke (2001) and Iroegbu (2000) affirmed that male students perform better than their female counterparts. Agbaje and Awodun (2013) and Abdulsalam-Nuhu (2011) concluded that both male and female students perform equally on achievement tests;.

Studies had been carried out on invariance like comparing the invariance of person parameter estimates based on CTT and IRT (Adedoyin 2010, Odili 2009, Nenty 2004, and Wiberg 2004); all the studies subjected the examinees to multiple choice tests. None of these studies to the best knowledge of the researcher dealt with Essay items and none of the literature reviewed have investigated on Invariance of item parameters of polytomous Geography items in Senior School Certificate Examination in Nigeria. This study is therefore, designed to cover the identified existing gap.

Statement of the Problem

Generally when test items are marked, the numbers of correct responses are counted to give the total score which is referred to as the ability. The scores obtained in WASSCE Physical Geography constitute part of the scores for grading students in geography which is used for certification. In most achievement tests, the difficulty levels of items are not usually considered when ability is estimated. This leads to variation in estimating the testees' ability, which shows lack of the principle of invariance. It is important that measurement instrument like test be invariant. This is to say that when a test is administered the characteristic of the items should be the same for all the examinees. For a test to achieve this stability there is a need for Item Response Theory (IRT). In order to take care of difficulty levels of items regardless of the testees that would respond to the items.

There have been series of studies in IRT. For example, Adewuni (2016) investigated Hierarchical cluster analysis of the dimensionality of Nigerian SSCE objective test in Government (2013-2014). The study employed Objective tests and was carried out in Nigeria. In the same vein, Akinboboye (2015) study employed Basic Education Certificate

Examination Mathematics multiple choice tests. The study was carried out in Nigeria. Similarly, Gadapa (2014) studied the invariance properties of IRT with 2pl and 3pl models. Multiple choice tests was adopted for the study and maximum likelihood estimation procedure was used with Bilog MG3. Also, Adedoyin (2010) study showed that IRT was invariant. This study employed the use of multiple choice tests in Mathematics at the Junior Secondary School in Botswana.

Furthermore, in Magno (2009) study, Multiple choice items were employed the study was carried out in the Philippines. In the same vein, Nkpone (2001) study employed multiple choice items in Physics Achievement Test for Senior Secondary Schools and was carried out in Nigeria. Adonu (2014) carried out a study on the Psychometric Analysis of WAEC and NECO Practical Physics Tests using Partial Credit Model. Essay items were used in the study which was carried out in Nigeria. Similarly, Ching-Fung (2002) carried out a study on Ability Estimation under different item parameterization and scoring models. The study employed the use of multiple choice item and essay items. The study was carried out in Texas. In the same vein, in 1993, Donoghue carried out a study using data from the 1991 field test of the NAEP Reading Assessment which was scored polytomously.

Most of these studies were on objective tests therefore the use of essay test would form the basis for this study to fill part of the gap.

Purpose of the Study

The purpose of this study was to investigate the invariance of Item Parameters of Polytomous Geography Items of Senior School Certificate Examination in Nigeria. Specifically, the study sought to: examine

(a) the examinees' ability estimates in SSCE Physical Geography;
(b) item difficulty estimates in SSCE Physical Geography;
(c) item discrimination estimates in SSCE Physical Geography;

(d) the extent of invariance of item difficulty estimates across male and female Geography students; and

(e) the extent of invariance of item difficulty estimates across Geography students in rural and urban areas.

Research Questions

The following research questions were raised to guide the conduct of the study:

(a) What were the examinees' ability estimates in physical Geography among senior secondary school Geography students in Nigeria?

(b) What were the item difficulty estimates in physical Geography among senior secondary school Geography students in Nigeria?

(c) What are the item discrimination estimates in physical Geography among senior secondary school Geography students in Nigeria?

(d) How invariant was the difficulty estimates of test items among male and female Geography students in Nigeria?

(e) How invariant was the difficulty estimates of test items among Geography students in rural and urban locations in Nigeria?

Research Hypotheses

The following hypotheses were formulated and tested in this study:

Ho_1: There is no significant invariance of item difficulty parameter estimates across different samples of male and female Geography students in Nigeria.

Ho_2: There is no significant invariance of item difficulty parameter estimates across different samples of Geography students in rural and urban school location in Nigeria.

Scope of the Study

The population for this study covered public senior school three students that offered Geography in Nigeria while the target population covered SSS3 Geography students in

Nigeria for the 2016/2017 academic session. Nigeria is divided into six geopolitical zones out of which three geopolitical zones were randomly selected. The randomly selected zones were South East, South West and North Central. From the selected zones, one state each was selected to make a total number of three states for the study. The three states were; Imo, Osun and Kwara.

There are 274 schools in Imo State, 255 schools in Osun State and 347 schools in Kwara State. A total number of 86 schools were sampled from all the three states using ten percent of the total number of schools from each state. All SSS 3 Geography students in each of the school selected constituted the sample for the study. This is because Geography is no more a compulsory subject, so very few students offer Geography in schools. A total number of 1546 geography students constituted the sample for the study. The instruments used were items in 2016 WASSCE Geography 3 (Practical and Physical Geography).

Operational Definition of Terms

The following terms were defined in the context of this work:

Physical Geography Achievement Test: These are WASSCE items selected from Practical and Physical Geography which consist of eight questions.

School Location: Examinees in rural and urban schools

Item parameter: this comprises of Item difficulty and item discrimination

Item difficulty: this describes where the item functions along the ability scale whether it is easy or hard.

Item discrimination: characteristic of an item to differentiate between examinees with high and low abilities.

Invariance: an item characteristic measuring ability equally across different samples of examinees or repeated at different time. This is what could be referred to as test stability.

Polytomous geography items: element of practical and physical geography.

Significance of the Study

The findings from the study might be useful to the classroom teachers, test developers, researchers, statisticians, statistical software developers, examination bodies, test users and the society. The results of the study might be useful to classroom teachers, as this will inform them on the possibility of the use of generalized partial credit model (GPCM) for the analysis of their polytomously scored items. This will go a long way to inform them about the advantages of using essay tests over multiple choice tests. The teachers would be informed on how to construct items that are sample free that do not depend on the ability of the respondents. The classroom teachers would be better informed as to what and how test should be constructed.

To the test developers, this study might inform them those items in test should not be scored the same hence the different items should be scored differently as examinees respond to different items with different ability and time. The advantages of polytomously scored tests would be learnt from the study.

The researchers might be enlightened on the use and advantage of IRT over CTT. This would serve as area for further research and researchers would find opportunities to think of researchable topics.

The examination bodies might be enlightened on the principle of invariance. And so, the results from this study would encourage examination bodies like WAEC to undertake rigorous item analysis before and after test administrations. To the guidance counselors, it might help them to identify the ability of the students so as to be able to place them in the appropriate group. It would help to expose the students' performance item by item and the possible reason for such performance for each item. The study would equip the guidance counselor with necessary information and data to diagnose students' strength and weaknesses since the study would have the data on their performance item by item.

For educational establishments, it might offer some explanation of examinees results through item by item response pattern for large scale testing purposes. The results of the study would serve as a tool for diagnosis of student's strengths and weaknesses by teachers and guidance counselors. The method of this study would involve identification of errors and factors/misconceptions leading to such errors. Hence, it would ensure improvement in the teacher's instructional strategies, coverage and practices.

The results of this study might help to establish the quality of examination conducted by WAEC. It would lead to formulations of more objective decisions in order to improve the quality of examinations. Examination bodies such as NECO and WAEC need the psychometric properties of test items in expressing the performance of the examinees. This might enable them to further improve upon test construction practices, administration and analysis. For this examination body, this study might give a clearer knowledge of IRT in the comparability of scores in practical examinations.

CHAPTER TWO

REVIEW OF RELATED LITERATURE

The review of relevant literature was carried out under the following sub-headings:

a. Classical Test Theory (CTT) and Item Response Theory (IRT);

b. Item Response Theory Models ;

c. Invariance in IRT;

d. Ability Estimation Methods in IRT;

e. Examinee Selected Item Model (ESIM);

f. Gender Differences in Geography Achievement;

g. Appraisal of the Literature Reviewed; and

h. Theoretical Background

Classical Test Theory and Item Response Theory

Classical Test Theory (CTT) had its origin from Charles Spearman's work on correlation coefficients. In his work, Spearman noted the existence of two types of correlations: a true correlation and an observed correlation (Courville, 2004). CTT is the approach that most readers have been exposed to throughout their education. In CTT person's answers to items in a test reflects the number of correct responses and it relates the individual's observed score to his or her location on the latent variable. Hambleton and Jones (1993) explained that examinees observed scores and true scores are not the same with ability scores. CTT item difficulty (p) and item discrimination (r) are so essential in the application of classical test model. This means that item discrimination indices will be high when there are heterogeneous examinee (examinees with different ability levels) samples, and will be low when there are homogeneous examinee samples. In practice, it is necessary that examinees from a single population be compared using results obtained from different test items as a result of having been administered different forms of the same test, or at least different subtests. This is difficult to do in CTT because it lacks the invariance principle. This dependency on the type of examinee that writes test in CTT limits the obtained scores in CTT. In practice this complicates any analyses. The reason for this dependency on test is due to the assumption of the CTT which cannot be measured directly. Furthermore, Wiberg (2004) explained that the true test score, T is the expected value of observed performance on the test of interest and is measured by the score achieved on a test but this score depends on more than the examinee's ability. Subsequently, Hambleton and Xing (2000) explained that because of the sample dependency, it makes comparison of the test result of the examinees difficult.

However, the shortcomings of CTT have led some measurement practitioners to opt for IRT. The reason for this change of emphasis by the psychometric and measurement community is to be able to solve measurement problems. These benefits include: (a) item statistics that are independent of the groups from which they were estimated; (b) scores describing examinee proficiency that is not dependent on test difficulty; (c) test models that provide a basis for matching test items to ability levels; and (d) test models that do not require strict parallel test for assessing reliability. Consequently, there are also benefits that can be derived through the application of CTT to measurement problems which includes: (a) the use of smaller sample sizes required for analyzes (a particular valuable advantage for field testing); (b) the use of simpler mathematical analysis when compared to item response theory; (c) a straight forward parameter estimation; and (d) CTT does not involve strict analysis.

Hambleton and Jones (1993) were of the opinion that in IRT, item analysis involves large sample size, complex mathematics and detection of items that do not fit the specified response model. The property of sample invariance in IRT means that test developers do not need a representative sample of the examinee population to calibrate test items. However, there is a need for heterogeneous and large examinee sample to insure proper item parameter estimate. Again, the test developer using IRT is faced with the problem of large sample sizes to obtain good item parameter estimate, the test developer must ensure that the examinee sample is of sufficient size to guarantee accurate item calibration.

The detection of poor items using item response theory is not as straight forward as the one done in CTT. This is because 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 that an adequate fit of model-to-data is essential for successful item analysis otherwise; items may appear poor as an object of poor model-fit (de Ayala, 2009; Hambleton, Swaminathan, & Rogers, 1991).

The main purpose of both CTT and IRT is determining numerical scores that estimate an individual's latent ability level. Although they have a common target, both primary theories of measurement differ significantly.

Similarly, Magno (2009) explained that CTT has a simple application procedure. In his study, a multiple choice test consisting of 70 chemistry items were administered to 219 students and the data collected were analysed using one parameter logistic model through conditional maximum likelihood. In another vain, Allen and Yen (2002) and Erguven (2014) explained that IRT is entirely independent of items. Also in 2014, Erguven's studied empirical evaluation and comparison of Classical Test Theory and Rasch Model. Findings from the study revealed that IRT gives more information than CTT.

However, Wiberg (2004) compared CTT with IRT in the Swedish driving school and employed 65 multiple-choice items to compare three IRT dichotomous models. A total of 5404 sample was used to examine invariance of item and person parameters. The study concluded that although IRT has some beneficial advantages over CTT, both theories are comparable.

Table 1: Main Differences Between Classical and Item Response Theories and Models

Area	Classical test theory	Item response theory
Model	Linear	Nonlinear
Level	Test	Item
Assumptions	Weak (i.e., easy to meet with test data)	Strong (i.e., more difficult to mee with test data)
Item-ability relationship	Not specified	Item characteristic functions
Ability	Test scores or estimated true score are reported on the test-score scale (or transformed test-score scale)	Ability scores are reported on the scale ∞ to +∞(or a transforme scale)
Invariance of item and person statistic	No-item and person parameters are sample dependent	Yes- item and person parameters are sample independent, if model fits the test data.
Item statistics	p, r	b, a, and c (for the three-paramete model) plus corresponding iter information functions.
Sample size (for item parameter e estimation)	200 to 500 (in general)	Depends on the IRT model but large samples, i.e., over 500, in general, ar needed.

Source: Hambleton and Jones (1993)

Although CTT has being in use in the measurement community for several years, the problems and limitations peculiar to this theoretical frame work made researchers to go in search of a better theory that would curb the problems and limitations of CTT. Hambleton and Jones (1993) noted that IRT is seen as a general arithmetical theory that examines the item and how performance relates to the abilities that are measured by the items in the test. De Ayala, (2009) described IRT as a system of models that creates one way of establishing the communication between underlying variables and their indicators which are the observed responses to items. Models incorporate a varying number of parameters, but one parameter all models have in common is the latent trait being measured, which is typically called the ability parameter and denoted by θ. IRT is a system of models that defines one way of establishing the correspondence between latent variables and their manifestations. It is not a theory in the traditional sense because it does not explain why a person provides a particular response to an item or how the person decides what to answer (de Ayala, 2009). Instead, IRT is like the theory of statistical estimation and is based on item by item analysis of a test. IRT can be discrete or continuous and can be dichotomously or polytomously scored; item score categories can be ordered or unordered; there can be one ability or many abilities underlying test performance. It links responses to ability, and item statistics are reported on the same scale as ability. The flexibility of IRT comes through knowing precisely where an item is doing its best. When test items are relatively difficult, then the test characteristics function explains how examinees with fixed ability can perform differently on two tests measuring the same ability, apart from the ubiquitous error scores. The test characteristic function connects ability scores in item response theory to true score in classical test theory because an examinee's expected test score at a given ability level is by definition the examinee's true score on that set of test items.

Also, if it is necessary to design a test with particular inherent characteristics for a specific examinee population, item response models permit the test developer to do that (Hambleton, Swaminathan, & Rogers, 1991). The need to build such tests is common for example, a test built to discriminate among less able students to select candidates for limited special-needs resources or a test built to discriminate among more-able students for the award of a scholarship or a test constructed for selection into tertiary institution like the post UTME constructed in Nigerian Universities. This is vital for certain modern testing applications, such as computerized adaptive testing.

Item Response Theory Models

There are two major models in IRT under which item response theory can be explained. Unidimensional and Multidimensional, unidimensional is measuring a single trait while multidimensional measures more than one trait. IRT can also be categorized on score responses polytomous model and dichotomous model.

Models for polytomous data: There are two major models for polytomous data which are Partial credit models and generalized credit models. **Partial credit models (PC)**; In partial credit model, when an item is given, there is an opportunity for the examinee to score part of the marks to be awarded to such item as it involves different steps to arrive at the answer to the item. For example 8+7/3=?

In the example above, the examinee is expected to carry out some mathematical operations. A scoring rubric for this item might be based on the assumptions that to correctly answer this item certain operations must be performed correctly and that it is not possible to correctly guess on these operations. The first operation is the addition of 8+7 and the second operation is the division of the number gotten from the first operation's result by 3. Zero points would be assigned for incorrectly performing the first operation. Because this item consists of two operations we can assign partial credit (e.g., 1 point) for correctly performing only the first operation and full credit for correctly performing the first and second operations (i.e., the 2 points reflects the number of correctly performed operations). We believe that if some credit is assigned for partially correct responses, the partially correct responses can provide useful information for estimating a person's location. One approach for modeling ordered (responses indicate more (or less) of what is being measured than other responses) polytomous data involves decomposing the responses into a series of ordered pairs of adjacent categories or category scores and then successively applying a dichotomous model to each pair Masters (1982).The amount of credit an answer is given is directly related to the degree of correctness of the response. As a result, higher credit indicates greater, say mathematics proficiency, than lower credit.

Generalized Partial credit (GPC): PC model states that all items on an instrument have equal discrimination but generalized partial credit (GPC) does not assume that there is equal discrimination. The GPC models contain a discrimination parameter that indicates the degree to which an item can differentiate among different values of θ. Muraki developed his model using, Masters' (1982) approach, response category over the previous one is governed by a dichotomous model. However, instead of using the dichotomous Rasch model as Masters did, Muraki used the 2PL model.

Models for dichotomous data: One parameter (Rasch model) logistic model; The 1PL and Rasch models require that items have a constant value for α, but allow the items to differ in their locations (difficulty). In this model, the discrimination parameter is fixed for all items. Some researchers see 1pl to be different from the Rasch model so discrimination α 1.0 is constant, whereas for the 1PL model the constant α does not have to be equal to 1.0 mathematically, the 1PL and the Rasch models are equivalent.

$$\rho(\theta)\,\frac{1}{1+e}-1(\theta-b)$$

Where θ= ability level

b = difficulty parameter

The values from one model can be transformed into the other by appropriate rescaling. The use of the Rasch model sets α to 1.0, and this constant value is absorbed into the metric used in defining the continuum. De Ayala (2009) explained that some researchers believed that, the Rasch model represents a different theoretical perspective than that embodied in the 1PL model. The 1PL model is focused on fitting the data as well as possible, given the model's constraints. In contrast, the Rasch model is a model used to build the variable of interest which is the standard by which one can create an instrument for measuring a variable (Wilson, 2005).

The two-parameter logistic model: This model contains item discrimination and item difficulty. It is represented with the formula below

$$p(\theta) = \frac{1}{1+e} - L = \frac{1}{1+e} - \alpha\,(\theta - b)$$

Where e= constant 2.718

b= difficulty parameter

$\alpha =$ discrimination parameter

The three-parameter logistic model: this model contains item discrimination, item difficulty and guessing. It is represented with the formula

$$P_i(\theta) = \gamma_i + \frac{1-\gamma_i}{1+\exp\{\alpha_i(\beta_i - \theta)\}}$$

3PL assumes that the examinee knows the correct answer of the item with probability equal to (3) or guesses the item correctly with probability. The 3PL model may be useful in applications other than educational testing. In many attitudinal surveys, there are items for which it makes sense to assume that all individuals have a probability that is bounded below by some non-zero number γ, regardless of the individual's ability.

We also have a four-parameter model (4PL), here 1-c in the 3PL is replaced by d- c. the 2PL uses b and a, the 3PL adds c, and the 4PL adds d.

Gadapa (2014) investigated into the invariance properties of item response theory with 2pl and 3pl models. The study showed that IRT items were invariant according to the group. Multiple choice tests was adopted for the study and maximum likelihood estimation procedure was used with Bilog MG3. Also, Nkpone (2001) utilized one and two parameter logistic models of IRT and also CTT in the development and standardization of physics achievement test for senior secondary students. The study revealed no significant difference exist among item parameter obtained using 1 PLM, 2 PLM and CTT in analysing the dichotomously scored physics achievement test.

Donoghue (1993) carried out a study using data from the 1991 field test of the NAEP Reading Assessment was scored polytomously while PARSCALE was used to analyze. Findings from the study showed that polytomous items yielded more information than dichotomous items.

Invariance in IRT

Invariance is the rudiment of item response theory which states that the examinee's ability with respect to the items used is stable. Invariance implies that, when an instrument like test is given to different groups of examinee say two, one should be able to say that the ability estimates of the two groups are the same. It is of interest to make comment about a particular assessment instrument for groups of examinees that share characteristics with another group that have being scored with the instrument to show the scores of different groups are comparable. In IRT, item parameters are invariant. Parameter indicates item parameters like discrimination, difficulty and guessing in the case of multiple choice test and examinee parameter in the case of theta level or ability estimates (θ) or test score.

Rupp and Zumbo (2004) described invariance to show that values of the parameters that are alike in different populations or across different conditions of interest, which is assessed when they are estimated repeatedly with different calibration samples. They explained further that, in measurement, one need to consider two different collections of population parameters because test data are the result of the intersection of item and examinee sample spaces and the model is the glue to bind the examinees and items together. When a test is utilized across subgroups of a population, measurement invariance is assumed. That is, it is assumed that the test measures the traits of interest in the same way across all groups. When this assumption is violated for a variable, the variable could be biased against one or more groups in a test. Meredith and Teresi and Borsboom (2006) noted that when measurement invariance is violated, then measurement is bias. When comparing means across groups, measurement invariance is a prerequisite for valid comparison of group differences because observed group mean differences (e.g. test scores) might not reflect the differences in latent factors (e.g. reading ability) if a certain type of measurement invariance does not hold. Similarly, Glockner-Rist and Hoijtink (2003) affirmed that the establishment of measurement invariance is particularly essential in assessing group difference on a measure. Adedoyin (2010) on the other hand, explained invariance to mean the variation following or preceding a distinguished phenomenon that does not differ, but share some properties with changes associated with similar phenomena. Invariance is a kind of connection that exists between two forms of mathematical objects, such that two parallel things correspond to one and the same object. The knowledge of invariance is the feature of being resistant to change under a set of alterations. From the scientific point of view, the concept of invariance has played a central role in the investigation of statements considered suitable to scientific laws. For example the concept of dimensional invariance has been widely used, through the method of dimensional analysis, in the search for lawful numerical relations among physical variables. Also in physics, invariants are usually quantities conserved (unchanged) and the physicist wants to be able to track what changes and what does not change under a set of transformation.

Horn and McArdle (1992) posited that measurement of invariance is whether or not, under different situations of observing and studying a phenomenon, measurement yields measures of the same attributes. If there is no evidence indicating presence or absence of measurement invariance then invariance does not exist, thus the basis for drawing scientific inference is severely lacking then differences between individuals and groups cannot be clearly interpreted. According to Nenty (2004), invariance is the base of objectivity in physical measurement, and the lack of it raises a lot of questions about the scientific nature of psychological measurement. Measurement which when repeated assigns significantly different values to the same characteristic of an individual, and for which such assignments depend on the particular set of items designed to measure the trait, cannot contribute to the growth of science or to the growth of objective knowledge in any area. Ojerinde (2013), established that invariance contains that the estimate of the parameters of an item across two or more groups of populations of interest that are different in their abilities must be the same. Also, measurement invariance is a necessary condition in the course of selection because fairness and equity cannot be achieved without measurement invariance. Meredith (1993) explained fairness as the equivalence of the conditional distribution of predictor over groups given any true score of an outcome variable. Equity exists when conditional distributions of an outcome variable are equivalent irrespective of group membership given any identical predictor score. If the measurement invariance does not hold, fairness and equity cannot exist across groups. Millsap and Kwok (2004) examined impact of measurement bias on the selection procedure and found that violation of measurement invariance could lead to unfair selection with regard to group membership.

Baker (2001) noted that when an examinee has an ability score of zero, this places him/her at the middle of the ability scale. This principle is the basis for a number of practical applications of the theory. In classical test theory when a certain test is given to a group of examinee, let say 10 items is given as a test to a group of examinee, after responding to the items, the test would be marked and scored. Scores like 2/10,3/10,5/10,8/10 etc. emerges from the test as the test result. In IRT, when a 10item test is given as in the one explained above for CTT. The test would be administered to a group of examinee after which the test is marked and scored. Here, the scores are estimated with maximum likelihood so that the ability is given on a scale of -3 to +3 so as to locate each examinee on the ability scale. In essence therefore, Gadapa (2014) investigated into invariance properties of item response theory with 2pl and 3pl models. The study showed that IRT items were invariant irrespective of the group while CTT items varied according to the group. Multiple choice tests was adopted for the study and maximum likelihood estimation procedure was used with Bilog MG3. Also the researcher is aware of Adedoyin, Nenty and Chilisa (2008) work titled investigating the invariance of item difficulty parameter estimates based on CTT and IRT in this study, objective test items were used for the study while multilog version 7.0 was used to estimate the IRT difficulty estimates while ANOVA was used to test for invariance. Findings from the study showed that IRT item difficulty parameter estimates were invariant across different independent samples of persons and across varying sample sizes of persons while CTT item difficulty parameter estimates across different independent samples of persons and across varying sample of persons were variant.

Ability Estimation Methods in IRT

In IRT, the aim of estimation is to obtain the parameter values which produce the curve that best fits the observed data. There are different methods used to estimate ability in IRT. Here are the methods used in ability estimation: Maximum likelihood procedure; and Bayesian modal estimation procedure.

Maximum Likelihood method (MLE): To score examinees with IRT, the Item Characteristic Curve (ICC) for a correct response and an incorrect response are multiplied to obtain a single function referred to as the likelihood function (LF). This represents the likelihood of each θ level given the responses of an examinee. For example, if an examinee has one correct answer and one incorrect answer to an item the ICCs are multiplied across θ to produce a symmetric graph in which the peak of the graph will serve as the maximum (Thompson, 2009).

A different location of the likelihood function leads to a different location of the maximum; this is how IRT provides precise θ estimates. There are three approaches commonly used to estimate θ within this framework. The major method of estimation is called maximum likelihood, called so because it simply finds the highest point on the likelihood function and returns the value of θ at which it occurs. In IRT, ability estimation in maximum likelihood estimation (MLE) is based on obtaining a value of estimated θ that maximizes the likelihood function or, equivalently the log likelihood function. De Ayala, (2009) and Hambleton, Swaminathan, & Rogers (1991), explained that parameter estimation techniques rely on maximizing a likelihood function.

It can be recall that in the 1PL-model, each item has a difficulty parameter, b, and all items are assumed to discriminate equally. Example Set a = 1 and omit it from the following calculations. Consider a test consisting of J = 5 dichotomously scored items with difficulties b1 = −1.9, b2 = −0.6, b3 = −0.25, b4 = 0.3, b5 = −0.45 and a 1PL-model. If this test were a mathematics test, item one would be considered an easy item, and items 2-5 would have an average difficulty level. Suppose an examinee has the response pattern (1, 1, 0, 0, 0); i.e., the examinee answered items 1 and 2 correctly and items 3-5 incorrectly. Ability parameter estimation is to determine which value of θ (the ability parameter) has the greatest likelihood of producing the response pattern (1, 1, 0, 0, 0). The estimation process begins with checking the chance of having the observed response to each item over a range of θ values. Maximizing the log likelihood function is equivalent to maximizing the likelihood function. De Ayala (2009) explained that, the log likelihood function was graphed, and a maximum was determined visually at ˆ θ = −0.85. In practice, a numerical approach such as Newton-Raphson is used to determine the maximum value.

This process is called Maximum Likelihood Estimation (MLE). One disadvantage of MLE is that examinees who answer all items correctly or incorrectly cannot have an ability estimate. In the first case, there are an infinite number of values above a certain threshold for θ, say above θ = 4, that would result in the same response pattern. There is no way of determining which value is most likely, and no value for ˆ θ is obtained. In other words, the log likelihood function has no unique maximum value. The situation is similar for an examinee who answers all items incorrectly.

However, MLE estimates also have several disadvantages. First, the bias of MLE estimates is positively correlated with θ on the ability scale (that is, outward bias). That is, MLE estimates are overestimated for high-level θs and underestimated for low-level θs (Kim & Nicewander, 1993). This bias is larger at negative θs than at positive θs. In addition, MLE might reach positive or negative infinity if the response pattern consists of all correct or incorrect item responses. Finally, the likelihood function might yield a number of local maxima (Thissen & Mislevy, 2000) and MLE may generate unreasonable values when an irregular item-response pattern is present.

Joint Maximum Likelihood Estimation (JML): this is a type of maximum likelihood estimation method where both persons and items parameters are estimated one after the other. It is straightforward, but has some problems associated with it and this limits its use. In JMLE one is simultaneously determining the item and person parameter that maximize the joint likelihood of the observed data.

Marginal Maximum Likelihood Estimation (MMLE): Estimates only item parameters. It is a method that addresses some of the difficulties of JMLE (de Ayala 2009). This process requires defining an approximate distribution for the population's ability parameters and as such, requires a high number of examinees. When item parameter estimates have been obtained and model-data fit is achieved, then one can proceed to estimate the person parameters using either MLE or a Bayesian approach. de Ayala (2009) stated that, in MMLE the items are considered fixed, but the persons are considered a random effect. It can be said that the MMLE is analogous to a mixed effects ANOVA model. The MMLE random factor is a mechanism for introducing population information. A popular method of obtaining marginal maximum likelihood estimates of parameters is the EM (Expectation-Maximization) algorithm. Many commercial programs, such as BILOG, PARSCALE, and Xcalibre (formerly MULTILOG), and the freely available program ICL obtain parameter estimates via some variation of this algorithm (Yen & Fitzpatrick, 2006). These estimates can be viewed as marginal maximum likelihood estimates, as one of the likelihood functions that this algorithm maximizes is derived from the marginal distribution of the observed responses (Hambleton & Jones, 1993). The EM algorithm uses both the observed data likelihood and the complete data likelihood.

Bayesian Estimation: Bayesian ability estimation methods differ from those of MLE in the sense that it is able to estimate ability for zero or perfect scores. This assumption would provide additional information to the person's response vector, about where we would expect the person to be located. That is, assuming that one had a normal population, the probability of observing persons located between –1 and 1 would be more specific than saying above 2. The essence of Bayesian method is that one has person location information in terms of a probability distribution prior to obtaining any observational data. This distribution is known as a prior distribution. Observational data is a distribution referred to as the posterior distribution, because it comes after collecting the observations. It is on the basis of the posterior distribution that we obtain the estimate of a person's location. Also the terms prior and posterior are relative (i.e. the posterior distribution can serve as the prior distribution in a second estimation cycle, and so on) (de Ayala, 2009).

Bayesian estimation procedure is a common variant of maximum likelihood, also called maximum a posteriori, or MAP, where this likelihood function is multiplied by an additional curve that represents an assumed population distribution. An additional variant is to take this Bayesian modified curve and rather than find the maximum point, find the average value as weighted by the function. This is referred to as Bayesian expectation a posteriori (EAP). Bayesian methods yield ability estimates that are a combination of the likelihood of getting an item correct or incorrect and the assumed distribution of a population's ability levels. Two Bayesian estimates are commonly applied in MML, i.e. maximum a posteriori (MAP) and expected a posteriori (EAP).

Maximum A posteriori (MAP): In maximum a posteriori (also known as Bayes Modal Estimate) one uses the mode of the posterior distribution as the θ. MAP approach is an iterative method like MLE and uses a continuous prior distribution. MAP θs exist for all response patterns and they suffer from greater regression toward the prior's mean than do the EAP estimates. EAP is noniterative and is based on numerical quadrature methods like those used in MMLE. This noniterative and efficient nature of EAP makes it faster than either MLE or MAP in estimating person locations. EAP uses a discrete prior distribution and the average squared error for EAP estimates over the population is less than that of MAP (as well as MLE) person location estimates. Also, the mathematics required for deriving the computational forms for person location estimation with any IRT model is simpler with EAP than with MAP.

Consequently, Maxwel, (2013) studied the use of Bayesian techniques with item response theory to analyse mathematics tests. The study compared the use of free software (Item Response Theory Command Language) to generally accepted commercial software (Xcalibre) in the analysis of College Algebra final exams. With both programs, a Bayesian approach was used. Bayes modal estimates were obtained for item parameters and EAP were obtained for ability parameters. Model-data fit analysis was conducted using two well-known chi-square fit statistics with no significant difference found in model-data fit. Findings from the study revealed that ability estimates from both programs were found to be nearly identical. Thus, when the assumptions of IRT are met, the freely available software program is an appropriate choice for the analysis of College Algebra final exams. Also, Glas (2006) study found that the ability estimates were similar to those obtained from MML estimate. Similarly, Norman (2010) found that Bayesian estimators are consistently biased on individual given ability respectively. Furthermore, Tsutakawa and Johnson (1988) studied on Ability Estimation via 3pl with partially known item parameters. In their study, the three parameter logistic model (3pl) was used to demonstrate ability estimation where data was collected from a 1987 American College Testing program (ACT) math Test. The results were compared with the more conventional approaches using maximum likelihood and empirical Bayes based on item parameters estimated by marginal maximum likelihood. The findings from the study was that when there is uncertainty in the item parameter, both maximum likelihood and empirical Bayes underestimate the variance of ability and therefore produce interval estimates which are too narrow and misleading.

Examinee Selected Item Model (ESIM)

In most achievement testing, examinees are required to complete all items. In some cases, examinees are allowed to respond to a certain number of items in a given set of items. This number of items is picked by the examinees and it can be called examinee-selected items (ESI). The test design is referred to as the examinee-selected Item (ESI) design. This is an IRT model in which examinees are required to respond to a specific number of items in a given set of items (e.g. responding to two items in five given items, leading to ten selection patterns) has the advantages of enhancing students learning motivation and reducing testing anxiety.

The ESI design is still commonly used in some high-stakes examinations because its boost motivation for learning. ESI design yields incomplete data (i.e. only the selected items are answered and the others having missing data). Although, it creates psychometric problems due to the selection effect but it has been demonstrated that missing data in unselected items are often MNAR (missingness not at random) which invalidates standard IRT models that hold the MAR (missingness at random). Wang, Wen, Jin and Qui (2012) proposed the examinee-selected item model (ESIM) to account for the selection effect in ES items in which numerous sources of selection effects were summarized by a latent variable (which was referred to as test wisdom). This is a way of tackling the selection effect in ESIM. Wang, et al (2012) suggested the use of Mplus and explained that this model requires the use of large sample size and a large number of ES items to yield reliable estimates for person parameters. It has been argued that missing data in the ESI design are missing not at random making normal item response theory (IRT) models to be inappropriate. Wang and Liu (2015) carried out an experimental study with 501 fifth graders who took two mandatory items and four pairs of mathematics (Dichotomous) items. In each pair of items, students were first asked to indicate which item they preferred to answer and then answer both items. In their study, it was observed that most students make a choice based on their perceived item easiness and familiarity.

Fitzpatrick and Yen (1995) noted that the selection behaviours differ across genders and grades or ethnic groups. This is because different groups of students have different degrees of test wisdom. Wainer and Thissen (1994) examined the score comparability in the ESI design and found the missing data to unselected items are missing not at random (MNAR) because the missing responses are related to examinees ability levels. Standard IRT models, which assume MAR, are no longer appropriate for ES items.

Wainer and Thissen (1994) noted that ESI may decrease test anxiety, raise motivation for learning, and encourage self-access and self-adaptive learning. An example is the senior secondary school certificate examination in Nigeria. Another example is the 2010 Geography, History and Mathematics subjects of the National Matriculation Entrance Tests in China, consisting of both mandatory and ES items. The ESI design brings challenges to psychometrics because allowing choice will inevitably cause problems in comparing scores obtained from different selection patterns because different items often (if not always) have different difficulties. Although IRT enables score comparison when different examinees respond to different items (e.g. computerized adaptive testing or large-scale assessment such as the program for international student Assessment, PISA), either the item difficulty has to be known in advance or the assumption of missing at random (MAR) has to be valid for those unselected items. Otherwise, score are no longer comparable when different examinees respond to different items.

Wang, Wainer, and Thissen (1995) conducted an experiment with the choose-one-Answer-All (COAA) design, in which examinees are required to indicate which item in a given pair of multiple choice items they prefer to respond and then answer both items. Such a design creates a complete data set with responses to all items, so the parameters of all items can be calibrated using conventional IRT models. They then treated the unselected items as missing data and evaluated the assumption of MAR in ES items. The results clearly point to MNAR, making standard IRT models inapplicable when the ESI design is implemented. Bridgeman, Morgan and Wang (1997) conducted the same experiment on constructed-response items and drew the same conclusion that the missing data to unselected items are MNAR. Bradlow and Thomas (1998) conducted simulation studies to demonstrate that the difficulties of easy items are systematically underestimated, whereas those for difficult items are systematically over estimated, which is an indication of MNAR. These experimental or stimulation studies lead to a conclusion that the item parameter equating is problematic and conventional IRT models are in appropriate for ES items.

It is of great interest to understand how examinees make a choice in the ESI design. Wang (1999) gathered empirical data from Hawaii and found that examinee’s perception of item difficulty is associated with their selection behaviour in ES items. Examinees tend to select easy and familiar items. Wainer and Thissen (1994) showed that the examinees do not always select items cleverly and less proficient examinees tend to do a worse job in selection than more proficient examinees. Allen, Holland and Thayer (2005) used logistic regression to assess the relationship between the target ability and selection behavior and found that the higher the ability level, the cleverer the choice is. There are some attempts to fitting conventional IRT models to empirical ES items. Lukhele, Wainer and Thissen (1993) used the graded responses model (GRM) to mandatory items and the nominal response model (NRM) to ES items in the chemistry subject of the college Boards' Advanced placement exams. No responses to unselected items were assumed to be MAR in their study. In reality, such knowledge is applicable only when missing data are observed (e.g. in the OCAA design) to sum up, it's very likely that nonresponses to unselected items are MNAR. If so, conventional; IRT models are no longer valid. Also Wang, Wen, Jin and Qui (2012) employed polytomous items which indicated that in practical situations ES items are polytomous (example essay items).

Gender Differences in Geography Achievement

Students' achievement in secondary school subjects is of interest to researchers, educationists and parents alike. Correlates of academic achievement in secondary school subjects in many disciplines have also been identified. The importance of considering the gender differences on geography achievement is based mainly on the socio-cultural variances between girls and boys (Abra, 1991). It was observed in our society that girls have been encouraged to go to school and be educated, while boys are expected to be active and dominant risk-takers. This view was supported by Hassan and Ogunyemi, (2008) that most boys are provided with toys that enhance their visual- spatial ability such as trucks, Legos (toys consisting of plastic building blocks and other components) and model. Spencer (2004) study also affirmed that the games of girls are often highly structured requiring turn taking and rules. Thus, social expectations and conformity pressures may create cultural blocks to girls. Fabunmi (2004) study discovered that gender composition has a significant relationship with students' academic performance and that gender composition has a significant influence on secondary school students' academic performance. There is therefore the need to find out if gender has effect on students' academic achievements in geography. Yusuf and Yakubu (2014) study revealed that female students perform better than male students in geography also stated that although there used to be more boys in geography class than girls but girls performed better than boys. Jacob and Linus (2017) studied effect of gender in Geography in Ganye Educational Zone using mastery learning strategy and conventional method for instruction. The quasi-experimental non-randomized pre-test, posttest control group design was used. The study used multi-stage random sampling technique at four levels to select 207 (120 Male and 87 Female) senior school two (SS II) students offering Geography in Ganye Educational Zone in Nigeria. A 40-item Geography Achievement Test (GAT) constructed by the researcher and validated by experts in Geography education was used to obtain data. Study revealed that female students exposed to mastery learning strategy performed better than male students exposed to mastery learning.

Oludipe (2012) conducted a study on Gender Difference in Basic Science using cooperative learning teaching strategy. A total number of one hundred and twenty (120) students from intact classes of three selected Junior Secondary Schools in three selected Local Government Areas of Ogun State, South-west Nigeria, were used in the study. The study employed a quasi-experimental design. Lesson note based on the jigsaw II cooperative learning strategy and Achievement Test for Basic Science Students (ATBSS) were the instruments used to collect the relevant data. The data collected was analyzed using descriptive and independent samples t-test statistical. It was observed that the academic achievement of male and female students at the pretest, posttest, and delayed posttest levels respectively were not significant. Falaye (2006) study did not reveal significant differences in female and male performances in Practical Geography Achievement Test (PGAT).

Okorie and Ezeh (2016) in their study Influence of Gender and Location on Students' Achievement in Chemical Bonding. Pre-test-post-test non-equivalent control group quasi-experimental design was used, with a population of 5,966 senior secondary class one (SS1) chemistry students in 57 senior secondary schools in Nsukka education zone of Enugu State, Nigeria. Purposive sampling technique was used to select 311 SS1 students from nine schools that were chosen. Intact classes were used. Instruments for data collection were Students' Interest Scale on Chemical Bonding (SISCB) and Chemical Bonding Achievement Test (CBAT). The instruments were validated and had reliability coefficients of 0.68 and 0.87 respectively. Regular chemistry teachers of the selected schools taught the students using instructional software package method of teaching (ISPMT). Result of the study showed that mean achievement score of female students was higher than that of result what tested with the hypothesis and the conclusion was that gender has no significant effect on the academic performance of male and female students in chemical bonding.

Appraisal of the Literature Reviewed

The literature reviewed was carried out on, classical test theory and IRT, Ability Estimation Methods in IRT, Models in IRT, and Invariance in IRT

IRT was seen by researchers such as Afolabi (2016), Ojerinde (2012), Wiberg (2011), Adedoyin (2010), Magno (2009), De Ayala (2009), Adedoyin, Nenty and Chillisa (2008), Nenty (2004), and Fan (1998) as a better measurement framework which gives the true ability of the examinee. This was because IRT is based on item by item analysis, which explains the ability of the testees and thus it is test independent. It curbs the disadvantages associated with CTT. In contrary, Courville (2004) and Fan (1998) reported that conclusions can still not be made that IRT is better that CTT because findings from their study were that IRT and CTT are comparable measurement framework in which one is not better than the other.

Classical test theory and Item response theory have been perceived as representing two different measurement frame works. Researchers such as Afolabi (2016), Ojerinde (2012), Wiberg (2011), Adedoyin (2010), Nenty (2004), Magno (2009) and Fan (1998) supported this claim. Studies have shown that IRT differs considerably from CTT in theory, and has some important theoretical advantages over CTT, as a result of this difference, it is expected that there would be differences between their item and person statistics. Fan (1998) and Courville (2004) were of different opinion as they state that both CTT and IRT are comparable in their item and person statistics. These differences are not entirely clear, so this study tends to establish if truly there is difference between IRT and CTT item and person parameter.

On ability estimation methods (maximum likelihood- joint maximum likelihood, conditional maximum likelihood and marginal maximum likelihood, Bayesian modal estimation, warm's weighted likelihood), it was evidence in the literature that maximum likelihood estimation had being the most frequently used estimation method. Researches like Awopeju and Afolabi (2016), Adonu (2014), Gadapa (2014), Magno (2009), Wiberg (2004), Courvlle (2004) used maximum likelihood estimation methods this was because their studies were on dichotomous test while researcher like Maxwel (2013) used the Bayesian ability estimation method. Since most of the studies on IRT dealt with objective items, this study is important because essay test was used and Bayesian ability estimation method was employed which formed part of the empirical studies in IRT.

As regards invariance, it was discovered from the literature that researchers such as Afolabi (2016), Ojerinde (2012), Adedoyin (2010) and Nenty (2009), were of the opinion that IRT has invariance across different subgroups. While researchers like Courville (2004), Lawson and Macdonald and Paunone (2002) Fan (1998), Stage (1998a), (1998b) and (1999), disagreed that CTT is also invariant across different subgroups. This shows that there is no agreement among the researchers as to whether CTT is invariant across different subgroups or not. This disagreement shows that there is need for more investigation regarding invariance. This necessitated this study using gender and school location variables.

From the literature reviewed, IRT has two models, which are dichotomous and polytomous. The literature reviewed shows that most researches were carried out using dichotomous models. Researchers like Afolabi (2016), Erguven (2014), Adegoke (2013), Ojerinde (2013), Adedoyin (2010), Magno (2009), Courville (2004), Wiberg (2004), and Npkone (2001) employed the use of dichotomous models in their studies while only few like Adonu (2014), Kim and Lee (2006), Ching-Fung (2002) and Donoghue (1994), employed polytomous model. Furthermore, those that employed polytomous model were not in geography; in Nigeria. This also constitute a gap, part of which this research work intends to fill. This research work is therefore designed to investigate the invariance of item parameters of polytomous Geography items in senior school certificate examination in Nigeria. Since there are limited studies on polytomous models in Nigeria, this creates a gap for further study.

Consequently, on the issue of gender difference researchers such as, Jacob and Linus (2017), Erguven (2014), Voyer and Voyer (2014), Yusuf and Yakubu (2013), revealed that generally, females do better than males because the females get higher grades than the males. while researchers like, Jacobs (2002) Emeka and Adegoke (2001), Iroegbu (2000), Kahle and Lakes (2003), affirmed that male students perform better than the female students on the other hand some researchers were of the opinion that both male and female students perform equally on achievement tests Agbaje and Awodun (2013), Abdulsalam-Nuhu (2011) .However, findings were inconclusive as to gender difference on achievement tests because researchers like Ezewu (1981), Adelowokan (1989), Adesoji (1991) and Akinbole (1991) contend that gender does not have impact on students' achievement. This study therefore seeks to find out more on gender differences in geography as few studies were carried out in Geography.

On school location, some studies like Bosede (2010), Akpan (2001) and Olonade (2000) reported that school location have influence on students' academic performance while Alokan 2010, Considine and Zappala (2002) were of the opinion that school location does not have influence on students' academic achievement. Ogunkola and Fayombo (2009), noted that urban students performed better than their rural counterparts while some researchers like Mberekpe, (2013) found that rural students performed better than their urban counterparts while Agbaje and Awodun (2013), were of the opinion that neither rural nor urban students can claim to be superior to one another as there were no significant difference in their performance. This also creates gap for further study.

Theoretical Background

There are two theories guiding this study, these are CTT and IRT.

Classical Test Theory has relatively weak assumptions which make it easy to apply in many testing situation (Hambleton et al, 1991). CTT is majorly focused on test-level, it also assumes that all items in a test are the same in terms of the item statistics (difficulty, discrimination and guessing).

Assumptions of CTT

CTT is based on a set of assumptions. These assumptions are that, in the population,

(1) The errors are uncorrelated with the trait scores for an instrument,

(2) The errors on one instrument are uncorrelated with the trait scores on a different instrument, and

(3) The errors on one instrument are uncorrelated with the error scores on a different instrument.

These assumptions are considered to be weak assumptions because they are likely to be met by any given data and because the items in CTT test and the ability are sample dependent. If a test constructed in CTT is simple then that means that the low ability student will pass alongside the higher ability students but if a test is hard that means that the higher ability students will pass and the lower ability will fail. This is not the case in IRT because simple test may be failed by higher ability students while hard test may be pass by lower ability students thus test here is not sample dependent.

Assumptions of IRT: Hambleton, Swaminathan & Rogers (1991), opined that, unlike Classical Test Theory, IRT has strong assumptions. For a unidimensional model, one with single dominant factor influencing these assumptions are:

- Unidimensional a single trait is measured θ;
- when the assumption of unidimensionality holds, the responses to a pair of items are statistically independent (this assumption is known as local independence);
- responses are shown on Item Response Function, or IRF;
- invariant across differing tests and differing sample populations; and
- estimates of ability parameters differ only to measurement error.

In the case of a typical final exam in a college setting, there are other factors that are involved: motivation, test anxiety, general cognitive skills, the ability to work quickly, etc., but these factors are assumed to play a relatively minor role. To say that the responses to a pair of items is statistically independent means that after taking into account the performance level of the examinee, the responses to the different items are in no way related. In other words, the responses are conditionally independent given the ability level.

Some assessments are not one dimensional. A clinician may use an instrument that measures a person's ability to adopt and adhere to a healthy diet. This ability is affected by the person's actual knowledge (a cognitive ability) of healthy food choices and the person's motivation to adopt such a diet (an affective dimension). In this case, the two dimensions may not interact. In contrast, a complicated word problem in mathematics may be two-dimensional: one factor is reading comprehension and the other is mathematical ability. In this case, the two factors do interact. Multidimensional models are treated as either non compensatory or compensatory, based on the interaction of the abilities being measured (de Ayale, 2009). Multidimensional IRT (MIRT) models are becoming more widely used, for example. The assumption of local independence can hold even when unidimensionality does not, provided the dominant factors are specified and can be conditioned on (Hambleton, Swaminathan & Rogers, 1991).

For the purpose of this study, the Item Response Theory was used (IRT). For any research on IRT, there should be model fit assessment and dimensionality assessment.

Dimensionality Assessment: In assessing dimensionality the interest is to know the number of dimensions that the test items will measure i.e. whether unidimension or multidimensions or the number of content-oriented factors. The traditional approach for assessing dimensionality involves the factor analysis of a correlation matrix. Whenever one factor analyzes a correlation matrix derived from binary data, there is a possibility of obtaining artifactual factor(s) that are related to the nonlinearity between the items and the common factors. These factors of curvilinearity have sometimes been referred to as difficulty factors and are not considered to be content-oriented factors (Ferguson, 1941; McDonald, 1967; McDonald & Ahlawat, 1974; Thurstone, 1938). To avoid extracting these difficulty factors, McDonald (1967) suggests the use of nonlinear factor analysis. Because our data are dichotomous, we use this nonlinear approach for our dimensionality analysis. This nonlinear strategy is implemented in the program NOHARM (Fraser, 1988; Fraser & McDonald, 2003). An alternative to using NOHARM is to use TESTFACT (Wood et al., 2003) and perform full information factor analysis (Muraki & Engelhard, 1985; Wood et al., 2003). We selected NOHARM over TESTFACT on the basis of research showing that NOHARM performs well in dimensionality recovery studies (e.g., De Champlain & Gessaroli, 1998; Finch & Habing, 2005; Knol & Berger, 1991) and because it is available at no cost. For additional information on approaches for assessing unidimensionality NOHARM (Normal Ogive Harmonic Analysis Robust Method) is a general program that takes advantage of the relationship between nonlinear factor analysis and the normal ogive model in order to fit unidimensional and multidimensional normal ogive models

In a multidimensional IRT, one of its assumptions is dimensionality assumption. This assumption states that the observations on the visible variables are a function of a set of continuous latent person variables. As is the case with the unidimensional models, the proper application of a MIRT model involves dimensionality assessment to determine the number of latent variables in the set. For instance, if the true state of nature is that the item responses are a function of three latent variables, then the dimensionality assessment should facilitate correctly specifying that the model have three latent person variables and not, for example, two. As is the case with the unidimensional models, violation of the dimensionality assumption is a matter of degree, and whether the resulting θs are useful and psychologically meaningful is a validity question.

Model-Data fit Assessment

This is an assessment that allows the researcher to know the exact model to choose for the data collected. BIGSTEPS (and WINSTEPS) are used for model-data fit assessment and it produces two fit statistics, INFIT and OUTFIT, for examining model-data fit. INFIT and OUTFIT provide information concerning differences in responses, depending on whether the discrepancies occur close to or farther away from the estimated parameter. These fit statistics are calculated for both persons and items at the individual person and item levels. Therefore, each individual and item has both an INFIT and OUTFIT statistic. For persons the statistics' calculations involve a sum across items, whereas for items the sum is across people. INFIT is a weighted fit statistic based on the squared standardized residual between what is observed and what would be expected on the basis of the model (i.e., a chi-square-like statistic). These squared standardized residuals are information weighted and then summed across observations (i.e., items or people); the weight is pj(1 – pj). These chi-square statistics are averaged to produce the INFIT mean-square statistic; the mean square is labeled MNSQ in the output. OUTFIT is also based on the squared standardized residual between what is observed and what would be expected, but the squared standardized residual is not weighted when summed across observations (i.e., items or people). As such, OUTFIT is an unweighted standardized fit statistic. As is the case with the INFIT statistic, the OUTFIT statistic is transformed to a mean square and labeled MNSQ in the output. These two statistics differ in their sensitivity to where the discrepancy between what is observed and what is expected occurs. For instance, and from a person fit perspective, responses on items located near the person's ˆ θ that are in line with what would be expected produce INFIT values close to 1 (given the stochastic nature of the model). However, responses on items located near the person's $\hat{\theta}$ that are not in line with what would be expected lead to large INFIT values. That is, INFIT is sensitive to unexpected responses near the person's $\hat{\theta}$. In contrast, OUTFIT has a value close to its expected value of 1 when responses on items located away from a person's $\hat{\theta}$ are consistent with what is predicted by the model (again, given the stochastic nature of the model). However, unexpected responses on items located away from a person's $\hat{\theta}$ (i.e., outlier responses) lead to OUT- FIT values substantially greater than 1. That is, OUTFIT is sensitive to, say, a high-ability person incorrectly responding to an easy item or a low-ability person correctly responding to a hard item. One has an analogous interpretation for these fit statistics when used for item fit analysis. The range of INFIT and OUTFIT is 0 to infinity with an expectation of 1; their distributions are positively skewed. Values that are above or below 1 indicate different types of misfit. For example, values substantially less than 1 may be indicative of dependency or over fit, whereas values substantially greater than 1 may reflect noise in the data. Although there are various interpretation guidelines, one guideline states that values from 0.5 to 1.5 are okay with values greater than 2 warranting closer inspection of the associated person or item. Smith, Schumacker, and Bush (1998) stated that using a common cutoff value does not necessarily result in correct Type I error rates. To take sample size into account when interpreting INFIT and OUTFIT by using $1 \pm 2/N$ and $1 \pm 6/N$ as cutoff values, respectively. Given INFIT's and OUTFIT's expectations and their range, it is clear that there is an asymmetry in their scales. Therefore, INFIT and OUTFIT are transformed to have a scale that is symmetric about 0.0. The result of this transformation is a standardized (0, 1) fit statistic, ZSTD. Such ZSTDs are obtained by using a cube root transformation of the MNSQs to make them normally distributed and to have a range from $-\infty$ to ∞. Good fit is indicated by INFIT ZSTD and OUTFIT ZSTD values close to 0. Because the ZSTDs are approximate t statistics, as sample size increases these t statistics approach z statistics. As such, values of +2 are sometimes used for identifying items or people that warrant further inspection; for inferential testing the null hypothesis is perfect model–data fit. In our output the standardized INFIT statistic is labeled INFIT ZSTD and the standardized OUTFIT statistic is labeled OUTFIT ZSTD. INFIT ZSTD and OUTFIT ZSTD correspond to the standardized INFIT MNSQ and OUTFIT MNSQ, respectively.

In carrying out study using IRT, the following steps need to be taken:

1. Check whether the assumptions of IRT hold if it does, check if the model chosen fit the data collected and also check if the data collected is unidimension or multidimension this is dimensionality assessment.
2. Determine whether you are using an essay test or an objective test if it is objective test, you can use any of the three dichotomous models (1pl, 2pl or 3pl) and if it is essay you can use any of the polytomous models (Nominal Categories Model (NCM), Partial credit model (PCM) ,Generalized partial credit model (GPCM) and Multiple-choice model)
3. Proceed to analysis using the IRT software like BILOG MG, MULTILOG or IRTPRO. BILOG MG this is used to analyse dichotomous tests. MULTILOG this software is used for polytomous tests and IRTPRO is software used for both dichotomous and polytomous tests and this can be used for tests of any dimension. When this is done the different ability estimation method can be determined.

This present study does not follow the normal IRT procedures because the test items were selected, thus, an examinee selected item model was used. Therefore, the study did not carry out dimensionality assessment and fit statistics.

CHAPTER THREE

RESEARCH METHODOLOGY

This chapter focuses on the method that was employed in carrying out this study. It was presented under the following sub-headings:

a. Research Design;
b. Population Sample and Sampling Techniques;
c. Instrumentation;
d. Procedure for Data Collection; and
e. Data Analysis Techniques

Research Design

The design for this study was a descriptive survey design. Check and Schutt (2010) explained survey research as an efficient method for systematically collecting data from a wide-range of individuals in educational settings. It is appropriate for this study because sample was selected from a large population to give answers to test items and their responses were analyzed to make generalization.

Population Sample and Sampling Techniques

The population of this study was all public senior secondary school students in Nigeria while the target population was all senior secondary school three students that offered Geography in Nigeria in 2016/2017 academic session. The target population for the study was 14,718 in Imo State, 8,433 in Osun State and 9,348 in Kwara State for 2061/2017 academic session. This information was collected from WAEC.

Multistage sampling procedure was used for the selection of the sample (respondents) at different stages of sampling. At the first stage, Nigeria was stratified into 6 geo-political zones (that is North East; North West; North Central; South West; South East; and South

South). At the second stage, simple random sampling technique was used to select three geo-political zones from the existing six geo-political zones in Nigeria. The three geo-political zones selected were South East, South West and North Central. At the third stage, simple random sampling technique was used to select a state each from the three geopolitical zones chosen to make a total of three states as sample for the study The three states selected were; Imo, Osun and Kwara.

There are 274 public senior secondary schools in Imo State, 255 public senior secondary schools in Osun State and 347 public senior secondary schools in Kwara State. Ten percent of the schools in each state were sampled using simple random sample. A total number of 86 schools were sampled and all SSS 3 geography students in each of the school selected constituted the sample for the study. This is because geography is no more a compulsory subject so very few students offer geography in schools. A total number of 1546 geography students made the sample size for the study.

Instrumentation

The instrument used for data collection was Essay items of Physical Geography which was selected from WASSCE 2016 Geography Paper 3 (Element of Practical and Physical Geography). The instrument, Physical Geography Achievement Test (PGAT) consisted of eight items and the students were instructed to answer questions one and any other three making a total of four questions in all. This method is referred to as Examinee Selected Response Model (ESIM), this is because the examinees were given the privilege of choice. Bradlow and Thomas (1998) said that this method gives the examinee the opportunity to answer items on a wide range of topics and also limit the time spent to complete the assessment process. Instrument for this study has the following component: Question 1 was on Map reading. Question 2 was on data representation three sub-items. Question 3 was on

Rocks which has two sub-items. Question 4 was on Landforms which has three sub-items. Question 5 was on Environmental resources (Water) which are made up of three sub-items. Question 6 was on Weather and Climate with three sub-items. Question 7 was on Solar System it has three sub-items. Question 8 was on Desert Landforms which has two sub-items. In all the total number of items was 23 items. The scripts were collected and marked by the researcher guided by the WAEC final marking rubric. The responses of the test were scored polytomously with the use of Generalized Partial Credit Model of IRT. The ability estimates was carried out using Bayesian method and the parameter estimates was estimated with Robust Maximum likelihood in Mplus software.

The validity and reliability of the instrument have been ascertained by WAEC. However, it must be noted that part this study is to ascertain the psychometric properties of the items to study the quality and acceptability of these items.

Procedures for Data Collection

The researcher collected a letter of introduction from the Head of Department, Department of Social Sciences Education Faculty of Education, University of Ilorin, Ilorin, Nigeria to seek for the permission of the selected school authorities (the principals and teachers) so that the students in the school would be allowed to respond to the instrument Physical Geography Achievement Test (PGAT). The instrument was administered with the help of the geography teachers in the selected schools. The scripts were collected from the students after the required time for the test and were later marked with the use of marking scheme. The questions were scored polytomously using the generalized partial credit model (GPCM)

Ethical Consideration: The researcher collected a letter of introduction from the Head of Department of Social Science Education, University of Ilorin, Ilorin to obtain information about the schools in the three selected states: Imo, Osun and Kwara States. The researcher took the letter to the Ministry of Education in the selected states.

In the administration of the test, the researcher collected a letter from the head of department to seek for the permission of the school authority (the principals and teachers) selected so that the students in the school were allowed to respond to the test. These question papers were administered with the help of the geography teachers in the schools. The researcher seeks approval of the Principals of the Schools selected to administer instrument to the students. The respondents (students of geography) were duly informed to seek for their consent before the administration of the test. All the information gathered were handled with utmost confidentiality. They were all informed that information gathered would be handled with utmost confidentiality.

Anonymity was maintained as the identity of the students who responded to the test items was not exposed. This was done by not using their names in the study. Also, plagiarism was addressed as all cited works were referenced. At the end of the study, the work would be published for all stakeholders to benefit from.

Data Analysis Techniques

Data collected were scored polytomously, using GPCM (in GPCM, the amount of credit an answer is given is directly related to the degree of correctness of the response for example a 2marks using GPCM will be 0, 1, 2 that is category score of 0, category score of 1 and category score of 2). Descriptive statistics (simple percentage, bar-chart, mean and median), Mplus version 7.4 with Robust maximum likelihood estimation for parameter estimates and Bayesian ability estimation method were used to answer all the research questions while t-test statistics was used to test the hypotheses at 0.05 level of significance.

CHAPTER FOUR

DATA ANALYSIS AND RESULTS

In this chapter, the analysis and results of the data collected for the study are presented in tables and figures for easy interpretation. The chapter begins with a presentation of the descriptive statistics of the senior secondary school (SSS) students (respondents) that participated in the study. Percentage was used to describe demographic data of the students (respondents) and this was followed by results. The scores obtained from Physical Geography Achievement Test (PGAT) administered to sampled students that offered Geography in 2016/2017 session in three different states from three geopolitical zones in Nigeria. Research question one was answered using the generalized partial credit model with Mplus 7.4 with the use of Bayesian method of ability estimation and descriptive statistics. Research question two, three and four were answered using the Robust maximum likelihood (MLR) estimate while the hypotheses were tested with the use of t- test.

Background characteristics of Students

Table 2: Distribution of Students on the Basis of Gender

Gender	**Frequency**	**Percentage (%)**
Male	758	49.03
Female	788	50.97
Total	1546	100

Table 2 shows that out of 1,546 students sampled for this study 758(49.03%) were males and 788 (50.97%) were females.

Table 3: Distribution of Students on the Basis of School Location

School location	**Frequency**	**Percentage (%)**
Rural	651	42.11
Urban	895	57.89
Total	1546	100

Table 3 shows that out of 1,546 students sampled for this study, 651 (42.11%) were from schools located in rural areas and 895 (57.89 %) were from schools located in urban areas. This implies that students from urban areas were more than the students in rural areas.

Answering Research Questions

Research Question 1: What were the examinees' ability estimates in Physical Geography Achievement Tests?

In order to answer this research question, the responses of the examinees to the Physical Geography Achievement test items were collated and analysed using generalized partial credit model with Mplus version 7.4. The estimation method used was the Bayesian Expected Posteriori (EAP). In the course of the estimation, the metric is set by defining the theta distribution to have a mean of 0 and standard deviation of 1.The results are presented in Table 4.

Table 4: Examinees' ability estimates in Physical Geography Achievement test

S/N	Ability	S/N	Ability	S/N	Ability	S/N	Ability
1	1.284	25	-0.196	1499	-0.783	1523	-1.088
2	1.305	26	-1.245	1500	-2.004	1524	-0.277
3	1.353	27	0.211	1501	-1.48	1525	-1.234
4	1.304	28	-0.197	1502	-1.32	1526	-0.401
5	1.881	29	-0.116	1503	-1.19	1527	-1.332
6	0.175	30	0.219	1504	-1.19	1528	-0.903
7	1.919	31	0.208	1505	-1.324	1529	0.078
8	1.629	32	0.196	1506	-1.17	1530	-1.147
9	1.484	33	0.218	1507	-0.956	1531	-1.088
10	0.955	34	0.221	1508	0.19	1532	0.08
11	0.218	35	0.047	1509	-1.332	1533	0.08
12	0.236	36	0.471	1510	-0.903	1534	-1.635
13	0.212	37	-0.742	1511	0.19	1535	-1.514
14	0.236	38	-0.014	1512	0.129	1536	-1.719
15	0.236	39	-0.075	1513	-0.454	1537	-1.388
16	0.236	40	0.104	1514	0.15	1538	-0.632
17	0.236	41	0.104	1515	0.15	1539	0.08
18	0.228	42	-0.014	1516	0.15	1540	-1.719
19	0.204	43	0.034	1517	0.19	1541	-0.632
20	0.229	44	0.013	1518	0.15	1542	-0.723
21	0.219	45	-0.507	1519	0.953	1543	-1.388
22	-0.176	46	-0.007	1520	0.08	1544	-1.088
23	0.236	47	0.103	1521	-0.454	1545	0.081
24	0.042	+ +	+ +	1522	-1.332	1546	0.08

Table 4 shows the abridged ability estimates of the examinees in the Physical Geography Achievement Test (see Appendix V for the full version). These estimates range between -3 to +3, with negative estimates (-3 to 0) showing low ability estimate and positive estimates (0 to 3) showing moderate to high ability estimates. The distribution of the examinees' ability estimates in the test is presented in the Figure 1.

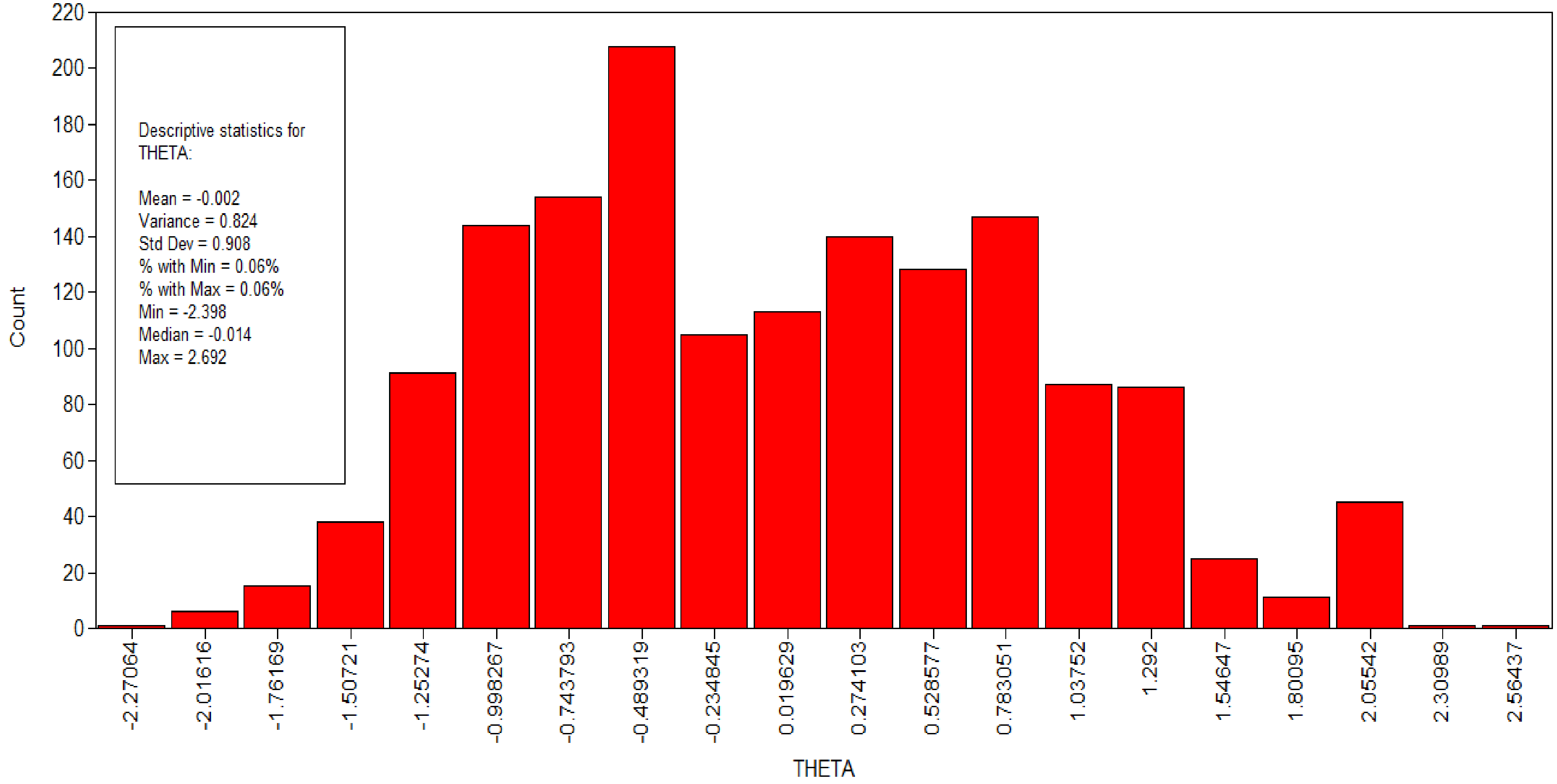


Figure 1: Distribution of abilities in Physical Geography Achievement Test

Figure 1 shows that the mean ability estimate in the test was -0.002. While 0.06% of the examinees who took the test had the least ability estimate (-2.398), 0.06% obtained the highest ability estimate (2.692). Also, the Figure shows that most (about 210 examinees) representing 13.58% of the examinees obtained ability score of -0.489. In all, 763 examinees representing 49.35% of the examinees who sat for the physical Geography Achievement Test obtained ability score below 0 (zero), while 50.65% of the examinees obtained ability score above 0. This was because for comparison, the average was set as 0 which implies that any score below 0 is poor while scores above 0 is a good performance. This implies that 50.65% performed well while the remaining 49.35% performed poorly.

Research Question 2: What were the item difficulty estimates of the Physical Geography Test?

To answer this research question, examinees' responses to the test items were collated and analysed using generalized partial credit model with Mplus 7.4. The estimation method used was the Robust Maximum Marginal Likelihood estimation (MLR). In all, there were eight broad questions out of which each examinee is expected to answer four. Because the privilege of choice which the examinees were given, missingness were not at random and traditional IRT which holds the assumption of missingness at random could not be used. As a result of this, the Examinee-selected item model (ESIM) was used in the estimation of the test items parameters. Although, there were eight questions in the test, each of the questions is made up of two or more test items. Consequently, there were 23 items found to make up the test. The step difficulty parameters and the overall difficulty parameters of the items are presented in Table 5

Table 5: Item Step Difficulties of Physical Geography Test

	Step difficulty									
Item	1	2	3	4	5	6	7	8	9	difficulty
1A	2.041	-1.855	1.556	-0.613	0.524					0.363
1B	1.619	-0.259	0.64	0.696	0.55	0.755	0.879	0.807	1.57	0.862
1C	4.327	-2.344	6.793	-3.219	2.451					1.966
1D	1.949	-0.838	2.453	-0.365						0.877
2A	7.481	1.256	-2.335	-7.778	7.174	0.313	1.651	-3.897	5.787	5.042
2B	7.427	0.94	2.86	-4.296						62.366
2C	-7.279	8.734								-1.591
3A	-1.171	-1.49	-0.984	1.846	0.956					0.403
3B	3.459	-2.02	2.037	-0.507	3.188	2.615				3.251
4A	-6.399	-2.105	-1.098	0.596	2.288	3				2.61
4B	3.507	-1.611	3.822	-0.614	0.353					2.103
4C	-0.79	0.695	2.526	-5.44						-2.011
5AI	3.837	2.626	0.013							2.631
5AII	-2.049	-3.692	-0.726	3.426	1.633	-3.903				-1.667
5B	4.601	-0.03	1.947	-1.523	4.182	1.527				3.475
6A	1.409	-4.07								-1.987
6B	1.467	-4.19	-1.25	-1.308	0.664					-1.187
6C	8.228	-1.655	0.393	-3.815	6.232	0.091	3.038	-2.466		3.149
7AI	-1.068	-1.689								-1.898
7AII	0.461	1.551	0.039	-2.823	-7.342					-2.371
7B	3.826	-3.038	1.297	-2.718	2.779	-4.135	5.681	-5.445		-0.572
8A	-1.037	-0.078	0.967	1.198	1.508	1.9				-0.484
8B	1.238	1.251	1.652	1.908	2.269	1.802	2.477	3.214		10.211

Table 5 shows the items' step difficulties or transition location parameters (is the point of transition from one category (score) to the next) of physical Geography test. The step difficulties indicate the point on the ability metric at which the probability of responding in two adjacent categories is equal (de Ayala, 2009) or point of intersection of two adjacent categories on the ability scale. The categories represent the arrays of scores obtained on the items. For items 1B and 2A, there were a total of 10 scores obtained by the examinees. They are: 0, 1, 2, 3, 4, 5, 6, 7, 8 and 9. For items 6C, 7B and 8B there were a total of 9 scores obtained by the examinees. The scores are: 0, 1, 2, 3, 4, 5, 6, 7 and 8. For items, 4A, 5AII, 5B and 8A, there were 7 total scores obtained. The scores are: 0, 1, 2, 3, 4, 5 and 6. For items 1A, IC, 3A, 4B, 6B, and 7AII, there were a total of 6 scores. The obtained scores are: 0, 1, 2, 3, 4 and 5. For items 1D, 2B, 3B and 4C, a total of 5 scores were obtained by the examinees on the test. The obtained scores include: 0, 1, 2, 3and 4. For item 5AI there were a total of 4 scores obtained by the examinees. These scores are: 0, 1, 2 and 3. And for items 2C, 6A, and 7AI, there a total of 3 scores: 0, 1 and 2 which were obtained by the examinees.

Furthermore, the table shows that item 1 (i.e., 1A), have five step difficulties. They are:$b_{0\,\&\,1} = 2.041, b_{1\&2} = -1.855,\ \ b_{2\,\&3} = 1.556,\ b_{3\,\&\,4} = -0.613,$ and $b_{4\,\&\,5} = 0.524$. Step difficulty 1 (i.e.,$b_{0\,\&\,1}$), showed that the transition point for examinees to have a score of 0 or a score of 1 was 2.041. This implies that examinees with ability 2.041 or greater would be able to move from score of 0 to 1. For step difficulty $b_{1\&2} = -1.855$ showed that the transition location at -1.855 is the transition between a category score of 1 and score of 2. This means that examinees with ability -1.855 or more would be able to move from a score of 1 to 2. Also, for step difficulty $b_{2\,\&3} = 1.556$ showed that the difficulty of examinees to move from a score of 2 to a score of 3 was 1.556 meaning that examinees with ability 1.556 or greater would be able to move from a score of 2 to a score of 3. Step difficulty $b_{3\,\&\,4} = -0.613$ showed that the difficulty of examinees to move from a score of 3 to a score of 4 was

-0.613 which implies that examinees with ability -0.613 or more would be able to move from a score of 3 to a score of 4. Step difficulty $b_{4\,\&\,5} = 0.524$ which showed that the difficulty of examinees to move from a score of 4 to a score of 5 was 0.524 this implies that examinees with ability 0.524 or greater would be able to move from a score of 4 to a score of 5. The result for this item showed that the transition location parameters were not in sequential order. For example, the probability of obtaining a score of 1 is substantially greater than that of obtaining a score of 2, but the probability of obtaining a score of 3 is greater than that of obtaining a score of 2. This showed that the item is not functioning as expected. As it was observed that examinees on higher ability level had low category scores while examinees with low ability had high category scores. See appendix VI for the category response curve for item 1A

Also Table 5 shows that item 2 (i.e., 1B), have nine step difficulties. They are: $b_{0\,\&\,1} = 1.619$, $b_{1\&2} = -0.259$, $b_{2\,\&3} = 0.64$, $b_{3\,\&\,4} = 0.696$, $b_{4\,\&\,5} = 0.55$, $b_{5\,\&\,6} = 0.755$, $b_{6\,\&\,7} = 0.879$, $b_{7\,\&\,8} = 0.807$, and $b_{8\,\&\,9} = 1.57$, Step difficulty 1 (i.e., $b_{0\,\&\,1}$) showed that the transition point for examinees to have a score of 0 or a score of 1 was 1.619 which means that for examinees to move from a score of 0 to a score of 1, the ability must be 1.619 or greater. That is examinees with ability of 1.619 or greater would be able to move from score of 0 to 1. Step difficulty $b_{1\&2} = -0.259$ showed that the transition location at -0.259 is the transition between a category score of 1 and score of 2. This implies that examinees with ability of -0.259 or higher would be able to move from a score of 1 to a score of 2. Step difficulty $b_{2\,\&3} = 0.64$ showed that the transition location at 0.64 is the transition between a category score of 2 and score of 3. This explained that movement from a score of 2 to a score of 3 will require an ability level of 0.64 or more. Step difficulty $b_{3\,\&\,4} = 0.696$ implies that to move from a score of 3 to a score of 4, the ability level of 0.696 is required. Step difficulty $b_{4\,\&\,5} = 0.55$ implies that the transition location at 0.55 is the transition between a category score of 4 and that of 5. This means that examinees with ability 0.55 or greater would be able to move from a score of 4 to a score of 5. Step difficulty $b_{5\,\&\,6} = 0.755$, showed that the transition location at 0.755 is the transition between a category score of 5 and score of 6 showed that the difficulty of examinees to move from score of 5 to a score of 6 was 0.755. This implies that examinees with ability of 0.755 or greater would be able to move from score of 5 to 6. Step difficulty $b_{6\,\&\,7} = 0.879$ showed that the transition point for examinees to have a score of 6 or a score of 7 was 0.879. This implies that examinees with ability 0.879 or greater would be able to move from score of 6 to 7.

Step difficulty $b_{7\,\&\,8} = 0.807$ showed that the transition location at 0.807 is the transition between a category score of 7 and score of 8. This implies that 50% of the examinees with ability of 0.807 or greater would be able to move from score of 7 to 8. Step difficulty $b_{8\,\&\,9} = 1.57$ and this showed that the difficulty of examinees to move from score of 8 to a score of 9 was 0.157. This implies that 50% of the examinees with ability 0.157 or greater would be able to move from score of 8 to 9. The result for this item showed that the transition location parameters were not in sequential order. For example, the probability of obtaining a score of 1 is higher than that of obtaining a score of 2, but the probability of obtaining a score of 3 is greater than that of obtaining a score of 2. This showed that the item is not functioning as expected. See appendix VI for the category response curve for item 1B

Furthermore, Table 5 shows that item 3 (i.e., 1C), have five step difficulties. They are: $b_{0\,\&\,1} = 4.327$, $b_{1\&2} = -2.344$, $b_{2\,\&3} = 6.793$, $b_{3\,\&\,4} = -3.219$, and $b_{4\,\&\,5} = 2.451$ Step difficulty 1 (i.e., $b_{0\,\&\,1}$) = 4.327, showed that the transition location at 4.327 is the transition between a category score of 0 and score of 1 this means that for examinees to move from a score of 0 to a score of 1 the ability must be 4.327 or greater. That is examinees with ability 4.327 or greater would be able to move from score of 0 to 1. Step difficulty $b_{1\&2} = -2.344$ showed that examinees must have ability of -2.344 or more to move from a score of 1 to a score of 2. Step difficulty $b_{2\,\&3} = 6.793$ showed that the transition point for examinees to have a score of 2 or a score of 3 was 6.793. This explains that movement from a score of 2 to a score of 3 will require an ability level of 6.793 or more. Step difficulty $b_{3\,\&\,4} = -3.219$ showed that the transition point for examinees to have a score of 3 or a score of 4 was -3.219. This implies that examinees with ability -3.219 or greater would be able to move from a score of 3 to a score of 4. Step difficulty $b_{4\,\&\,5} = 2.451$ transition location at 2.451 is the transition between a category score of 4 and that of 5. This implies that examinees with ability 2.451. The result for this item showed that the transition location parameters were not in sequential order. For example, the probability of obtaining a score of 1 is substantially greater than that of obtaining a score of 2, but the probability of obtaining a score of 3 is greater than that of obtaining a score of 2. This showed that the item is not functioning as expected. See appendix VI for the category response curve for item 1C

In the same vain, Table 5 shows that item 4 (i.e., 1D), have four step difficulties. They are:$b_{0\,\&\,1} = 1.949$, $b_{1\&2} = -0.838$, $b_{2\,\&3} = 2.453$ and $b_{3\,\&\,4} = -0.365$. Step difficulty 1 (i.e.,$b_{0\,\&\,1}$) showed that the transition point for examinees to have a score of 0 or a score of 1 was 1.949, showed that the difficulty of examinees to move from score of 0 to a score of 1 was 1.949. This implies that examinees with ability 1.949 or greater would be able to move from score of 0 to 1. For step difficulty $b_{1\&2} = -0.838$ showed that the difficulty of examinees to move from score of 1 to a score of 2 was -0.838 meaning those examinees with ability -1.838 or more would be able to move from a score of 1 to 2. Also for step difficulty $b_{2\,\&3} = 2.453$ showed that the difficulty of examinees to move from a score of 2 to a score of 3 was 2.453 meaning those examinees with ability 2.453 or greater would be able to move from a score of 2 to a score of 3. Step difficulty $b_{3\,\&\,4} = -0.365$ showed that the difficulty of examinees to move from a score of 3 to a score of 4 was -0.365 which implies that examinees with ability -0.365 or more would be able to move from a score of 3 to a score of

4. The result for this item showed that the transition location parameters were not in sequential order. For example, the probability of obtaining a score of 1 is substantially greater than that of obtaining a score of 2, but the probability of obtaining a score of 3 is greater than that of obtaining a score of 2. This showed that the item is not functioning as expected. As it was observed that examinees on higher ability level had low category scores while examinees with low ability had high category scores. See appendix VI for the category response curve for item 1D.

Still on Table 5, it shows that item 5 (i.e., 2A), have nine step difficulties. They are: $b_{0\,\&\,1} = 7.481$, $b_{1\&2} = 1.256$, $b_{2\,\&3} = -2.335$, $b_{3\,\&\,4} = -7.778$, $b_{4\,\&\,5} = 7.174$, $b_{5\,\&\,6} = 0.313$, $b_{6\,\&\,7} = 1.651$, $b_{7\,\&\,8} = -3.897$, and $b_{8\,\&\,9} = 5.787$, Step difficulty $b_{0\,\&\,1}$, showed that the transition point for examinees to have a score of 0 or a score of 1 was 7.481 which means that for examinees to move from a score of 0 to a score of 1 the ability must be 7.481 or greater. So examinees with ability 7.481 or greater would be able to move from score of 0 to 1. Step difficulty $b_{1\&2} = 1.256$ showed that the transition point for examinees to have a score of 1 or a score of 2 was 1.256. This implies that examinees with ability 1.256 or would be able to move from 1 to 2. Step difficulty$b_{2\,\&3} = -2.335$, showed that the difficulty of examinees to move from score of 2 to a score of 3 was -2.335. This implies that examinees with ability -2.335 or greater would be able to move from score of 2 to 3. Step difficulty $b_{3\,\&\,4} = -7.778$ showed that the difficulty of examinees to move from score of 3 to a score of 4 was -7.778. This implies that examinees with ability -7.778 or greater would be able to move from score of 3 to 4. Step difficulty $b_{4\,\&\,5} = 7.174$ showed that the difficulty of examinees to move from score of 4 to a score of 5 was 7.174. This implies that examinees with ability 7.174 or greater would be able to move from score of 4 to 5. Step difficulty $b_{5\,\&\,6} = 0.313$ showed that the difficulty of examinees to move from score of 5 to a score of

6 was 0.313. This implies that examinees with ability 0.313 or greater would be able to move from score of 5 to 6.

Step difficulty $b_{6\,\&\,7} = 1.651$ showed that the difficulty of examinees to move from score of 6 to a score of 7 was 1.651. This implies that examinees with ability 1.651 or greater would be able to move from score of 6 to 7. Step difficulty $b_{7\,\&\,8} = -3.897$ showed that the difficulty of examinees to move from score of 7 to a score of 8 was -3.897. This implies that examinees with ability -3.897 or greater would be able to move from score of 7 to 8. Step difficulty $b_{8\,\&\,9} = 5.787$ showed that the difficulty of examinees to move from score of 8 to a score of 9 was 5.787. This implies that examinees with ability 5.787 or greater would be able to move from score of 8 to 9. The result for this item showed that the transition location parameters were not in sequential order. For example, the probability of obtaining a score of 1 is substantially greater than that of obtaining a score of 2, but the probability of obtaining a score of 3 is lesser than that of obtaining a score of 2. This showed that the item is not functioning as expected. See appendix VI for the category response curve for item 2A.

For item 6, table 5 shows that item 6 (i.e., 2B), have four step difficulties. They are: $b_{0\,\&\,1} = 7.427$, $b_{1\&2} = 0.94$, $b_{2\,\&3} = 2.86$, and $b_{3\,\&\,4} = -4.296$, , Step difficulty $b_{0\,\&\,1} = 7.427$, means that for examinees to move from a score of 0 to a score of 1 the ability must be 7.427 or greater. That is examinees with ability 7.427 or greater would be able to move from score of 0 to 1. Step difficulty $b_{1\&2} = 0.94$ showed that the transition point for examinees to have a score of 1 or a score of 2 was 0.94. This implies that examinees with ability 0.94 or greater would be able to move from 1 to 2. Step difficulty$b_{2\,\&3} = 2.86$, means that for examinees to move from a score of 2 to a score of 3 the ability must be 2.86 or greater. That is 50% of the examinees with ability 2.86 or greater would be able to move from score of 2 to 3. Step difficulty$b_{3\,\&\,4} = -4.296$, means that for examinees to move from a score of 3 to a score of 4 the ability must be -4.296 or greater. That is 50% of the examinees with ability -4.297 or greater would be able to move from score of 3 to 4. The result for this item showed that the transition location parameters were not in sequential order. For example, the probability of obtaining a score of 1 is substantially greater than that of obtaining a score of 2, and the probability of obtaining a score of 3 is greater than that of obtaining a score of 2 while the probability of obtaining the score of 4 is lesser than the probability of obtaining of the scores of 0, 1, 2 or 3. This showed that the item is not functioning as expected. See appendix VI for the category response curve for item 2B.

Also, table 5 shows that item 7 (i.e., 2C), have two step difficulties. They are: $b_{0\,\&\,1} = -7.279$ and $b_{1\&2} = 8.734$. Step difficulty$b_{0\,\&\,1} = -7.279$, showed that the transition point for examinees to have a score of 0 or a score of 1 was -7.279. That is examinees with ability 7.279 or greater would be able to move from score of 0 to 1. Step difficulty $b_{1\&2} = 8.734$ showed that examinees must have ability of 8.734 or more to move from a score of 1 to a score of 2. This implies that examinees with ability 8.734 or greater would be able to move from 1 to 2. The result for this item showed that the transition location parameters were in sequential order. For example, the probability of obtaining a score of 1 is lesser than that of obtaining a score of 2. This can be regarded as a good item and this shows that the item is functioning as expected. See appendix VI for the category response curve for item 2C.

Also, table 5 shows that item 8 (i.e., 3A), have five step difficulties. They are:$b_{0\,\&\,1} = -1.171$, $b_{1\&2} = -1.49$, $b_{2\,\&3} = -0.984$ $b_{3\,\&\,4} = 1.846$, and $b_{4\,\&\,5} = 0.956$. Step difficulty 1 (i.e.,$b_{0\,\&\,1}$), showed that for examinees to move from score of 0 to a score of 1 their ability must be -1.171 or greater. This implies that examinees with ability -1.171 or greater would be able to move from score of 0 to 1. For step difficult $b_{1\&2} = -1.49$ showed that the transition point for examinees to have a score of 1 to a score of 2 was -1.49 meaning that examinees with ability -1.49 or more would be able to move from a score of 1 to 2. Also for step difficulty $b_{2\,\&3} = -0.984$ showed that the difficulty of examinees to move from a score of 2 to a score of 3 was -0.984 meaning that examinees with ability -0.984 or greater would be able to move from a score of 2 to a score of 3. Step difficulty $b_{3\,\&\,4} = 1.846$ showed that the difficulty of examinees to move from a score of 3 to a score of 4 was 1.846 which implies that 50% of the examinees with ability 1.846 or more would be able to move from a score of 3 to a score of 4. Step difficulty $b_{4\,\&\,5} = 0.956$ showed that the difficulty of examinees to move from a score of 4 to a score of 5 was 0.956 this implies that 50% of the examinees with ability 0.956 or greater would be able to move from a score of 4 to a score of 5. The result for this item showed that the transition location parameters were not in sequential order. For example, the probability of obtaining a score of 4 is greater than that of obtaining a score of 5. This shows that the item is not functioning as expected. See appendix VI for the category response curve for item 3A.

Furthermore, table 5 shows that item 9 (i.e., 3B), have six step difficulties. They are:$b_{0\,\&\,1} = 3.459$, $b_{1\&2} = -2.02$, $b_{2\,\&3} = 2.037$ $b_{3\,\&\,4} = -0.507$, $b_{4\,\&\,5} = 3.188$ and $b_{5\&\,6} = 2.615$. Step difficulty (i.e.,$b_{0\,\&\,1}$), showed that the transition point for examinees to have a score of 0 or a score of 1 was 3.459. This implies that examinees with ability 3.459 or greater would be able to move from score of 0 to 1. For step difficult $b_{1\&2} = -2.02$ showed that the difficulty of examinees to move from score of 1 to a score of 2 was -2.02 meaning that examinees with ability -2.02 or more would be able to move from a score of 1 to 2.

Also for step difficulty $b_{2\,\&3} = 2.037$ showed that the difficulty of examinees to move from a score of 2 to a score of 3 was 2.037 meaning that 50% of the examinees with ability -2.037 or greater would be able to move from a score of 2 to a score of 3. Step difficulty $b_{3\,\&\,4} = -0.507$ showed that the difficulty of examinees to move from a score of 3 to a score of 4 was -0.507 which implies that 50% of the examinees with ability -0.507 or more would be able to move from a score of 3 to a score of 4. Step difficulty $b_{4\ \&\ 5} = 3.188$ showed that the difficulty of examinees to move from a score of 4 to a score of 5 was 3.188 this implies that 50% of the examinees with ability 3.188 or greater would be able to move from a score of 4 to a score of 5. Step difficulty $b_{5\&\ 6} = 2.615$ showed that the transition point for examinees to have a score of 5 or a score of 6 was 2.615 which implied that examinees with ability 2.615 or greater would be able to move from a score of 5 to a score of 6. The result for this item showed that the transition location parameters were not in sequential order. For example, the probability of obtaining a score of 1 is greater than that of obtaining a score of 2 and the probability of obtaining a score of 3 is greater than the probability of obtaining a score of 2. This shows that the item is not functioning as expected. See appendix VI for the category response curve for item 3B.

Again, table 5 shows that item 10 (i.e., 4A), have six step difficulties. They are:$b_{0\ \&\ 1} = -6.399$, $b_{1\&2} = -2.105$, $b_{2\ \&3} = -1.098$ $b_{3\ \&\ 4} = 0.596$, $b_{4\ \&\ 5} = 2.288$ and $b_{5\&\ 6} = 3$. Step difficulty (i.e.,$b_{0\ \&\ 1}$), showed that the transition point for examinees to have a score of 0 or a score of 1 was -6.399. This implies that examinees with ability -6.399 or greater would be able to move from score of 0 to 1. For step difficult $b_{1\&2} = -2.105$ showed that the difficulty of examinees to move from score of 1 to a score of 2 was -2.105 meaning that 50% of the examinees with ability -2.105 or more would be able to move from a score of 1 to 2. Also for step difficulty $b_{2\ \&3} = -1.098$ showed that the difficulty of examinees to move from a score of 2 to a score of 3 was -1.098 meaning that 50% of the examinees with ability -1.098 or greater would be able to move from a score of 2 to a score of 3. Step difficulty $b_{3\ \&\ 4} = 0.596$ showed that the difficulty of examinees to move from a score of 3 to a score of 4 was 0.596 which implies that 50% of the examinees with ability 0.596 or more would be able to move from a score of 3 to a score of 4. Step difficulty

$b_{4\ \&\ 5} = 2.288$ showed that the difficulty of examinees to move from a score of 4 to a score of 5 was 2.288 this implies that 50% of the examinees with ability 2.288 or greater would be able to move from a score of 4 to a score of 5. Step difficulty $b_{5\&\ 6} = 3$ showed that the transition point for examinees to have a score of 5 or a score of 6 was 3 which implied that examinees with ability 3 or greater would be able to move from a score of 5 to a score of 6. The result for this item showed that the transition location parameters were in sequential order. For example, the probability of obtaining a score of 1 is lesser than that of obtaining a score of 2, 3, 4, 5 and 6. This step difficulty is which is in increasing order shows that the item is functioning as expected. See appendix VI for the category response curve for item 4A.

In another vain, table 5 indicates that item 11 (i.e., 4B), have five step difficulties. They are:$b_{0\ \&\ 1} = 3.507$, $b_{1\&2} = -1.611$, $b_{2\ \&3} = 3.822$ $b_{3\ \&\ 4} = -0.614$, and $b_{4\ \&\ 5} = 0.353$. Step difficulty (i.e.,$b_{0\ \&\ 1}$), showed that the transition point for examinees to have a score of 0 or a score of 1 was 3.507. This implies that examinees with ability 3.507 or greater would be able to move from score of 0 to 1. For step difficult $b_{1\&2} = -1.611$ showed that the difficulty of examinees to move from score of 1 to a score of 2 was -1.611 meaning that 50% of the examinees with ability -1.611 or more would be able to move from a score of 1 to 2. Also for step difficulty $b_{2\ \&3} = 3.822$ showed that the difficulty of examinees to move from a score of 2 to a score of 3 was 3.822 meaning that 50% of the examinees with ability -3.822 or greater would be able to move from a score of 2 to a score of 3. Step difficulty $b_{3\ \&\ 4} = -0.614$ showed that the difficulty of examinees to move from a score of 3 to a score of 4 was -0.614 which implies that 50% of the examinees with ability -0.614 or more would be able to move from a score of 3 to a score of 4. Step difficulty $b_{4\ \&\ 5} = 0.354$ showed that the difficulty of examinees to move from a score of 4 to a score of 5 was 0.354 this implies that 50% of the examinees with ability 0.354 or greater would be able to move from a score of 4 to a score of 5. The result for this item showed that the transition location parameters were not in sequential order. For example, the probability of obtaining a score of 1 is greater than that of obtaining a score of 2 but the probability of obtaining a score of 3 is greater than the probability of obtaining a score of 2. Also, the probability of obtaining 4 is lesser than the probability of obtaining 3. This shows that the item is not functioning as expected. See appendix VI for the category response curve for item 4B.

Table 5 indicates that item 12 (i.e., 4C), have four step difficulties. They are:$b_{0\ \&\ 1} = -0.79$, $b_{1\&2} = 0.695$, $b_{2\ \&3} = 2.526$ and $b_{3\ \&\ 4} = -5.44$,. Step difficulty (i.e.,$b_{0\ \&\ 1}$) which is -0.79, showed that for examinees to move from score of 0 to a score of 1 their ability must be -0.79 or greater. This implies that 50% of the examinees with ability -0.79 or greater would be able to move from score of 0 to 1. For step difficult $b_{1\&2} = 0.695$ showed that the difficulty of examinees to move from score of 1 to a score of 2 was 0.695 meaning that 50% of the examinees with ability 0.695 or more would be able to move from a score of 1 to 2. Also for step difficulty $b_{2\ \&3} = 2.526$ showed that the difficulty of examinees to move from a score of 2 to a score of 3 was 2.526 meaning that 50% of the examinees with ability 2.526 or greater would be able to move from a score of 2 to a score of 3. Step difficulty $b_{3\ \&\ 4} = -5.44$ showed that the difficulty of examinees to move from a score of 3 to a score of 4 was -5.44 which implies that 50% of the examinees with ability -5.44 or more would be able to move from a score of 3 to a score of 4. The result for this item showed that the transition location parameters were not in sequential order. For example, the probability of obtaining a score of 1 is lesser than that of obtaining a score of 2 and the probability of obtaining a score of 3 is greater than the probability of obtaining a score of 2. Also, the probability of obtaining 4 is lesser than the probability of obtaining 3. This shows that the item is not functioning as expected. Figure 13 illustrates the step difficulty for item 4C.

Also, table 5 indicates that item 13 (i.e., 5AI), have three step difficulties. They are:$b_{0\,\&\,1} = 3.837$, $b_{1\&2} = 2.626$, and $b_{2\,\&3} = 0.013$. Step difficulty (i.e.,$b_{0\,\&\,1}$) 3.837, showed that for examinees to move from score of 0 to a score of 1 their ability must be 3.837 or greater. This implies that 50% of the examinees with ability 3.837 or greater would be able to move from score of 0 to 1. For step difficult $b_{1\&2} = 2.626$ showed that the transition location at 2.626 is the transition between a category score of 1 and score of 2. This mean those examinees with ability 2.626 or more would be able to move from a score of 1 to 2. Also for step difficulty $b_{2\,\&3} = 0.013$ showed that the difficulty of examinees to move from a score of 2 to a score of 3 was 0.013 meaning that 50% of the examinees with ability 0.013 or greater would be able to move from a score of 2 to a score of 3. The result for this item showed that the transition location parameters were in decreasing order or reverse order. For example, the probability of obtaining a score of 1 is greater than that of obtaining a score of 2 and the probability of obtaining a score of 3 is lesser than the probability of obtaining a score of 2. This shows that the item is not functioning as expected. See appendix VI for the category response curve for item 5AI.

Table 5 shows that item 14 (i.e., 5AII), have six step difficulties. They are:$b_{0\,\&\,1} = -2.049$, $b_{1\&2} = -3.692$, $b_{2\,\&3} = -0.726$, $b_{3\,\&\,4} = 3.426$, $b_{4\,\&\,5} = 1.633$ and $b_{5\,\&\,6} = -3.903$. Step difficulty (i.e.,$b_{0\,\&\,1}$) which is -2.049, showed that for examinees to move from score of 0 to a score of 1 their ability must be -2.049 or greater. This implies that examinees with ability -2.049 or greater would be able to move from score of 0 to 1. For step difficult $b_{1\&2} = -3.625$ showed that the difficulty of examinees to move from score of 1 to a score of 2 was -3.692 meaning that 50% of the examinees with ability -3.692 or more would be able to move from a score of 1 to 2. Also for step difficulty $b_{2\,\&3} = -0.726$ showed that the transition location at -0.726 is the transition between a category score of 2 and score of 3 that the difficulty of examinees to move from a score of 2 to a score of 3 was -0.726 meaning that examinees with ability -0.726 or greater would be able to move from a score of 2 to a score of 3. Step difficulty $b_{3\,\&\,4} = 3.426$ showed that the difficulty of examinees to move from a score of 3 to a score of 4 was 3.426 which implied that 50% of the examinees with ability 3.426 or more would be able to move from a score of 3 to a score of 4. Step difficulty $b_{4\,\&\,5} = 1.633$ showed that the difficulty of examinees to move from a score of 4 to a score of 5 was 1.633 meaning that 50% of the examinees with ability 1.633 or greater would be able to move from a score of 4 to a score of 5. Step difficulty $b_{5\,\&\,6} = -3.903$ showed that the transition location at -3.903 is the transition between a category score of 5 and score of 6 meaning that examinees with ability -3.903 or greater would be able to move from a score of 5 to a score of 6. The result for this item showed that the transition location parameters were not in sequential order. For example, the probability of obtaining a score of 1 is higher than that of obtaining a score of 2 and the probability of obtaining a score of 3 is greater than the probability of obtaining a score of 2. Also, the probability of obtaining a score of 6 is lesser than the probability of obtaining 0, 1, 2, 3, 4, and 5. This shows that the item is not functioning as expected. See appendix VI for the category response curve for item 5AII.

On the other hand, table 5 indicates that item 15 (i.e., 5B), have six step difficulties. They are:$b_{0\,\&\,1} = 4.601$, $b_{1\&2} = -0.03$, $b_{2\,\&3} = 1.947$, $b_{3\,\&\,4} = -1.523$, $b_{4\,\&\,5} = 4.182$ and $b_{5\,\&\,6} = 1.527$. Step difficulty $b_{0\,\&\,1} = 4.601,$ showed that for examinees to move from score of 0 to a score of 1 their ability must be 4.601 or greater. This implies that 50% of the examinees with ability -4.601 or greater would be able to move from score of 0 to 1. For step difficult $b_{1\&2}$ which is -0.03 showed that the difficulty of examinees to move from score of 1 to a score of 2 was -0.03 meaning that 50% of the examinees with ability of -0.03 or more would be able to move from a score of 1 to 2. Also for step difficulty $b_{2\,\&3} = 1.947$ showed that the difficulty of examinees to move from a score of 2 to a score of 3 was 1.947 meaning that 50% of the examinees with ability 1.947 or greater would be able to move from a score of 2 to a score of 3. Step difficulty $b_{3\,\&\,4} = -1.523$ showed that the difficulty of examinees to move from a score of 3 to a score of 4 was -1.523 which implied that 50% of the examinees with ability -1.523 or more would be able to move from a score of 3 to a score of 4. Step difficulty $b_{4\,\&\,5} = 4.182$ showed that the transition location at 4.182 is the transition between a category score of 4 and score of 5. That is difficulty of examinees to move from a score of 4 to a score of 5 was 4.182 meaning that examinees with ability 4.182 or greater would be able to move from a score of 4 to a score of 5. Step difficulty $b_{5\,\&\,6} = 1.527$ showed that the difficulty of examinees to move from a score of 5 to a score of 6 was 1.527 meaning that 50% of the examinees with ability 1.527 or greater would be able to move from a score of 5 to a score of 6. The result for this item showed that the transition location parameters were not in sequential order. For example, the probability of obtaining a score of 1 is higher than that of obtaining a score of 2 and the probability of obtaining a score of 3 is greater than the probability of obtaining a score of 2. Also, the probability of obtaining a score of 6 is lesser than the probability of obtaining 1. See appendix VI for the category response curve for item 5B.

Also, table 5 indicates that item 16 (i.e., 6A), has two step difficulties. $b_{0\,\&\,1} = 1.409$ and $b_{1\&2} = -4.07$, Step difficulty $b_{0\,\&\,1}$ which is 1.409 showed that for examinees to move from score of 0 to a score of 1 their ability must be 1.409 or greater. This implies that 50% of the examinees with ability 1.409 or greater would be able to move from score of 0 to 1. For step difficult $b_{1\&2} = -4.07$ showed that the difficulty of examinees to move from score of 1 to a score of 2 was -4.07 meaning that 50% of the examinees with ability -4.07 or more would be able to move from a score of 1 to 2. The result for this item showed that the transition location parameters were not in sequential order. For example, the probability of obtaining a score of 1 is substantially greater than that of obtaining a score of 2. This showed that the item is not functioning as expected. See appendix VI for the category response curve for item 6A.

Furthermore, Table 5 shows that item 17 (i.e., 6B), have six step difficulties. They are:$b_{0\ \&\ 1} = 1.467,\ b_{1\&2} = -4.019,\ b_{2\ \&3} = -1.25,\ b_{3\ \&\ 4} = -1.308,\ b_{4\ \&\ 5} = 0.664.$ Step difficulty $b_{0\ \&\ 1} = 1.467,$ showed that for examinees to move from score of 0 to a score of 1 their ability must be 1.467 or greater. This implies that 50% of the examinees with ability 1.467 or greater would be able to move from score of 0 to 1. For step difficult $b_{1\&2} = -4.019$ showed that the difficulty of examinees to move from score of 1 to a score of 2 was -4.019 meaning that examinees with ability -4.019 or more would be able to move from a score of 1 to 2. Also for step difficulty $b_{2\ \&3}$ which is – 1.25 showed that the difficulty of examinees to move from a score of 2 to a score of 3 was -1.25 meaning that examinees with ability -1.25 or greater would be able to move from a score of 2 to a score of 3. Step difficulty $b_{3\ \&\ 4} = -1.308$ showed that the difficulty of examinees to move from a score of 3 to a score of 4 was -1.308 which implied that examinees with ability -1.308 or more would be able to move from a score of 3 to a score of 4. Step difficulty $b_{4\ \&\ 5} = 0.664$ showed that the difficulty of examinees to move from a score of 4 to a score of 5 was 0.664 meaning that 50% of the examinees with ability 0.664 or greater would be able to move from a score of 4 to a score of 5. The result for this item showed that the transition location parameters were not in sequential order. For example, the probability of obtaining a score of 1 is higher than that of obtaining a score of 2 and the probability of obtaining a score of 3 is greater than the probability of obtaining a score of 2. Also, the probability of obtaining a score of 4 is lesser than the probability of obtaining 3. This shows that the item is not functioning as expected. See appendix VI for the category response curve for item 6B.

Also, table 5 indicates that item 18 (i.e., 6C), have eight step difficulties. They are:$b_{0\,\&\,1} = 8.228$, $b_{1\&2} = -1.655$, $b_{2\,\&3} = 0.393$, $b_{3\,\&\,4} = -3.815$, $b_{4\,\&\,5} = 6.232$, $b_{5\,\&\,6} = 0.091$, $b_{6\,\&\,7} = 3.038$, $b_{7\,\&\,8} = -2.466$. Step difficulty $b_{0\,\&\,1} = 8.228$, showed that the transition location at 8.228 is the transition between a category score of 0 and score of 1 showed that for examinees to move from score of 0 to a score of 1 their ability must be 8.228 or greater. This implies that examinees with ability 8.228 or greater would be able to move from score of 0 to 1. For step difficult $b_{1\&2} = -1.655$ showed that the difficulty of examinees to move from score of 1 to a score of 2 was -1.655 meaning that 50% of the examinees with ability -1.655 or more would be able to move from a score of 1 to 2. Also for step difficulty $b_{2\,\&3} = 0.393$ showed that the difficulty of examinees to move from a score of 2 to a score of 3 was 0.393 meaning that examinees with ability 0.393 or greater would be able to move from a score of 2 to a score of 3. Step difficulty $b_{3\,\&\,4} = -3.815$ showed that the difficulty of examinees to move from a score of 3 to a score of 4 was -3.815 which implied that examinees with ability -3.815 or more would be able to move from a score of 3 to a score of 4. Step difficulty $b_{4\,\&\,5}6.232$ showed that the transition location at 6.232 is the transition between a category score of 4 and score of 5 meaning that examinees with ability 6.232 or greater would be able to move from a score of 4 to a score of 5. Step difficulty $b_{5\,\&\,6} = 0.091$ showed that the transition location at 0.091 is the transition between a category score of 5 and score of 6 meaning that examinees with ability 0.091 or greater would be able to move from a score of 5 to a score of 6. Step difficulty $b_{6\,\&7} = 3.038$ showed that the difficulty of examinees to move from a score of 6 to a score of 7 was 3.038. meaning that 50% of the examinees with ability 3.038 or greater would be able to move from a score of 6 to a score of 7.

Step difficulty $b_{7\,\&\,8} = -2.466$ showed that the transition location at -2.466 is the transition between a category score of 7 and score of 8. meaning that examinees with ability

-2.466 or greater would be able to move from a score of 7 to a score of 8. The result for this item showed that the transition location parameters were not in sequential order. For example, the probability of obtaining a score of 1 is higher than that of obtaining a score of 2 and the probability of obtaining a score of 3 is higher than the probability of obtaining a score of 2. Also, the probability of obtaining a score of 8 is lesser than the probability of obtaining 1. This shows that the item is not functioning as expected. See appendix VI for the category response curve for item 6C.

On the other hand, table 5 indicates that item 19 (i.e., 7AI), have -1.068 step difficulties. They are:$b_{0\,\&\,1} = -1.068$, and $b_{1\&2} = -1.689$, Step difficulty $b_{0\,\&\,1} = -1.068$, showed that the transition location at -1.068 is the transition between a category score of 0 and score of 1. This implies that examinees with ability -1.068 or greater would be able to move from score of 0 to 1. For step difficult $b_{1\&2} = -1.689$ showed that the difficulty of examinees to move from score of 1 to a score of 2 was -1.689 meaning that examinees with ability -1.689 or more would be able to move from a score of 1 to 2. The result for this item showed that the transition location parameters were not in sequential order. For example, the probability of obtaining 1 is greater than the probability of obtaining a score of 2. This shows that the item is not functioning as expected. See appendix VI for the category response curve for item 7AI.

For item 20, table 5 shows that item 20 (i.e., 7AII), have five step difficulties. They are:$b_{0\,\&\,1} = 0.461$, $b_{1\&2} = 1.551$, $b_{2\,\&3} = 0.039$, $b_{3\,\&\,4} = -2.823$, and $b_{4\,\&\,5} = -7.342$ Step difficulty $b_{0\,\&\,1} = 0.461$, showed that the transition location at 0.461 is the transition between a category score of 0 and score of 1. This implies that examinees with ability 0.461 or greater would be able to move from score of 0 to 1. For step difficult $b_{1\&2} = 1.551$ showed that the difficulty of examinees to move from score of 1 to a score of 2 was 1.551

meaning that 50% of the examinees with ability 1.551 or more would be able to move from a score of 1 to 2. Also for step difficulty $b_{2\,\&3} = 0.039$ showed that the difficulty of examinees to move from a score of 2 to a score of 3 was 0.039 meaning that examinees with ability 0.039 or greater would be able to move from a score of 2 to a score of 3. Step difficulty $b_{3\,\&\,4} = -2.823$ showed that the difficulty of examinees to move from a score of 3 to a score of 4 was -2.823 which implied that examinees with ability -2.823 or more would be able to move from a score of 3 to a score of 4. Step difficulty $b_{4\,\&\,5}$ showed that the transition location at -7.342 is the transition between a category score of 4 and score of 5. meaning that examinees with ability -7.342 or greater would be able to move from a score of 4 to a score of 5. The result for this item showed that the transition location parameters were not in sequential order. For example, the probability of obtaining a score of 1 is higher than that of obtaining a score of 4 and 5. This shows that the item is not functioning as expected. See appendix VI for the category response curve for item 7AII.

Also, table 5 indicates that item 21 (i.e., 7B), have eight step difficulties. They are:$b_{0\,\&\,1} = 3.826$, $b_{1\&2} = -3.038$, $b_{2\,\&3} = 1.297$, $b_{3\,\&\,4} = -2.718$, $b_{4\,\&\,5} = -2.779$, $b_{5\,\&\,6} = -4.135$, $b_{6\,\&\,7} = 5.681$, $b_{7\,\&\,8} = -5.445$. Step difficulty $b_{0\,\&\,1} = 3.826$, showed that for examinees to move from score of 0 to a score of 1 their ability must be 3.826 or greater. This implies that 50% of the examinees with ability 3.826 or greater would be able to move from score of 0 to 1. For step difficult $b_{1\&2} = -3.038$ showed that the transition location at -3.038 is the transition between a category score of 1 and score of 2. Meaning those examinees with ability -3.038 or more would be able to move from a score of 1 to 2. Also for step difficulty $b_{2\,\&3} = 1.297$ showed that the transition location at1.297 is the transition between categories score of 2 and a score of 3.Meaning those examinees with ability 1.297 or greater would be able to move from a score of 2 to a score of 3. Step difficulty $b_{3\,\&\,4} = -2.718$ showed that the difficulty of examinees to move from a score of 3 to a score of 4 was -2.718 which implied that 50% of the examinees with ability -2.718 or more would be able to move from a score of 3 to a score of 4.

Step difficulty $b_{4\,\&\,5} = 2.779$ showed that the difficulty of examinees to move from a score of 4 to a score of 5 was 2.779 meaning that 50% of the examinees with ability 2.779 or greater would be able to move from a score of 4 to a score of 5. Step difficulty $b_{5\,\&\,6} = -4.135$ showed that the difficulty of examinees to move from a score of 5 to a score of 6 was -4.135 meaning that 50% of the examinees with ability -4.135 or greater would be able to move from a score of 5 to a score of 6. Step difficulty $b_{6\,\&7} = 5.681$ showed that the difficulty of examinees to move from a score of 6 to a score of 7 was 5.681. meaning that 50% of the examinees with ability 5.681 or greater would be able to move from a score of 6 to a score of 7. Step difficulty $b_{7\,\&\,8} = -5.445$ showed that the difficulty of examinees to move from a score of 7 to a score of 8 was -5.445. meaning that 50% of the examinees with ability -5.445 or greater would be able to move from a score of 7 to a score of 8. The result for this item showed that the transition location parameters were not in sequential order. For example, the probability of obtaining a score of 1 is higher than that of obtaining a score of 2 and the probability of obtaining a score of 3 is higher than the probability of obtaining a score of 4. Also, the probability of obtaining a score of 8 is lesser than the probability of obtaining 1. This shows that the item is not functioning as expected. See appendix VI for the category response curve for item 7B.

Furthermore, table 5 indicates that item 22 (i.e., 8A), have six step difficulties. They are:$b_{0\,\&\,1} = -1.037$, $b_{1\&2} = -0.078$, $b_{2\,\&3}0.967$, $b_{3\,\&\,4} = 1.198$, $b_{4\,\&\,5} = 1.508$, $b_{5\,\&\,6} = 1.9$. Step difficulty $b_{0\,\&\,1} = -1.037$ showed that for examinees to move from score of 0 to a score of 1 their ability must be -1.037 or greater. This implies that 50% of the examinees with ability -1.037 or greater would be able to move from score of 0 to 1. For step difficult

$b_{1\&2} = -0.078$, showed that the difficulty of examinees to move from score of 1 to a score of 2 was -0.078 meaning that 50% of the examinees with ability -0.078 or more would be able to move from a score of 1 to 2. Also for step difficulty $b_{2\,\&3} = 0.967$ showed that the difficulty of examinees to move from a score of 2 to a score of 3 was 0.967 meaning that 50% of the examinees with ability 0.967 or greater would be able to move from a score of 2 to a score of 3. Step difficulty $b_{3\,\&\,4} = 1.198$ showed that the difficulty of examinees to move from a score of 3 to a score of 4 was 1.198 which implied that 50% of the examinees with ability 1.198 or more would be able to move from a score of 3 to a score of 4. Step difficulty $b_{4\,\&\,5} = 1.508$ showed that the difficulty of examinees to move from a score of 4 to a score of 5 was 1.508 meaning that 50% of the examinees with ability 1.508 or greater would be able to move from a score of 4 to a score of 5. Step difficulty $b_{5\,\&\,6} = 1.9$ showed that the transition location at 1.9 is the transition between a category score of 5 and score of 6. Meaning those examinees with ability 1.9 or greater would be able to move from a score of 5 to a score of 6. The result for this item showed that the transition location parameters were in sequential order. For example, the probability of obtaining a score of 1 is lesser than that of obtaining a score of 2 and it follows in sequential order like that. The probability of obtaining a score of 6 is the highest in the step difficulties. This shows that the item is functioning as expected. See appendix VI for the category response curve for item 8A.

Also, table 5 further indicates that item 23 (i.e., 8B), have eight step difficulties. They are:$b_{0\,\&\,1} = 1.238$, $b_{1\&2} = 1.251$, $b_{2\,\&3} = 1.652$, $b_{3\,\&\,4} = 1.908$, $b_{4\,\&\,5} = 2.269$, $b_{5\,\&\,6} = 1.802$, $b_{6\,\&\,7} = 2.477$, $b_{7\,\&\,8} = 3.214$. Step difficulty$b_{0\,\&\,1} = 1.238$, showed that for examinees to move from score of 0 to a score of 1 their ability must be 1.238 or greater. This implies that 50% of the examinees with ability 1.238 or greater would be able to move from score of 0 to 1. For step difficult $b_{1\&2} = 1.251$, showed that the transition location at 1.251. is the transition between categories score of 1 and score of 2. Meaning those examinees with ability 1.251 or more would be able to move from a score of 1 to 2. Also for step difficulty $b_{2\ \&3} = 1.652$ showed that the difficulty of examinees to move from a score of 2 to a score of 3 was 1.652 meaning that 50% of the examinees with ability 1.652 or greater would be able to move from a score of 2 to a score of 3. Step difficulty $b_{3\ \&\ 4} = 1.908$ showed that the difficulty of examinees to move from a score of 3 to a score of 4 was 1.908which implied that 50% of the examinees with ability 1.908 or more would be able to move from a score of 3 to a score of 4. Step difficulty $b_{4\ \&\ 5} = 2.269$ showed that the difficulty of examinees to move from a score of 4 to a score of 5 was 2.269 meaning that 50% of the examinees with ability 2.269 or greater would be able to move from a score of 4 to a score of 5.

Step difficulty $b_{5\ \&\ 6} = 1.802$ showed that the transition location at 1.802 is the transition between a category score of 5 and score of 6. Meaning those examinees with ability 1.802 or greater would be able to move from a score of 5 to a score of 6. Step difficulty $b_{6\ \&7} = 2.477$ showed that the transition location at 2.477 is the transition between a category score of 6 and score of 7. Meaning that examinees with ability 2.477 or greater would be able to move from a score of 6 to a score of 7. Step difficulty $b_{7\ \&\ 8} = 3.214$ showed that the difficulty of examinees to move from a score of 7 to a score of 8 was 3.214. meaning that 50% of the examinees with ability 3.214 or greater would be able to move from a score of 7 to a score of 8. The result for this item showed that the transition location parameters were not in sequential order. For example, the probability of obtaining a score of 4 is higher than that of obtaining a score of 6 and the probability of obtaining a score of 5 is higher than the probability of obtaining a score of 6. This shows that the item is not functioning as expected. See appendix VI for the category response curve for item 8B.

In summary, based on the step difficulties of the items, it was observed that items 2C, 3A and 8A were good items because the step difficulties were sequential while the remaining items were not good enough due to the non-sequential of their step difficulties.

From Table 5, the difficulty column represents the overall difficulty indices of the test items. These estimated difficulty parameters are useful in judging the difficulty or appropriateness of the test for a given set of examinees. These estimates indicate how easy or difficult the items were for the students. Easier items have lower (negative) difficulty indices and more difficult items have higher (positive) indices. On the table, item 2B was the most difficult, followed by 8B, 2A, 5B, 3B and by 6C. The estimated difficulty parameters were 62.366, 10.211, 5.042, 3.475, 3.251 and 3.149 respectively. The easiest item as presented on the table was 7AII, followed by 4C, 6A, 7AI and by 5AII. The estimated difficulty parameters of these items were: -2.371, -2.011, -1.987, -1.898 and -1.667 respectively.

More importantly, the table showed that items 2B, 8B, 2A, 5B, 3B and 6C were poor items as their difficulty parameters (62.366, 10.211, 5.042, 3.475, 3.251 and 3.149 respectively) were outside the range (-3 to 3) for which items difficulty parameter estimates are considered good (Baker, 2001; Hambleton & Jones, 1993; De Mars, 2010).

Research question 3: What are the Items Discrimination Estimates of The Physical Geography Tests?

To answer this research question, examinees' responses to the test items were collated and analysed using generalized partial credit model with Mplus 7.4. The estimation method used was the Robust Maximum Marginal Likelihood estimation (MLR). In all, there were eight broad questions out of which each examinee is expected to answer four. Because the privilege of choice which the examinees were given, missingness were not at random and traditional IRT which holds the assumption of missingness at random could not be used. As a result of this, the Examinee-selected item model (ESIM) was used in the estimation of the test items parameters. Although, there were eight questions in the test, each of the questions is made up of two or more test items. Consequently, there were 23 items found to make up the test. The item discrimination of the Geography test are shown in Table 6

Table 6: Item discrimination of the Geography test

Item	Discrimination
I1A	3.315
I1B	4.364
I1C	0.496
I1D	0.735
I2A	0.091
I2B	0.006
I2C	0.157
I3A	-0.034
I3B	0.162
I4A	0.041
I4B	0.166
I4C	0.143
I5AI	0.41
I5AII	0.141
I5B	0.114
I6A	0.224
I6B	0.321
I6C	0.055
I7AI	0.393
I7AII	0.151
I7B	0.095
I8A	0.095
I8B	0.035

Table 6 indicates the item discrimination of the PGAT, items that has a discrimination value of 0.4 and above discriminate well while items that has discrimination values of less than 0.4 do not discriminates well. I1B discriminates most followed by I1A, I1D, I1C and I5A1 the other items in the test showed poor discrimination. This is because the higher discrimination values do a better job of discriminating among the examinees than items with smaller discriminations (de Ayala, 2009).

Research Question 4: How invariant is the difficulty estimate of test items among male and female students?

To answer this research question, the scores of all the 1546 students was divided into two (male and female). The male scores were subjected to generalized partial credit model with Mplus 7.4 so as to obtain the difficulty estimate of test items of male students. The female scores were subjected to generalized partial credit model with Mplus 7.4 so as to obtain the difficulty estimate of test items of female students. The estimation method used was the Robust Maximum Marginal Likelihood estimation (MLR). The difficulty estimates for male and female is shown in table 7

Table 7: Difficulty Estimates among Male and Female students' samples

Items	Male	Female
I1A	0.501	0.421
I1B	1.189	0.852
I1C	1.571	1.791
I1D	0.743	0.847
I2A	9.621	5.912
I2B	2.966	62.37
I2C	-0.95	-1.18
I3A	0.183	-1.07
I3B	2.746	2.495
I4A	46.51	1.621
I4B	3.608	1.795
I4C	-1.47	-3.97
I5AI	3.682	2.005
I5AII	-0.89	-2.54
I5B	4.091	2.626
I6A	-1.35	-2.33
I6B	-1.12	-1.05
I6C	10.62	1.924
I7AI	-1.36	-1.69
I7AII	-1.5	-2.91
I7B	-0.48	-0.49
I8A	-0.07	-0.3
I8B	1.863	5.439
Mean	3.509	3.155
STD	9.918	13.14

Table 7 indicates the difficulty parameter estimates of Geography test items. Column 2, titled male represents the difficulty estimates of the test items among the male students who took the test and column 3, titled female represents the difficulty of the test items among the female students sample. The table shows that on average, the test items were more difficult among the male students (Mean = 3.509, SD = 9.918) than among the female students (Mean = 3.155, SD = 13.14). The individual consideration of the items to show the level of invariance is presented in Figure 2. The difficulty estimates for both male and female are plotted in the graph as shown in figure 2. In figure 2, where there is disparity in the point of male and female shows that the item is not invariant that is the item functioned differently among male and female and where there is no disparity in the point of male and female shows there is invariant that is the item functioned the same for both male and female.

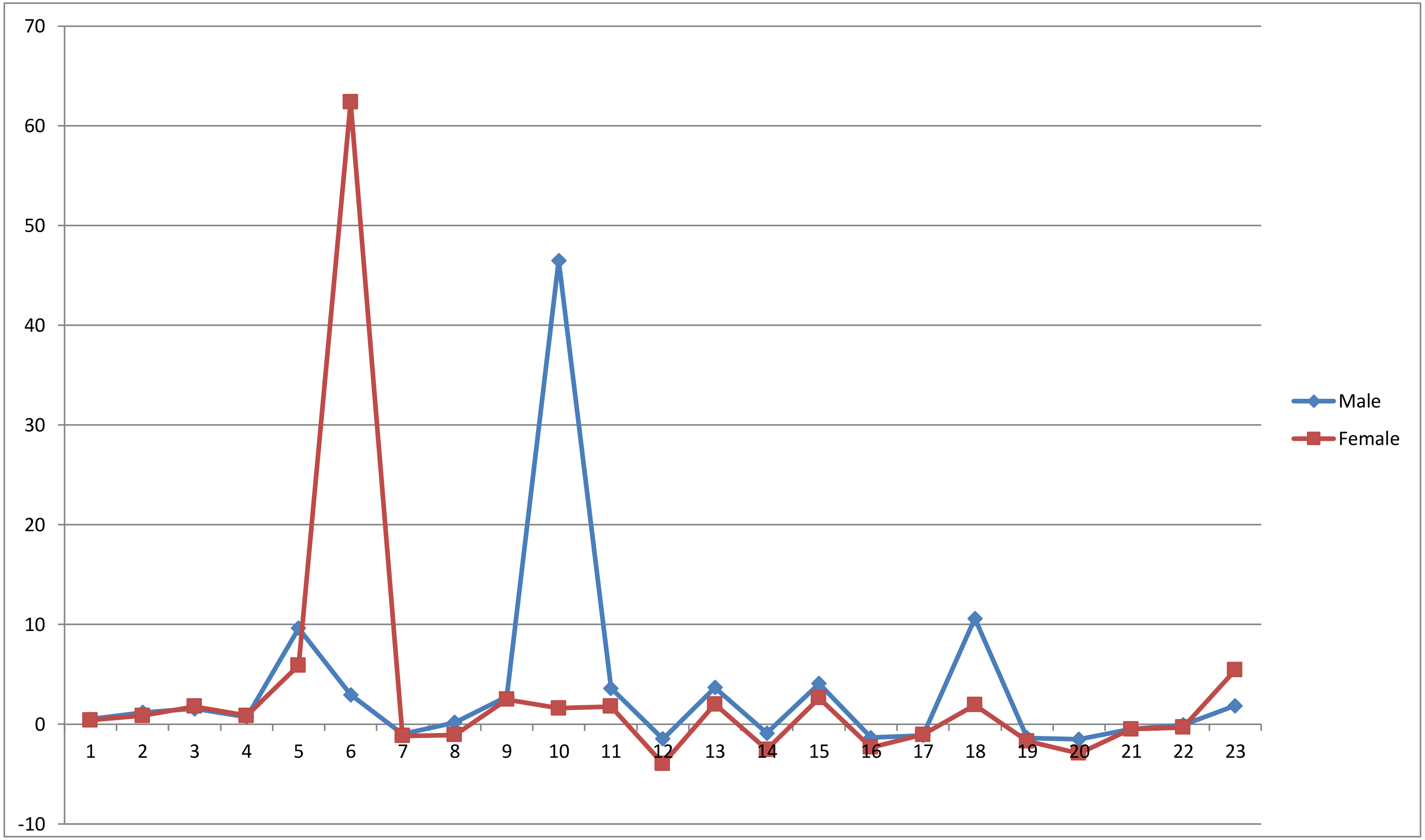


Figure 2: Distribution of difficulty estimates of Physical Geography test items in male and female samples

Figure 2 indicates that the difficulty estimates of ten items (items 5, 6, 10, 11, 12, 13, 14, 15, 18, and 23) appeared to obviously vary with respect to gender. The results revealed that the test items appeared to be marginally more difficult for the male sample than the female sample when considered on the average. However, when the items were considered individually, five of the items 5, 6, 10, 18 and 23 (representing about 22% of the test items) showed obvious variation in the difficulty estimates of the test items in both male and female samples. The implication of the results is that with respect to gender, five of the Physical Geography test items were not invariant (see appendix VII for the item characteristic curves that illustrate the degree of invariant for male and female students)

Research Question 5: How invariant is the difficulty estimate of the items among rural and urban school students?

To answer this research question, the scores of all the 1546 students was divided into two (rural and urban school students). The rural school student scores were subjected to generalized partial credit model with Mplus 7.4 so as to obtain the difficulty estimate of test items of rural school students. The urban school students' scores were subjected to generalized partial credit model with Mplus 7.4 so as to obtain the difficulty estimates of test items of urban school students' scores. The estimation method used was the Robust Maximum Marginal Likelihood estimation (MLR). The difficulty estimates for rural and urban school students male and female is shown in table 8

Table 8: Difficulty Estimates among Rural and Urban School Samples

Item	Rural	Urban
I1A	0.3	0.515
I1B	0.673	1.073
I1C	1.434	3.066
I1D	0.5	1.365
I2A	-6.87	2.55
I2B	3.124	-2.636
I2C	-0.851	-5.629
I3A	-1.11	-0.247
I3B	1.572	9.986
I4A	-1.258	0.758
I4B	-12.181	1.892
I4C	-3.65	-1.391
I5AI	3.504	2.83
I5AII	-4.525	-0.85
I5B	-15.218	1.912
I6A	-2.605	-1.828
I6B	-2.04	-0.887
I6C	-2.27	2.585
I7AI	-1.287	-2.914
I7AII	-1.247	-6.348
I7B	-2.5	0.063
I8A	-0.429	0.131
I8B	1.446	-8.635
Mean	-1.97774	-0.11474
STD	4.419633	3.753513

Table 8 shows the difficulty parameter estimates of Geography test items among students of schools located in rural and urban. Column 2, titled rural represents the difficulty estimates of the test items among students of rural schools who took the physical Geography Achievement test and column 3, titled urban, represents the difficulty of the test items among the students of urban schools sample. The table shows that the test items were more difficult for students of urban schools (Mean = -0.11, SD = 3.754) than for the students of rural schools (Mean = -1.98, SD = 4.420). Figure 4 presents the distribution of the difficulty estimates of the test items among students of rural and urban school. The difficulty estimates for both rural and urban are plotted in the graph as shown in figure 3.

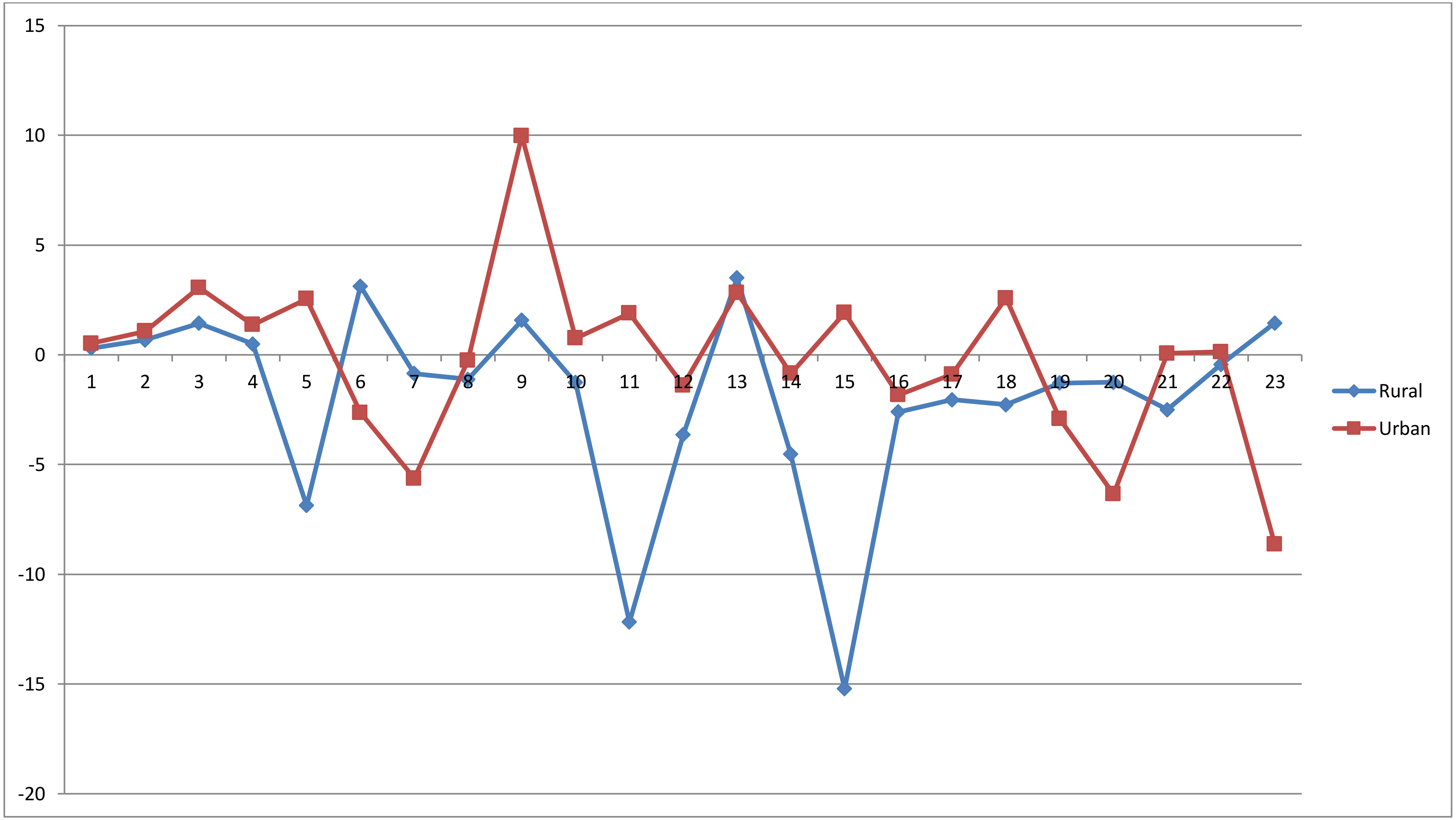


Figure 3: Distribution of difficulty estimates of Physical Geography test items among students of rural and urban

In figure 3, where there is disparity in the point of rural and urban shows that the item is not invariant that is the item functioned differently in rural and urban and where there is no disparity in the point of rural and urban shows there is invariant that is the item functioned the same for both rural and urban. It shows that 14 items (3, 5, 6, 7, 9, 10, 11, 12, 14, 15, 18, 19, 20, 21 and 23) of the Geography test had difficulty estimates in rural schools sample that were obviously at variance with the items' difficulty estimates in the urban schools sample. The results revealed that the test items appeared to be more difficult for the students of urban sample than their rural sample counterparts when considered on the average. More importantly, on the individual ground, most of the items showed dissimilar difficulty parameter estimates in the two contrasting samples. The implication of the results is that the extent of variation of the difficulty of the Physical geography test items with respect to school location was very large. (See appendix VIII for the item characteristic curves that illustrate the degree of invariant for rural and urban schools)

Testing of Hypotheses

$Ho_{1:}$ *There is no significant invariance of item difficulty parameter estimates across different samples of male and female students.*

H_1: *There is a significant invariance of item difficulty parameter estimates across different samples of male and female students.*

In order to assess whether the difference observed in the difficulty estimates of the items in the male and female samples was statistically significant, paired sample t-test statistic was conducted and the result is presented in Table 9.

Table 9: Paired sample t-test of difficult estimates Among Male and Female students Samples

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std	Std. Error	95% Conf. Int.				
					Lower	Upper			
Pair 1	male - female	.35387	16.04690	3.34601	-6.58533	7.29307	.106	22	.917

Table 9 indicates that the difference observed in the mean difficulty estimates of the Geography test items in the male and female sample was not statistically significant, $t = 0.106$, $p > 0.05$. Thus, there is a significant invariance of item difficulty parameter estimates across different samples of male and female students. The implication of this result is that the difficulty estimates of the Geography test items were statistically the same in the male and female students' samples.

Ho_1: *There is no significant invariance of item difficulty parameter estimates across different samples of rural and urban school location.*

H_1: *There is a significant invariance of item difficulty parameter estimates across different samples of rural and urban school location.*

In order to assess whether the difference observed in the difficulty estimates of the items in the urban and rural was statistically significant, paired sample t-test statistic was conducted and the result is presented in Table 10

Table 10: Paired sample t-test of difficulty estimates among students of rural and urban schools

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std.	Std. Error	95% Confid. Inter				
					Lower	Upper			
Pair 1	rural - urban	-1.86300	6.09533	1.27096	-4.49882	.77282	-1.466	22	.157

Table 10 showed that the difference observed in the mean difficulty estimates of the Geography test items in the rural and urban sample was not statistically significant, $t(22) = -1.466$, $p > 0.05$ 0.05. Thus, there is a significant invariance of item difficulty parameter estimates across different samples of examinees of rural and urban school location. The implication of this result is that the difficulty estimates of the Geography test items were statistically the same in the rural and urban school location samples.

Summary of Findings

Based on the data collected, collated, analyzed and interpreted, the following findings were obtained:

8. the examinees ability estimates in physical Geography showed that 49.35% of the examinees who sat for the Physical Geography Achievement Test had performance below average while 50.65% of the examinees had performance above average. This showed that the performance of the examinees in the PGAT was average performance.
9. item 2B was the most difficult with difficulty index of 62.366 (mean = 3.719)
10. item 1B discriminated most with discrimination index of 4.364 (mean = 0.5076)
11. five out of 23 items representing 22% of the test items showed obvious variation in the difficulty estimates of the test items across gender;
12. 14 out of 23 items were more difficult for students of urban schools (Mean = -0.11, SD = 3.754) than the students of rural schools (Mean = -1.98, SD = 4.420);
13. there was a significant invariance of item parameter estimates across male and female geography students; and
14. there was a significant invariance of item parameter estimates across rural and urban geography students.

CHAPTER FIVE

DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS

This chapter contains the discussion of the results presented in chapter four, conclusions of the study based on the findings, recommendations to the relevant stakeholders and suggestions for further study.

Discussion of findings

It was observed from the results obtained that 49.35% of the examinees had ability of less than zero while 50.65% of the examinees had ability of above 0 which implies that the performance of students in the PGAT were low. These findings are in agreement with studies by Adeyegbe, (2002), Onipede, (2003), Adeyemi, (2010) and Adewuni (2016). Adeyegbe (2002) found that there was a decline in students' performance in Senior Secondary Certificate Examinations. In the same vain, Onipede (2003) reported that students' performance was below expectations in Senior School Certificate Examinations (SSCE) in many subjects. Adeyemi (2010) also said that performance of students in senior secondary certificate Examinations were low. It was also reported by Adewuni (2016) that the performance of students in NECO Senior School Certificate Examinations Multiple-Choice Tests in Government for the years 2013 and 2014 were low.

The reason for the low performances may be because the test was not constructed in line with the IRT construction techniques. Also, the poor performance may be attributed to the fact that some of the questions in the test do not conform to the assumptions of IRT. For example the items were not locally independent (this is an assumption of IRT which states that responses to an item should not lead to the answer of another item) in the PGAT responses to an item lead to the responses of another item especially in the first part of the test. In the same vein, if an examinee fails an item then it will lead to the failure of the next item. The items in a test should be locally independent. Supporting this claim was Adedoyin (2010), who disclosed that IRT ability estimate is better than the ability estimate in CTT because of the assumption of local independence.

The result showed that out of the 23 items that constituted the test, 3 items (that is 2C, 4A and 8A) functioned as expected because the step difficulties of these items were in sequential order and item 2B was the most difficult. It was also observed that the high ability examinees scored low marks while the low ability examinees scored high marks at the different category score. The reason for this result could be because essay tests used by examination bodies were not trial tested to find out the psychometric properties of the test items before it was administered to the students. As different items in a test have different difficulty levels. This result is in agreement with study by Korobko, Glas, Bosker, and Luyten (2008). They found that a lower GPA may not necessarily mean that the examinee performs less well than students who have higher GPAs but the examinees with lower GPAs may simply be taking courses and studying in fields with more stringent grading standards. Korobko,et al (2008) explained that this is one of the problems of score adjustment for the comparison of students and schools performance. Johnson, (2003) noted that combining student grades through simple averaging schemes to obtain grade point averages (GPAs) results in systematic bias against students enrolled in more rigorous curricula.

This study revealed that the test items were more difficult for male than their female counterparts 49.09% males and 50.90% females. The reason may be that females generally obtain higher grades than their males counterparts this is because boys tend to show some level of superiority when they are in class especially SS3 and this superiority is shown by not writing notes, attending lessons, writing assignments etc. while the females are always in the class to attend lessons, write notes etc. all these behaviours lead to poor performance of male students in schools. This result is in line with the work of Yusuf and Yakubu (2014) which revealed that female students perform better than male students in geography stating that although there used to be more boys in geography class than girls but girls performed better than boys. This interpretation is also in agreement with study by Jacob and Linus (2017) which revealed that female students exposed to mastery learning strategy performed better than male students exposed to mastery learning. This finding contradicted that of Voyer, Voyer and Bryden (1995) affirmed that male students outperform female students on spatial perception and visualization. Also, against the findings of this study is Gender geography (2010) suggested that a strong masculine bias exists in maps and little is mentioned of women in geography.

It was revealed that the discrimination of the items in the test was low as only two items had the good values for item discrimination items 1a and 1b with 3.315 and 4.364 respectively. Item 3a has negative discrimination which indicated an item where individuals with lower abilities have a higher probability of obtaining a response of 1 than individuals with higher abilities (de Ayala,2009) items that has a discrimination value of 0.4 and above discriminate well while items that has discrimination values of less than 0.4 do not discriminates well. I1B discriminate most followed by I1A, I1D, I1C and I5A1 the other items in the test showed poor discrimination. This is because the higher discrimination values do a better job of discriminating among the examinees than items with smaller discriminations (de Ayala, 2009). This result is in line with Hambleton and Jones (1993) stated that items with high discriminating power contribute more to measurement precision than items with lower discriminating power and that items tend to make their best contribution to measurement precision around their b value on the ability scale.

The study showed that rural students performed better than the urban students this may be as a result of great involvement in urban activities like watching of football, watching of films, internet activities, like the socio media, etc. that the urban students engaged themselves in has resulted in the poor performance of the students. This result is in line with Eugene and Ezeh (2016) in their study found that rural students performed better than the urban students they attributed their performances to the fact that urban students get carried away with social media activities which distract them from academic work, they further found out that from the interview made from the student that they have ICT facilities and electricity regularly which make the students browse, chat or play games with their cellphones or computers while teaching and learning is ongoing in the class. some of the students were of the opinion that the students in rural schools face their studies and study harder because they lack the technologies and other social amenities that distract the students in urban. This result is also in line with Mberekpe, (2013) in his study effect of students improvised instructional materials on senior secondary school student achievement in biology, findings from the study showed that rural students performed better than the urban students in biology when taught using student improvised instructional materials. This result is against Owoeye and Yara (2011) study School Location and Academic Achievement of Secondary School in Ekiti State, Nigeria the researcher found that urban students had better educational achievement than their rural counterparts also Agbaje and Awodun (2014) in their study impact of school location on Academic Achievement of science students in senior secondary school certificate examination found that students in urban performed better than the students in rural school location.

Finding of the result showed that most of the items were invariant except for five items that showed an obvious variation among males and females' samples which indicated that the five items were not invariant among males and females. The reason may be because the test items were constructed based on CTT but the study used IRT to estimate ability. This work is in line with the work of Adedoyin, Nenty and Chilisa (2008) that investigated the invariance of item difficulty parameter estimates based on CTT and IRT. Findings from the study showed that IRT item difficulty parameter estimates were invariant across different independent samples of persons and across varying sample sizes of persons while CTT item difficulty parameter estimates across different independent samples of persons and across varying sample of persons were variant.

Findings of the study showed that 14 items (3, 5, 6, 7, 9, 10, 11, 12, 14, 15, 18, 19, 20, 21 and 23) of the Geography test had difficulty estimates in rural schools sample that were obviously at variance with the items' difficulty estimates in the urban schools sample. The results revealed that the test items appeared to be more difficult for the students of urban sample than for their rural sample counterpart when considered on the average. More importantly, on the individual ground, most of the items showed dissimilar difficulty parameter estimates in the two contrasting samples. The implication of the results is that that the extent of variation of the difficulty of the Physical geography test items with respect to school location was very large. This result is in line with the work of Adedoyin, Nenty and Chilisa (2008) that investigated the invariance of item difficulty parameter estimates based on CTT and IRT. Findings from the study showed that IRT item difficulty parameter estimates were invariant across different independent samples of persons and across varying sample sizes of persons while CTT item difficulty parameter estimates across different independent samples of persons and across varying sample of persons were variant.

Finding from the result showed that there was no significant difference in the difficulty estimates of the Geography test items in the male and female samples. The implication of this result is that the difficulty estimates of the Geography test items were statistically the same in the male and female students' samples. This result is in line with Yusuf and Afolabi's work (2010) on the effects of Computer Assisted Instruction (CAI) on secondary school students' performance in Biology finding from the study revealed that there is no significant difference in male and female performance. Also in line with Okorie and Ezeh (2016) works' influence of gender and location on students' Achievement in chemical Bonding shows that there is no significant difference in male and female performance. This study contradicted the study of Filgona and Sababa (2017) in their study effect of gender on senior secondary school academic achievement in geography where mastery learning strategy and conventional method are used for instruction. The study revealed that achievement of female student improved significantly compared to their male counterpart in the use of mastery learning strategy.

Finding from the result showed that there was no significant difference in the difficulty estimates of the Geography test items in rural and urban schools. This result is in line with Considine and Zappala (2002) found that geographical location do not significantly predict outcomes in performance. Musbau and Johnson (2010) in their study influence of school sex, location and type on student academic performance revealed that school location had no significant influence on student academic performance. This result contradicted Owoeye and Yara (2010) showed significant differences between students' academic achievement in rural and urban schools in senior secondary school certificate examinations. Also researchers like Agbaje and Awodun (2014) works revealed a statistical difference in the achievement of students in rural and urban schools. Also Ogunkola and Fayombo (2009) revealed a statistical difference in favour of urban schools as compared to rural schools. Okorie and Ezeh (2016) study revealed that school location has statistical significant effect on students' achievement in chemistry.

Conclusions

It could be concluded from the findings that the invariance of item parameters of polytomous geography items in senior school certificate examination in Nigeria for 2016 WAEC paper 3 was low among schools located in rural and urban. This was because most of the items (14 items out of 23 items) were not invariant which implies that the test was not good for group comparison.

Implications of the study - The implications of the findings of this study are that it exposes:

1. examination bodies like WAEC to the use of IRT polytomous scoring model and the need to research on other subject areas;
2. examination bodies like WAEC, educational researchers and evaluators to invariance and the need for test items to be invariant;
3. classroom teachers and educational researchers to ability estimation methods in IRT;
4. classroom teachers and educational researchers to the use of polytomous scoring model and the need to use free responses in IRT;
5. the advantages of using generalized partial credit model(GPCM); and
6. teachers, educational researchers to the use of examinee selected items (ESI) which is a way of reducing test anxiety in students.

Recommendations

Based on the findings and conclusions drawn in this study the researcher puts forward the following recommendations.

1. Psychometricians, test developers, teachers and examination bodies should be trained in item response theory framework. This will enable the advantage of the framework and its overall usage to be appreciated and make popular in our locality.
2. Government should encourage our examination bodies such as WAEC, NECO, NABTEB, NTI etc. to adopt IRT measurement framework which will reduce measurement problems encountered like scores comparison, test equating etc.
3. The different IRT software used in analyzing should be made available for student researchers so as to ease research studies in schools.
4. Classroom teachers and lecturers should be trained in psychometric properties of items so as to improve the standard of testing in schools
5. The government, ministries of education and high profile stakeholders in education should procure thevarious IRT analytical software and sponsor the training of individuals to learn the analysis using IRT framework. This way, been made popular and the interpretation of the results and usage of the theory would be simplified.

Limitations of the Study

This study sampled 1546 SS3 geography students of 2016/2017 session but other students could be used as sample. It was carried out in three states from three geopolitical zones but it could be carried out in all the six geopolitical zones using more than one state from each of the geopolitical zone to have a larger sample. An aspect of geography was studied but a whole geography examination which includes human and regional geography, objective questions and the practical (map reading and physical geography) may be studied.

Suggestion for Further Study

This study was carried out to investigate the invariance of item parameters of polytomous geography items in senior school certificate examination in Nigeria. The sample size could be increased by other researchers in future studies. Also, other subject like chemistry, biology, English etc. could be used instead of geography by other researchers in future study.

In this study, generalized partial credit model was used future study may use other polytomous scoring model. This study used the selected examinee response model(ESIM) which deviated from the conventional IRT other researchers in the future could use the IRT which means the test items will be answer all.

REFERENCES

Abiri, J. O. O. (2006). *Elements of evaluation measurement and statistical techniques in education*. Ilorin: Unilorin Press.

Abul, J. N. (2007). *Effect of fieldwork on students' achievement in environmental education contents in senior secondary school geography*. An unpublished master's thesis; University of Nigeria, Nsukka.

Adedoyin O. O. (2010). Investigating the invariance of person parameter estimates based on classical test and item response theories. *International Journal of Education Science, 2(2), 107-113.*

Adeyegbe, S. O. (2002). *How students and examiners perform at WAEC examinations* Vanguard, Thursday, 19th December, pp: 22.

Adeyemi, T. O. (2010). A comparative study of students' academic performance in public examinations in secondary schools in Ondo and Ekiti States, Nigeria. *Current Research Journal of Economic Theory, 3(2), 36-42.*

Adewuni, D. A. (2016). *Hierarchical cluster analysis of the dimensionality of Nigerian senior school certificate examination objective tests in government (2013-2014).* An unpublished Ph.D. thesis; University of Ilorin, Ilorin.

Adonu, I. I. (2014). *Psychometric analysis of WAEC and NECO practical Physics tests using partial credit model.* An unpublished Ph.D. thesis; University of Nigeria, Nsukka.

Agbaje, R. O. & Awodun, A. O. (2013). Impact of school location on academic achievement of science students in senior secondary school certificate examination. *International Journal of Scientific and Research Publications, 4(9), 1-3.*

Allen, M. & Yen, W. (2002). *Introduction to measurement theory*. Long Grove, IL: Waveland Press.

Ali, A. A. (1996). *Fundamentals of Research in Education*. Awka: Meks Publishers.

Alokan, F. B. (2010). Influence of sex and location on relationship between problems and academic performance. *The Social Sciences (TSS), 5(4), 340-345.*

Anastasi, A. & Urbina, A. (2002). *Psychological testing (7th Edition)***.** New Delhi, India: Pearson Education Pvt. Ltd.

Bandele, S. O. (2002). Administration of continuous assessment in tertiary institutions in Nigeria. *Journal of Educational Foundation and Management*, 1(1), 289-296.

Baker, F. B. (2001). *The Basics of Item Response Theory.* ERIC Clearinghouse on Assessment and Evaluation. Retrieved from http://echo.edres.org:8080/irt/baker/.

Baykul, Y. (2000). Eğitimdevepsikolojideölçme: Klasik testteorisiveuygulaması. Ankara: ÖSYM.

Bock, R. D. & Aitkin, M. (1981). Marginal maximum likelihood of item parameters: Application of an EM Algorithm. *Psychometrika, 46(4), 443-459.*

Bjorner, J. B., Smith, K. J., Stone, C. A. & Sun, X. (2007). IRTFIT: *A macro for item fit and local dependence tests under IRT models*. Quality Metric Incorporated. Retrieved from http://outcomes.cancer.gov/areas/measurement/irtfitmacrouser'sguide.pdf.

Bradlow, E. T. & Thomas, N. (1998). Item response theory models applied to data allowing examinee choice. *Journal of Education and Behavioral Statistics*, 23, 236-243.

Brown, T. A. (2006). *Confirmatory Factor Analysis for Applied Research*. New York: The Guilford Press.

Cavanagh, R. F. & Waugh, R. (2011). *Applications of Rasch measurement in learning environments research*. Netherlands: Sense Publishers.

Cantrell, C. E. (1999). *Item response theory: Understanding the one-parameter Rasch model*. Paper presented at the Southwest Educational Research Association, Austin.

Ching-Fung, B. (2002). *Ability Estimation under different item parameterization and scoring models.* An unpublished Ph.D. thesis; University of North Texas, Texas.

Considine, G. & Zappala, G. (2002). The influence of social economic disadvantage in academic performance of school students in Australia *Journal of Sociology* 38, 127-148.

Courville, G. T. (2004). *An empirical comparison of item response theory and classical test theory item/person statistics.* An unpublished Ph.D. thesis; Texas A & M University.

Crocker, L. & Algina, J. (1986). Introduction to classical and modern test theory. *Harcourt Brace Jovanovich College Publishers: Fort Worth, 527 pp.*

Dania, O. P. (2014). Effect of gender on students' academic achievement in secondary school social studies. *Journal of Education and Practice, 5(21)*, 78-85.

De Ayala, R. J. (2009). *The theory and practice of item response theory.* New York: The Guilford Press.

De Champlin, A. F. (2010). A primer on classical test theory and item response theory for assessments in medical education. *Medical Education, 44, 109-117.* Retrieved from doi:10.1111/j.1365-2923.2009.03425.x

De Mars, C. (2010). *Item response theory. Understanding statistics measurement*. City: Oxford University Press.

Donoghue, J. R. (1993). An empirical examination of the IRT information of polytomously scored reading items under the generalized partial credit model. *Journal of Educational Measurement, 31, 295–311.*

Embretson, S. E. & Reise, S. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum Publishers.

Ergüven, M. (2014). An empirical evaluation and comparison of classical test theory and Rasch model. *Journal of Education, 3 (1), 33-38.*

Esere, M. O. & Idowu A. I. (2009). *Continuous assessment practices in Nigerian schools: A review.* University of Ilorin, Nigeria.

Ezeudu, A. S., Olaowei, G. G. & Umeifekwem, E. J. (2014). School location versus academic achievement in Geography: Does reflective inquiry instructional technique has effect? *Indian Journal of Research PARIPEX, 3(9), 209-216.*

Fan, X. (1998). Item response theory and classical test theory: An empirical comparison of their item/person statistics. *Educational and psychological Measurement, 58(3), 357-374.*

Falmagne, J. (1989). A latent trait theory via a stochastic learning theory for a knowledge space. Psychometrika, 54, 283–303.

Federal Republic of Nigeria (2014). *National Policy on Education*. Abuja: Nigeria National Educational Research Council NERC Press.

Field, A. (2005), *Reliability Analysis*. In discovering statistics using SPSS. London.

Filgona, J. & Sababa, L.K. (2017). Effect of gender on senior secondary school students' academic achievement in Geography in Ganye educational zone, Nigeria. *European Journal of Education Studies, 3(4), 394-410.*

Foley, B. P. (2010). *Improving IRT parameter estimates with small sample sizes*: *Evaluating the efficacy of a new data augmentation technique.* An unpublished Ph.D. theses; College of Education and Human Sciences, Paper 75. Retrieved from http://digitalcommons.unl.edu/cehsdiss/75.

Fotaris, P., Mastoras, T., Mavridis, I. & Manitsaris, A. (2011). Identifying potentially flawed items in the context of small sample IRT analysis. *International Journal on Advances in Intelligent Systems, 4, 31–42*. Retrieved from http://www.iariajournals.org/intelligent systems/.

Fox, J. (2010). *Bayesian item response modeling: Theory and applications*. New York: Springer

Furr, R. M. & Bacharach, V. R. (2007). *Item Response Theory and Rasch Models*. In M. Ergüven (2014). An empirical evaluation and comparison of classical test theory and Rasch model. *Journal of Education, 3(1), 33-38.*

Gasper, C. (2010). Allowing examinee choice in educational testing. *Metodoloskizveki, 7(2), 151-166.*

Gao, S. (2011). *The exploration of the relationship between guessing and latent ability in IRT models*. Dissertations. Graduate School Southern Illinois University, Carbondale. Retrieved from http://opensiuc.lib.siu.edu/dissertations.

Glas, C. A. W. (2006). *Bayesian estimation methods for multidimensional models for discrete and continuous responses*. The Netherlands: Law School Admission Council, University of Twente, Enschede.

Gronlund, N. E. (1976). *Measurement and evaluation in teaching.* London: Macmillan Publishing Co.

Guyer, R. & Thompson, N. (2011). *Item response theory parameter recovery using Xcalibre^{TM} 4.1. Technical Report.* Assessment Systems Corporation. Retrieved from www.assess.com/docs/Xcalibre 4.1 tech report.pdf.

Hambleton, R. K. & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practices, 12, 38–47.*

Hambleton, R. K., Swaminathan, H. & Rogers, H. J. (1991). *Fundamentals of Item Response Theory.* (D. S. Foster, Ed.) USA: Sage Publications.

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

Hambleton, R. K., Robin, F. & Xing, D. (2000). Item response models for the analysis of educational and psychological test data. In H. Tinsley & S. Brown (Eds.), Handbook of applied multivariate statistics and modelling. San Diego, CA: Academic Press.

Henson, R. K. (1999). *Understanding the one-parameter Rasch model of item response theory*. Paper presented at the annual meeting of the Southwest Educational Research Association, San Antonio.

Hooper, D. J., Coughlan, & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *The Electronic Journal of Business Research Methods, 6, 53-60.* Retrieved from www.ejbrm.com.

Hoijtink, H. & Glockner-Rist, A. (2003). *The best of both worlds: factor analysis of dichotomous data using item response theory and structural equation modeling*, 10(4), 544–565.

Horn, J. L. & Mc Ardle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research, 18, 117-144.*

Iwena, O. A. (2012). *Essential Geography for senior secondary schools* (6th edition). Anambra: Tonad Publishers Ltd.

Johnson, M. S. (2007). Marginal Maximum Likelihood Estimation of Item Response Models in *R. Journal of Statistical Software, 20(10), 1-24.* Retrieved from http://www.jstatsoft.org/

Kerlinger, F. N. & Lee, H. B. (2000). *Foundations of Behavioural Research*. Philadelphia: Harcourt College Publishers.

Kim, E. S. & Myeongsun, Y. (2011). Testing measurement invariance: A comparison of multiple-group categorical CFA and IRT, structural equation modeling: *A*

Multidisciplinary Journal, 18(2), 212-228. Retrieved from DOI:10.1080/10705511.2011.557337.

Kim, S. -H. (2007). Some posterior standard deviations in item response theory. *Educational and Psychological Measurement, 67, 258-279.*

Kim, J. K. & Nicewander, W. A. (1993). Ability estimation for conventional tests. *Psychometrica Journal of Quantitative Psychology, 58(4), 587-599.*

Korobko, O. B., Glas, C. A. W., Bosker, R. J. & Luyten, J. W. (2008). Comparing the difficulty of examination subjects with item response theory. *Journal of Educational Measurement, 45(2), 139–157.*

Le, D. -T. (2013). *Applying item response theory modeling in educational research*. An unpublished Graduate Theses and Dissertations. Retrieved from http://lib.dr.iastate.edu/etd.

Liu, C. & Wang, W. (2017). Non-ignorable missingness item response theory models for choice effect in examinee selected items. *British Journal of Mathematical and Statistical Psychology, 70, 499–524.*

MacDonald, P. & Paunonen, S. V. (2002). A Monte Carlo comparison of item and person parameters based on item response theory versus classical test theory. *Educational and Psychological Measurement, 62, 921–943.*

Magno, C. (2009). Demonstrating the difference between classical test theory and item response theory using derived test data. *The International Journal of Educational and Psychological Assessment, 1(1), 1-11.*

Maxwell, M. (2013). *Using Bayesian techniques with item response theory to analyze Mathematics tests.* A Ph.D. thesis; University of Alabama, USA.

Mberekpe, A. C. (2013). *Effect of students improvised instructional materials on senior secondary school students' achievement in Biology.* An M.Ed. project, University of Nigeria, Nsukka.

McBride, N. (2001). *An item response theory analysis of the scales from the international personality item pool and the NEO personality inventory*-revised. An M.Sc. project. Virginia Polytechnic Institute and State University.

Meng, H. (2007). *A comparison study of IRT calibration methods for mixed-format tests in vertical scaling.* A Ph.D. thesis; University of Iowa. Retrieved from http://ir.uiowa.edu/etd/338.

Mislevy, R. J. & Bock, R. D. (1990). *Item analysis and test scoring with binary logistic models:* computer software. Chicago: Scientific Software.

Musibau, A. Y. & Johnson, T. A. (2010). The influence of school sex, location and type on students' academic performance. *International Journal of Edu. Sci., 2(2), 81-85.*

Muthen, L. K. & Muthen, B. O. (2006). *Mplus user's guide. Fourth editions.* Los Angeles, CA: Muthen & Muthen.

Nenty, H. J. (2004). *From classical test theory (CTT) to item response theory (IRT): An introduction to a desirable transition*. In: O. A. Afemikhe, J. G. Adewale (Eds.): Issues in Educational Measurement and Evaluation in Nigeria. pp. 372-384. Retrieved from http://www.academia.edu/15263615.

Nkpone, H. L. (2001). *Application of latent trait models in the Development and Standardization of a Physics Achievement test for senior secondary schools*. An unpublished Ph.D. thesis; University of Nigeria, Nsukka.

Nwana, O. C. (1979). *Measurement and Evaluation for Teachers. Objectives in Education*. Nelson Africa. Hong Kong.

Odili, J. N. (2009). Demonstrating the difference between classical test theory and item response theory using derived test data. *The International Journal of Educational and Psychological Assessment, 1(1), 1-11.*

Odili, J. N. (2013). Reforming educational measurement for national security: The role of item response theory in test fairness. *China Education Review 3(12), 924-932.*

Ogunkola, B. J. & Fayombo, G. A. (2009). Investigating the combined and relative effects of some students related variables on science achievement among secondary school students in Barbados. *European Journal of Scientific Research, 37(3), 481-489.*

Ojerinde, D. (2013). *Classical test theory (CTT) vs item response theory (IRT): An evaluation of the comparability of item analysis results.* Lecture Presentation at the Institute of Education University of Ibadan.

Okafor. G. A. (2015). Essay test as a predictor of students' achievement in objective test in senior secondary school certificate Geography. *International Journal of Scientific & Engineering Research, 6(11), 598-600.* Retrieved from http://www.ijser.org.

Okorie, E. U. & Ezeh, D. N. (2016). Influence of gender and location on students' achievement in chemical bonding. *Mediterranean Journal of Social Sciences, 7(3), 309-318.*

Oludipe, D. I. (2012). Gender difference in Nigerian junior secondary students' academic achievement in Basic Science. *Journal of Educational and Social Research, 2(1), 93-99. Retrieved from DOI:10.5901/jesr.2012.02.01.93.*

Omotoso, D. (1992). *How to make 3-3 secondary school component of the new system of education work in Ondo State.* Paper presented to the All Nigeria Conference of principal of secondary schools, pp; 2-12.

Onipede, H. (2003). *National Development Hinges on Quality Education.* The Comet, Thursday, 2nd January pp: 21.

Onwuka, U. (1981). *Curriculum Development for Africa*. Africana FEP Publishers Ltd. Nigeria.

Osunde, (1997). Measurement of Educational Objectives. In S. A. Makinde, (ed.) Educational theory and practice in Nigeria. Lagos: New Era Publications 67-73.

Owoeye, J. S. & Yara, P. O. (2011). School location and academic achievement of secondary school in Ekiti State, Nigeria. *Asian Social Science, 7(5), 170-175.* Retrieved from doi.10.5539/ass.v7n5p170.

Oyesola, S. E. (1986). *Guidance for 6-3-3-4 system of education*: A new approach. Ibadan: University Press Ltd.

Progar1, Š. & Sočan, G. (2008). An empirical comparison of item response theory and classical test theory. *Psihološkaobzorja / Horizons of Psychology, 17*(3)*, 5-24.*

Rupp, A. A. & Zumbo, B. D. (2013). A note on how to quantify and report whether IRT parameter invariance holds: When Pearson correlations are not enough. Educational and Psychological Measurement *65*(4*)*, 588-599. Doi: 10.117/0013164403261051

Rupp. A. A & Zumbo, B. D. (2006). Understanding Parameter Invariance in Unidimensional IRT Models. Educational and Psychological Measurement *66*(1), 63-84. Doi: 10.1177/0013164404273942

Rust, J. & Golombok, S. (1999). *Modern psychometrics: the science of psychological assessment*. Routledge, 2nd Edition

Rowley, G. L. & Traub, R. E. (1977). Formula scoring, number right scoring, and test-taking strategy. *Journal of Educational Measurement, 14(1), 15-22.*

Samejima, F. (1976). The *Graded response model of the latent trait theory and tailored testing***.** In C. L, Clark (Ed.), proceedings of the first Conference on Computerized Adaptive Testing (Personnel Research and Development Center, PS-75-6) (pp. 5-15). Washington, DC

Thissen, D. (2000). *Reliability and measurement precision.* In H. Wainer (Ed.), Computer adaptive testing: A primer (2nd Ed., pp. 159-184). Westport, CT: Praeger Publishers.

Thompson, N. A. (2009). *Ability estimation with item response theory*. Minnesota: Assessment Systems Corporation.

Timminga, E. (1995). Optimum examinee samples for item parameter estimation in item response theory: A multi-objective programming approach. *Psychometrika, 60, 137-154.*

Traub, R. (1997). Classical test theory in historical perspective. *Educational Measurement: Issues and Practice, 16, 8-14.*

Tsung-Han, H. O. (2010). *A comparison of item selection procedures using different ability estimation methods in computerized adaptive testing based on the generalized partial credit model.* An unpublished Ph.D. thesis; University of Texas, Austin.

Van der L. W. J., Klein, E. R. H. & Fox, J. –P. (2010). IRT parameter estimation with response times as collateral information. *Applied Psychological Measurement, 34(5), 327–34.* Retrieved from DOI: 10.1177/0146621609349800 http://apm.sagepub.com.

Varma, S. (2013). *Preliminary item statistics using point Biserial correlation and p-values.* Retrieved from Educational Data Systems: http://www.eddata.com/resources/publications/EDS_Point_Biserial.pdf.

Voyer, D. & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin, 140(4), 1174-1204.* Retrieved from http://dx.doi.org/10.1037/a0036620.

Wang, W. & Wilson, M. (2005). Exploring local item dependence using a random-effects facet model. *Applied Psychological Measurement, 29, 296–318.*

Wainer, H. & Thissen, D. (1994). On examinee choice in educational testing. *Review of Educational Research,Spring 64(1), 159-195*.

Wen, C., Jin, K., Qui, K. & Wang, L. (2012). Item response models for examinee-selected items. *Journal of Educational Measurement, 49(4), 419-445.*

Wilberg, M. (2004). *Classical test theory vs. item response theory*: *An evaluation of the theory test in the Swedish driving –license test. Universitet UMEA.*

Yen, W. M. & Fitzpatrick, A. R. (2006). Item response theory. In R. L. Brennan (Ed.) *Educational Measurement*, (4th Edition, 111–153). Westport, CT: Praeger Publishers.

Yi-Ting, H. & MengYeali C. S .S. (n.d.). *An interpretable statistical ability estimation in web-based learning environment*. National Taiwan University, Taiwan. pp. 230-237 Retrieved from *d97008@im.ntu.edu.tw.

Yoloye, T. W. (2004). *That we may learn better*. Ibadan: Stirling Hordon Publishers Ltd.

Yusuf, M. O. & Afolabi, A. O. (2010). Effects of computer assisted instruction (CAI) on secondary school students' performance in Biology. *The Turkish Online Journal of Educational Technology, 9(1), 62-69.*

Yusuf, Y. O. & Yakub, K. O. (2013). Comparative study of gender difference performance in Geography: A case study of some selected schools in Zaria Inspectorate Division of Kaduna, Nigeria. *Educational Research and Review, 8(5), 179-185.* Retrieved from DOI:10.5897/ERR2013.1421.

Zembar, M. J. & Blume, L. B. (2009). Excerpt from Middle Childhood Development: A Contextual Approach p. 212-215.

Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF)): Logistic regression modeling as a unitary framework for binary.* Ottawa, Canada.

APPENDIX I

PHYSICAL GEOGRAPHY ACHIEVEMENT TEST (PGAT)

ELEMENT OF PRATICAL AND PHYSICAL GEOGRAPHY

Answer questions one and any three others.

1. Study the map extract provided on a scale of 1.50,000 and answer the following questions.
   a. In your answer booklet, draw the outline of the mapped area to a scale of 1.150,00.
   b. In your new outline, insert and name;
      i. The main road from UGEP junction to the north- western end of the map
      ii. CROSS RIVER
      iii. 100 *ft* contour line in the south – eastern end of the map.
   c. Describe the drainage of the mapped area.
   d. Using evidences from the map, identify any **two** possible economic activities in the mapped area.
2. Use the data in the table below to answer the questions that follow:

   Country B's population figures (in millions) by State in the year 2005 .

State	Population (millions)
A	40
B	35
C	25
D	50
TOTAL	150

a. construct a pie chart to represent the data

b. Highlight any **two** problem in the use of pie charts

c. State **one** other method that could be used to represent the above information.

3. a**.** Highlight **three** differences between plutonic and volcanic rocks.

   b. With the aid of diagrams, describe the mode of formation of Crater Lake.

4. a. State three characteristics of limestone region.

   b. Describe the formation of limestone pillar**.**

   c. In what **two** ways is limestone important to man?

5. a. i. Highlight **two** ways in which water resources are classified.

   ii. In what three ways can water resources be used?

   iii. Describe three ways by which man's activities limit the use of water as a resources.

6. a. Define the term *climate*

   b. Name the instrument used in measuring each of the following climate elements:

   i. Temperature;

   ii. Pressure;

   iii. Rainfall;

   iv. Sunshine;

   v. Humidity.

   c. In what **four** ways is the way study of weather and climate important to man?

7. a. i. Define the term Solar System

   ii. List any **five** components of the solar system.

b. Outline any **four** characteristics of the earth.

8. a. Outline **three** differences between *deflation* and *abrasion.*

b. Using specific examples, describe the mode of formation of a deflation hollow.

APPENDIX II

SCHEME1

(a) **New outline**

Length- 12.3 cm (± 0.2cm) = (12.1cm- 12.5cm) = 2 marks

Breadth- 12.3 cm (± 0.2cm) = (12.1cm – 12.5cm) = 2 marks

New scale- 1:150,000 or 1/150,000 = 2 marks

(b) **Insertions and naming**

(i) The main road at 3 marks = 3 marks

(Correct insertion: beginning and end = 1 mark; alignment = 1 mark; naming = 1 mark)

(ii) CROSS RIVER at 3 marks = 3 marks

(Correct insertion: beginning and end = 1 mark; alignment = 1 mark; naming = 1 marks)

(iii) 100ft contour line at 3 marks = 3 marks

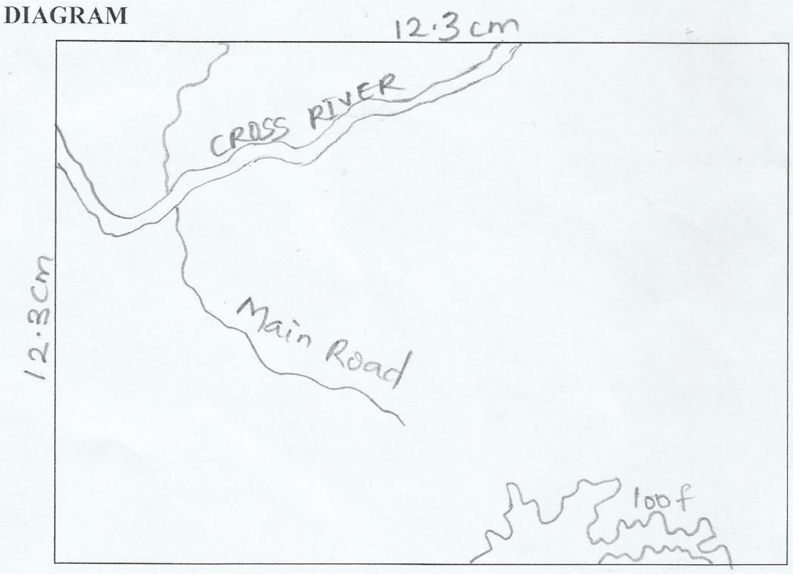
(Correct insertion: beginning and end = 1 mark; alignment = 1 mark; naming 1 marks)

(c) **Description of drainage of the mapped area**

- Well drained
- Major river is CROSS RIVER
- CROSS RIVER is at the lower stage
- Drainage density of the area is high
- North-western corner of the area is liable to flooding
- The major river is located in the north western part of the map
- Some rivers like River Ekuru flow southwards
- North flowing rivers drain into the CROSS RIVER
- Watershed at the centre of the maps
- Major tributary is CROSS RIVER is river ASU
- Drainage pattern is dendritic
- Some of the water courses are intermittent

Any 3 points well explained at 2 marks each = 6 marks

DIAGRAM



(d) **Possible economic activities in the mapped area**

- Farming due to presence of scattered cultivation
- Lumbering due to presence of forest
- Fishing due to presence of rivers
- Transport services due to presence of roads and rivers
- Plantation agriculture due presence of ENDC rubber Estate

Any 2 points well explained at 2 marks each = 4 marks
Mere listing = 1 mark each.
(a) = 6 marks; (b) = 9 marks; (c) marks; (d) = marks; Total= 25 marks.

SCHEME 2

(a) Construction of pie chart
Calculate of sector valve
State A= 40/150 x 360 = 96^o =1 mark (calculation = + mark, answer = + mark)
State B= 35/150 x 360 = 84^0= 1 mark (calculation = +mark, answer = + mark)
State C = 25/150 x 360 = 60^0= 1 mark (calculation = +mark, answer = + mark)
State D = 50/150 x 360 =120^0 = 1 mark (calculation = + mark, answer = + mark)

SCHEME 2 a

A pie chart representing 2005 population figure of country B in millions

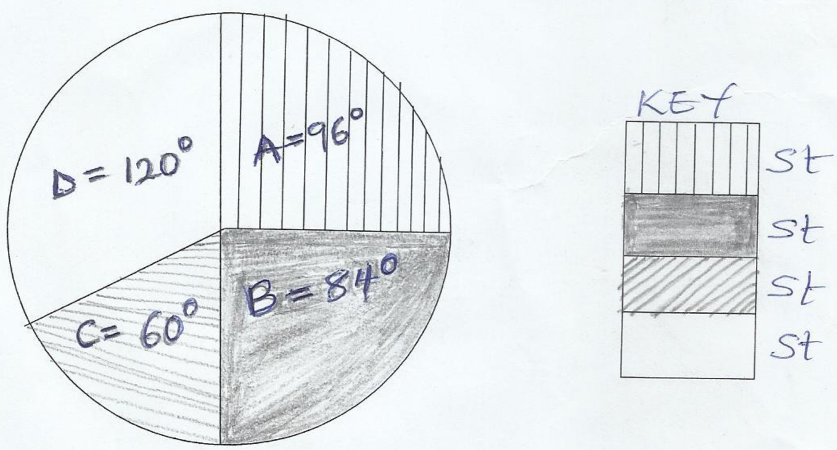


Diagram = 4 marks (4 sectors at 1 mark each = 4 marks)
Title = 1 mark

Diagram = 4 marks (4 sectors at 1 mark each = 4 marks)

Title = 1 mark

(b) **Problems in the use of pie charts**

- Can only show one distribution
- Comparison of sectors is difficult
- Mathematical calculate is tasking
- Construction is difficult if sectors are many
- Angular measurements are difficult
- Vision is blurred when too many items are involved
- Difficulty in determining actual figures represented

(c) Other methods that could be used to represent the information

- Simple bar chart
- Percentage bar charts
- Divided /component bar chart
- Pictogram

Any 1 point at 2 marks = 2 marks

(a) = 9 marks; (b) = 4 marks; (c) = 2 marks; total = 15 marks.

<u>SCHEME 3</u>

(a) Differences between plutonic and volcanic rocks

- Plutonic rocks are formed beneath or within the earth's crust while volcanic rocks are formed on the surface of the earth
- Plutonic rocks have large crystals while volcanic rocks have small crystals
- Plutonic rock are coarse grained while volcanic rocks are fine grained
- Plutonic rocks are also called intrusive rocks while volcanic rocks are also called extrusive rocks
- Plutonic rocks are mostly basic rocks while volcanic rocks are acidic rocks

- During the process of formation, cooling is slower for plutonic rocks while it is faster for volcanic rocks
- Landforms associated with plutonic rocks are batholiths, sills and dykes while those associated with volcanic rocks are composite cones, lava plateaux, etc.
- Examples of plutonic rocks includes, granites, diorites and gabbro; while examples of volcanic rocks includes basalt, rhyolite andesite and obsidian.

Any 3 points well differentiated at 2 marks each = 6 marks

N:B If tabulated mark out of half

b) Mode of formation of crater lakes

- Weakness develops in earth crust
- Molten magma is trapped within the earth crust
- Cracks and joints develop in the area of crustal weakness
- Magma under pressure is ejected through a vent
- Magma is composed of ash and cinder
- A cone is built around the vent by the materials
- The vent is blocked
- Subsequent eruption blows off the top of the cone
- A depression is formed
- The development is called crater
- Rain water/melting ice collects in the crater
- This forms a crater lake
- Subsequent forceful eruptions blow off the top cone and crater
- A larger depression called caldera is created
- Water collects in the caldera to form a caldera lake
- E.g. Lake Payam in Jos Plateau, Nigeria; Lakes Bosumtwi near Kumasi, Ghana lake in Australia; Lakes Shala in Ethiopia; Lake Toba in Sumatra, Lake Ore, USA

Any 7 points at 1 mark each each = 7 marks

2 diagrams at 1 mark each = 2 marks

(a) = 6 marks; (b) = 9 marks; Total = 15 marks

SCHEME 4

a) Characteristics of limestone's regions

- Rocks in the region are sedimentary
- The rocks are of organic origin
- Absence of luxuriant vegetation
- Presence of depression/gorges
- Presence of valleys
- Solubility of rock materials
- Presence of scrubs and grassland vegetation
- Presence of underground drainage
- Presence of caves
- Presence of pillars, stalactites and stalagmites, etc
- Presence of sink holes and swallow holes
- Absence of surface drainage
- Presence of rugged topography
- Jointed rocks
- Stream disappear into swallow holes
- Presence of resurgence streams
- Presence of grikes, clints and limestone pavement
- Examples are found in Yugoslavia, Ewekoro, Nkalagu, Sokoto, Nauli etc.

Any 3 points well explained at 2 marks each = 6 markss

Mere listed = 1 mark each

b) Formation of limestone pillar

- Formed in limestone regions
- Formed within a cave
- Rain water combined with atmospheric cardon dioxide (CO_2) to form weak carbonic acid
- Acid water dissolves limestone
- Dissolution causes calcium bicarbonate
- The dissolution process wears out materials along limestone joint

- joints widen and deepen to make rivers drain underground
- intensification of rock dissolution leads to the formation of underground caves
- water drips into caves from roof of caves
- water evaporates leaving behind calcium carbonate residue
- calcium carbonate residues hang from roofs of caves as stalactites
- some calcium carbonate residues grow upwards from the floors of caves as stalagmites
- continuous growth of stalactites and stalagmites join together to form new features called limestone pillars

Any 5 points at 1 mark each= 5marks

c) Ways in which limestone is important to man

- Raw materials for cement industry
- used as alloy in smelting tin and iron
- provides grazing land
- Source of underground water
- Beautiful scenery for tourism
- Portions of collapsed gorges and caverns are fertile
- provides shelter for refuge in times of emergency

Any 2 points well explained at 2marks each =4 marks

Mere listing =1 mark each

a) =6 marks; b) =5marks; c) =4 marks; Total =15 marks

<u>SCHEME 5</u>

a) (i) **Ways in which water resources are classified**

- According to location =underground vs. surface water resources
- According to composition = salt Vs. fresh water
- According to size = e.g. ocean, seas, lakes, ponds, rivers, springs, stream etc
- According to state = stagnant vs. running water

Any 2 classes or types at 1+ mark each =3 marks

(ii) **Ways in which water resources can be used**

- sporting
- irrigation
- sources of sea food
- domestic uses
- transportation
- H.E.P generation
- Industrial use
- Recreation
- Tourism
- Sources of minerals
- Research
- boundaries

Any 3 points well explained at 2marks each =6 marks

Mere listing = 1 mark each

(ii) **Ways by which man's activities limit the use of water resources**

- discharge of industrial waste into water bodies
- fishing with chemicals
- oil spillage
- mining activities
- waste disposal
- application of chemical fertilizers
- discharge of toxic substances into water bodies
- sharing of same water source with animals
- damming of rivers up stream
- dumping untreated sewage inside water bodies
- discharge of effluent from ships inside water

Any 3 points well explained at 2marks each =6 marks

Mere listing = 1 mark each

a) = 9 marks ; b) = 6 marks Total =15 marks

SCHEME 6

(a) Definition of climate

- The average weather condition of a place over a long period of time usually 30 to 40 years. 2 marks

(b) Instruments for measuring climate elements

- Thermometer - temperature
- Barometer - pressure
- Rain gauge - rainfall
- Sun dial/sun shine recorder/ Campbell stoke - sunshine
- Hygrometer - humidity

5 points well matched at 1 mark each = 5 marks

(C) Ways in which the study of weather and climate is important to man

- Human settlement
- Health
- Environmental hazards
- Clothing
- Vegetation
- Agriculture
- Cultural practices
- Soil formulation
- Transportation
- Communication
- Building types
- Weather forecast
- planning of activities
- Sporting activities

Any 4 points well explained at 2marks each=8marks

Mere listing=1 mark each

a) =2marks; b)=5marks; c)= 8marks. Total=15marks

SCHEME 7

a) i.) Definition of Solar system

The solar system consists of the sun and the nine planets with their satellites.

=2 marks

ii.) **Components of the solar system**

- the sun
- Mercury
- Venus
- Earth
- Jupiter
- Saturn
- Uranus
- Neptune
- Pluto
- Satellites
- Asteroids
- Meteors

Any 5 points at 1 mark each=5marks

b.) **Characteristics of the Earth**

- Spherical or geoid
- Has more oxygen than other planets
- Only planet which supports life
- Has one natural satellite
- Is the 5th largest planet

- rotates on its axis
- rotates eastwards(from west to east)
- revolves round the sun
- completes a rotation in 24 hours
- Completes a revolution in 365 1/5 days
- Has different seasons
- The third planet from the sun
- The surface is made up of 29% land and 71% water
- Is about 148.8 million km away from the sun
- Has outer solid crust
- Possesses an inner molten core
- Surrounded by gaseous envelope
- The zone of life on the earth is at the interface of hydrosphere, lithosphere, lithosphere and the atmosphere
- Axis is inclined at angle of $661/2^0$
- Has equatorial circumference of 40,085 km
- Has polar circumference of 39,955km
- Has equatorial diameter of 12,762 km
- Is the most dense planet
- Has polar diameter of 12,722km

Any 4 points well explained at 2 marks each = 8 marks

a) =7; b) =8 Total =15marks

SCHEME 8

(a) **Differences between deflation and abrasion**

- Deflation involves blowing away of loose rock wastes while abrasion involves hauling rocks against rock surfaces

- Deflation involves rolling loose rock materials along the ground where as abrasion involves polishing, scratching and wearing away of rock surfaces.

- Deflation usually results in lowering of land surfaces while abrasion is most effective at the base of rocks

- Deflation is associated with winds while abrasion is associated with winds, water, and waves

- Deflation produces land forms such as deflation hollows and oasis, while abrasion produces landforms such as rock pedestal, zeugens, yardangs etc.

- Abrasion occurs everywhere while deflation occurs only in arid and semi-arid regions

Any 3 points well differentiated at 2 marks each = 6 marks

NB: if tabulated mark out of half

(a) Mode of formation of deflation hollow

- A feature of arid region

- A producer of wind deflation

- A basin or saucer-shaped desert landform

- Wind exerts frictional force on loose rocks particles on the surface

- Involves the lifting of loose materials from the ground

- It also involves scooping away of loose sandy materials

- The loose materials lifted are blown away

- This action consequently lowers the land surface

- Wind eddies help to deepen and enlarge the hollows particularly in soft rock

- The enlarged hollows are called depressions or deflation hollows

- They vary in size from some few meters to several kilometres

- The depressions may reach water table to produce Oasis or swamps

- E.g. Faiyum in Egypt, Quattarra in Egypt, Baharia in Egypt,Farafra in Egypt, Dakhla in Egypt, Wyoming in USA.

Any 7 points at 1 mark each = 7 marks

Any 2 examples at 1 mark each = 2 marks

a) = 6 marks; b) =9 marks. Total = 15 marks

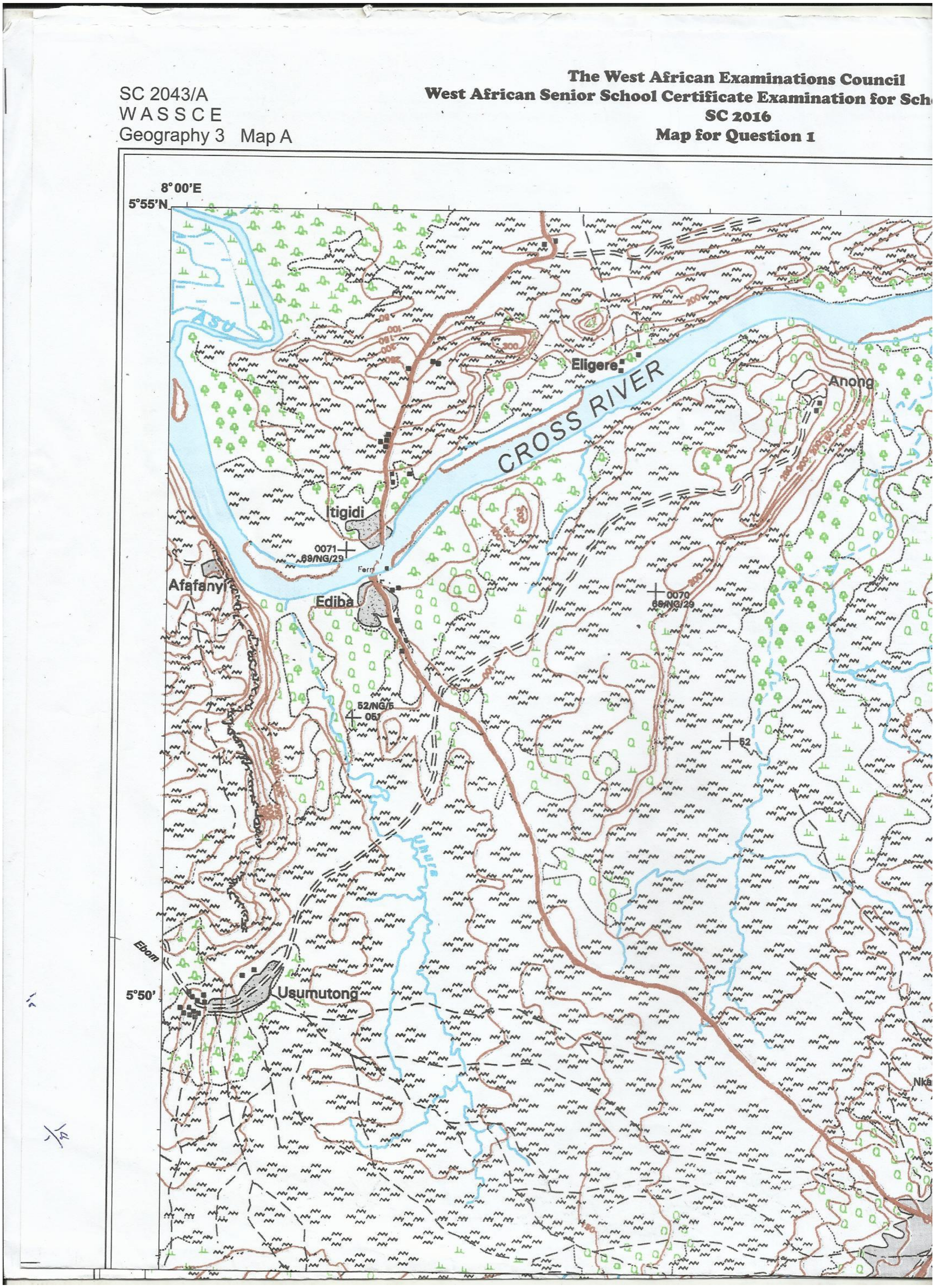
APPENDIX III

ANSWER SHEET

Name of School……………………………………………………………………………………..

Sex: Male( ) Female ( )   Age………………………………………..

APPENDIX IV

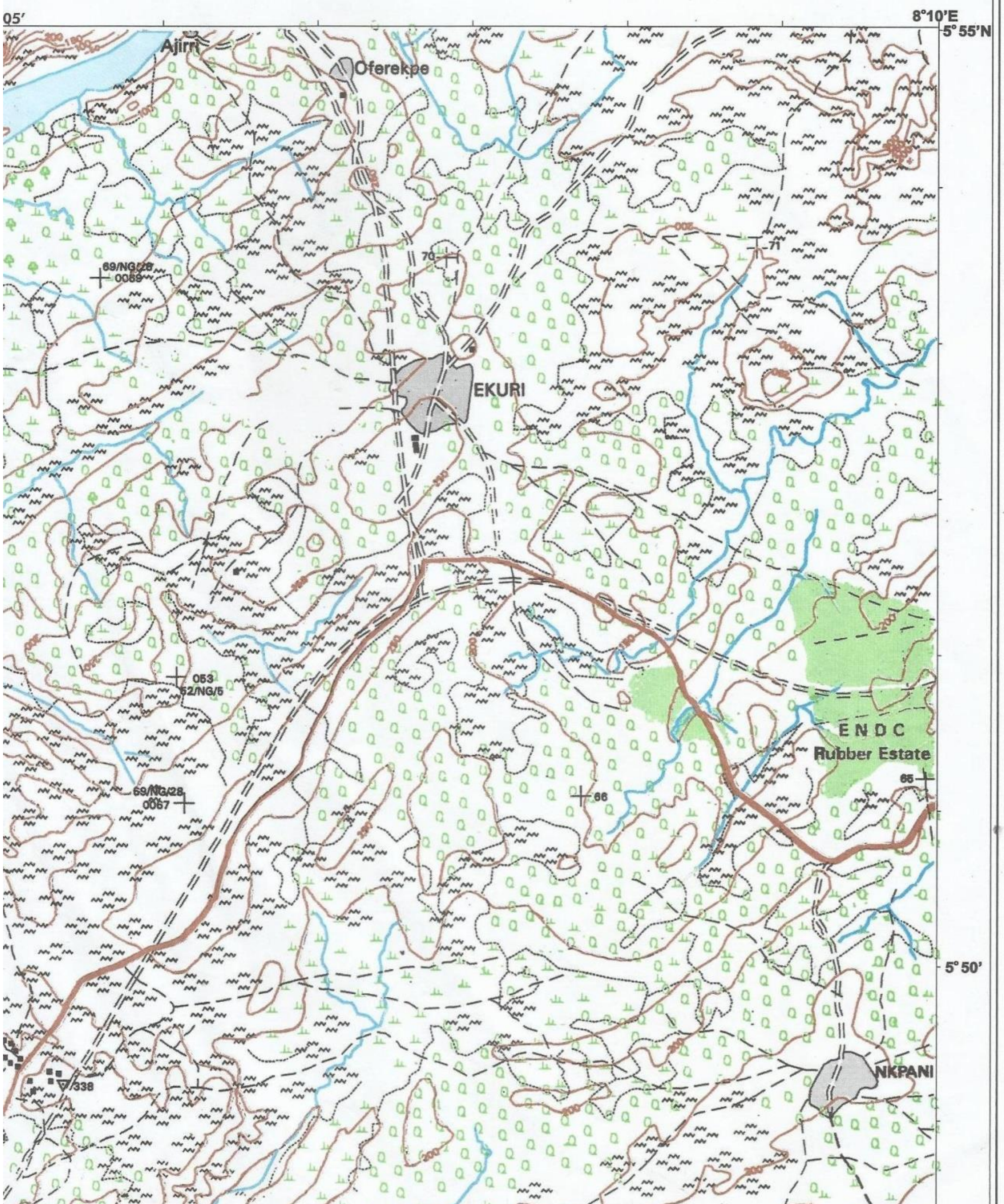


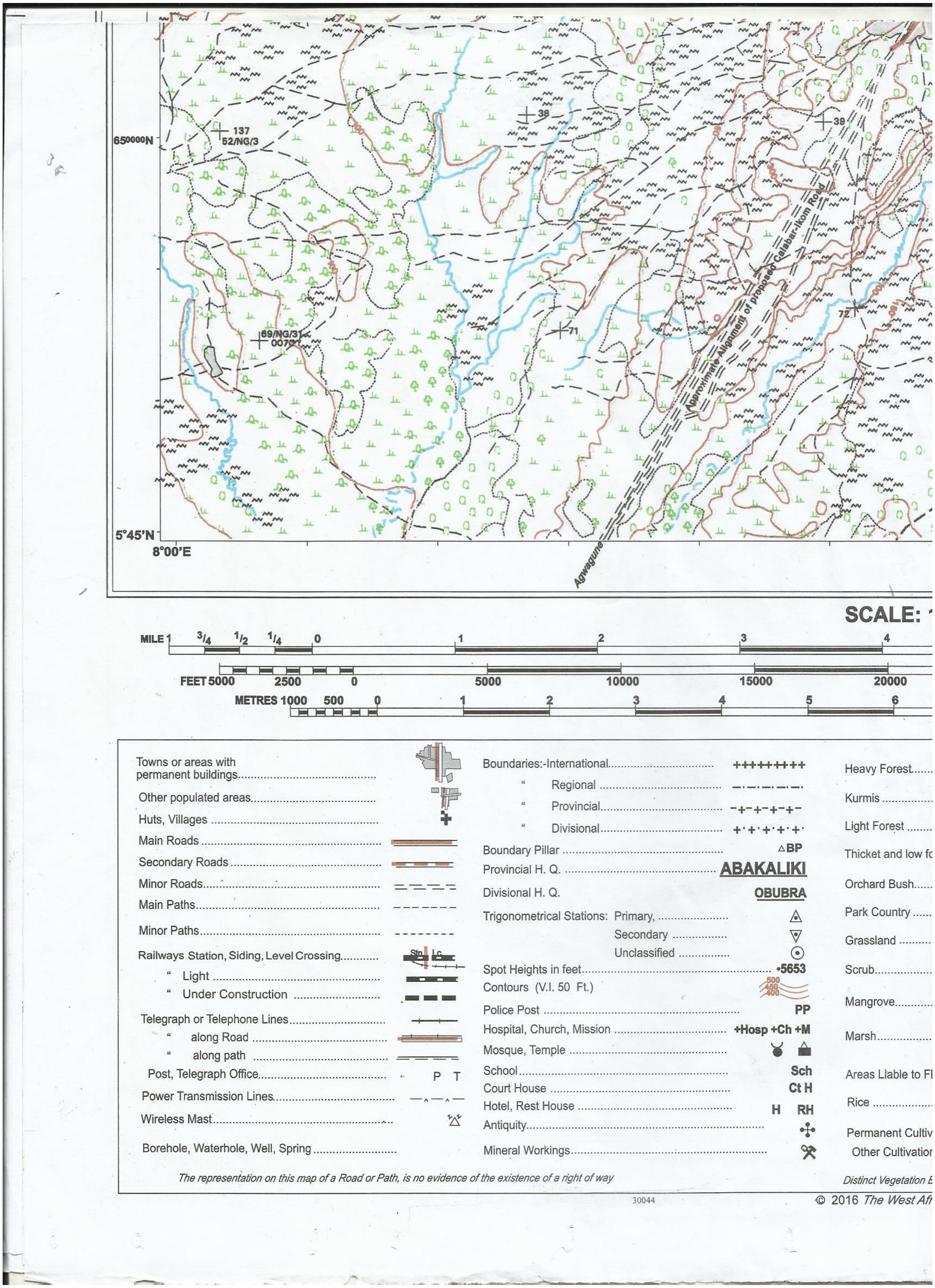
SC 2043/A
W A S S C E
Geography 3 Map A
The West African Examinations Council
West African Senior School Certificate Examination for Sch
SC 2016
Map for Question 1
8° 00'E
5°55'N
ASU
CROSS RIVER
Eligere
Anong
Itigidi
0071
89/NG/29
Afafanyi
Ediba
0070
89/NG/29
52/NG/5
052
52
Ebom
5°50'
Usumutong
Nka

Candidates

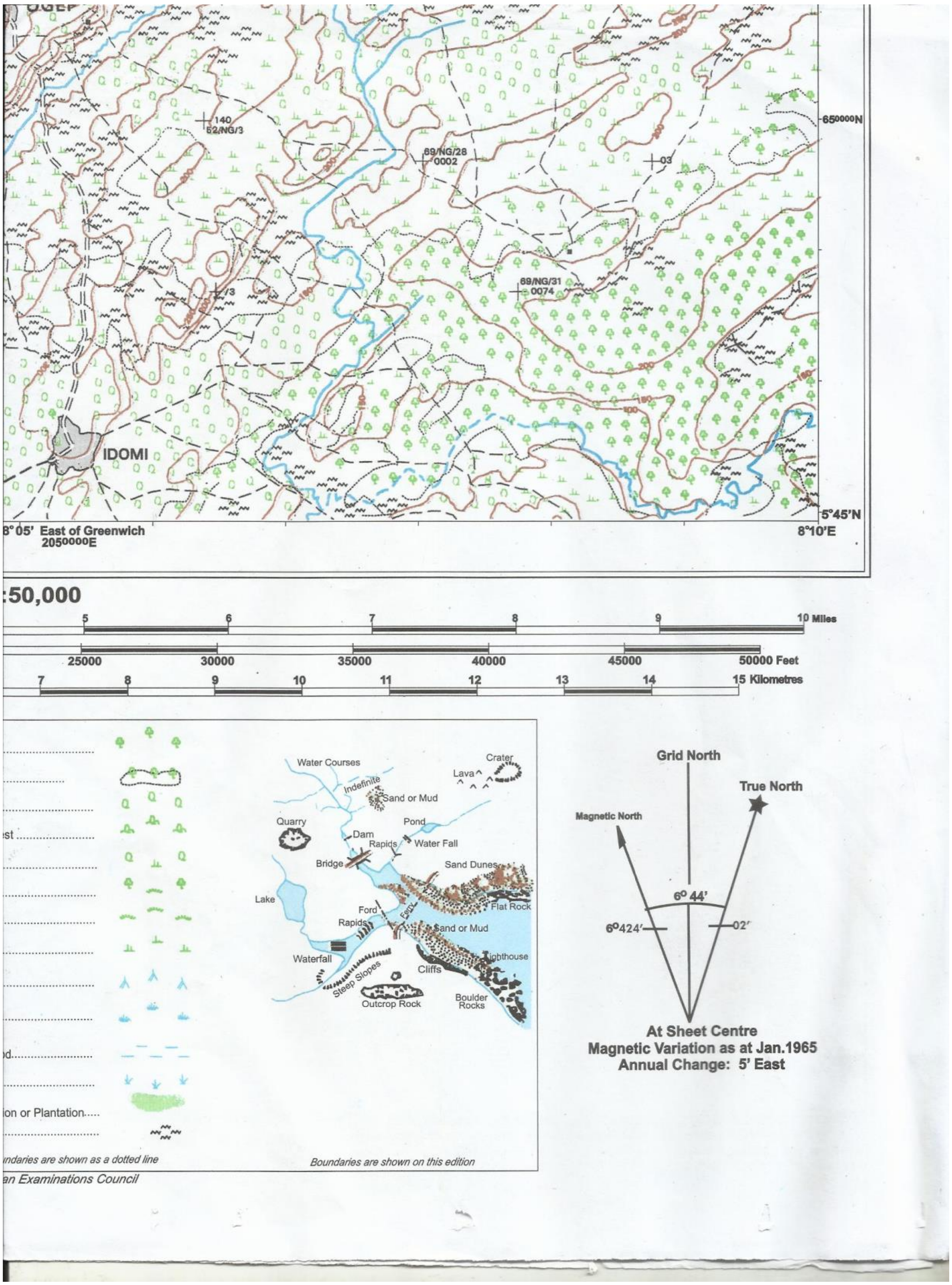
NOTE

Each box should contain a colour. If any box is white, it means that the map is incomplete and should be returned to the invigilator for replacement.





137
52/NG/3
38
39
650000N
69/NG/31
0079
71
72
Approximate Alignment of proposed Calabar-Ikom Road
5°45'N
8°00'E
Agwagune
SCALE:
MILE 1 3/4 1/2 1/4 0 1 2 3 4
FEET 5000 2500 0 5000 10000 15000 20000
METRES 1000 500 0 1 2 3 4 5 6
Towns or areas with permanent buildings
Other populated areas
Huts, Villages
Main Roads
Secondary Roads
Minor Roads
Main Paths
Minor Paths
Railways Station, Siding, Level Crossing
" Light
" Under Construction
Telegraph or Telephone Lines
" along Road
" along path
Post, Telegraph Office
P T
Power Transmission Lines
Wireless Mast
Borehole, Waterhole, Well, Spring
Boundaries:-International
" Regional
" Provincial
" Divisional
Boundary Pillar
BP
Provincial H. Q.
ABAKALIKI
Divisional H. Q.
OBUBRA
Trigonometrical Stations: Primary,
Secondary
Unclassified
Spot Heights in feet
•5653
Contours (V.I. 50 Ft.)
500
450
400
Police Post
PP
Hospital, Church, Mission
+Hosp +Ch +M
Mosque, Temple
School
Sch
Court House
Ct H
Hotel, Rest House
H RH
Antiquity
Mineral Workings
Heavy Forest
Kurmis
Light Forest
Thicket and low fo
Orchard Bush
Park Country
Grassland
Scrub
Mangrove
Marsh
Areas Liable to Fl
Rice
Permanent Cultiv
Other Cultivation
The representation on this map of a Road or Path, is no evidence of the existence of a right of way
Distinct Vegetation B
30044
© 2016 The West Afr



IDOMI
140
62/NG/3
69/NG/28
0002
69/NG/31
0074
650000N
5°45'N
8° 05' East of Greenwich
2050000E
8°10'E
:50,000
10 Miles
25000
30000
35000
40000
45000
50000 Feet
15 Kilometres
ion or Plantation
ındaries are shown as a dotted line
an Examinations Council
Water Courses
Indefinite
Sand or Mud
Crater
Lava
Quarry
Dam
Pond
Rapids
Water Fall
Bridge
Sand Dunes
Lake
Flat Rock
Ford
Rapids
Ferry
Sand or Mud
Waterfall
Cliffs
Lighthouse
Steep Slopes
Outcrop Rock
Boulder Rocks
Boundaries are shown on this edition
Grid North
True North
Magnetic North
6° 44'
6°424'
02'
At Sheet Centre
Magnetic Variation as at Jan.1965
Annual Change: 5' East

APPENDIX V

EXAMINEE ABILITY ESTIMATES

S/N	Ability	std Error
	1.284	0.228
2	1.305	0.234
3	1.353	0.245
4	1.304	0.233
5	1.881	0.376
6	0.175	0.197
7	1.919	0.386
8	1.629	0.312
9	1.484	0.277
10	0.955	0.161
11	0.218	0.186
12	0.236	0.181
13	0.212	0.187
14	0.236	0.181
15	0.236	0.181
16	0.236	0.181
17	0.236	0.181
18	0.228	0.183
19	0.204	0.189
20	0.229	0.182
21	0.219	0.186
22	-0.176	0.325
23	0.236	0.181
24	0.042	0.239
25	-0.196	0.331
26	-1.245	0.661
27	0.211	0.188
28	-0.197	0.329
29	-0.116	0.297
30	0.219	0.185
31	0.208	0.188
32	0.196	0.191
33	0.218	0.186
34	0.221	0.185
35	0.047	0.238
36	0.471	0.14
37	-0.742	0.54

S/N	Ability	std Error
38	-0.014	0.259
39	-0.075	0.283
40	0.104	0.219
41	0.104	0.219
42	-0.014	0.259
43	0.034	0.242
44	0.013	0.252
45	-0.507	0.459
46	-0.007	0.257
47	0.103	0.219
48	-0.626	0.494
49	0.137	0.209
50	-0.487	0.445
51	-0.715	0.532
52	0.123	0.213
53	0.119	0.214
54	-1.923	0.726
55	-0.742	0.54
56	-0.932	0.587
57	-0.932	0.587
58	-0.69	0.532
59	-0.652	0.519
60	-0.561	0.485
61	-0.771	0.56
62	-0.878	0.592
63	-0.903	0.65
64	-0.657	0.526
65	-0.923	0.605
66	-0.693	0.519
67	-0.768	0.558
68	-0.677	0.528
69	-0.767	0.558
70	-0.779	0.545
71	-0.553	0.471
72	-1.107	0.657
73	-1.011	0.622
74	-0.753	0.549
75	-0.878	0.592
76	-0.793	0.562
77	-1.02	0.624
78	-1.139	0.657

79	-0.813	0.578
80	-1.02	0.624
81	-0.923	0.605
82	-0.623	0.512
83	-0.691	0.529
84	-0.712	0.544
85	0.255	0.176
86	0.973	0.164
87	0.672	0.138
88	1.144	0.195
89	0.616	0.137
90	0.628	0.137
91	0.526	0.137
92	-0.047	0.271
93	-0.232	0.344
94	1.577	0.304
95	-1.089	0.649
96	0.4	0.148
97	0.803	0.146
98	0.557	0.136
99	0.232	0.182
100	0.231	0.182
101	0.21	0.188
102	0.22	0.185
103	0.473	0.139
104	0.222	0.185
105	0.22	0.185
106	0.575	0.136
107	0.79	0.145
108	0.811	0.147
109	0.223	0.184
110	0.222	0.185
111	0.233	0.182
112	0.233	0.182
113	0.231	0.182
114	0.231	0.182
115	0.575	0.136
116	0.229	0.183
117	0.811	0.147
118	0.811	0.147
119	0.231	0.182
120	0.233	0.182
121	1.027	0.172
122	0.241	0.18
123	0.23	0.183
124	0.222	0.185
125	0.241	0.18
126	0.238	0.181
127	1.027	0.172
128	0.239	0.18
129	0.565	0.136
130	0.81	0.147
131	0.231	0.182
132	0.231	0.182
133	-0.12	0.301
134	0.018	0.248
135	-1.103	0.633
136	0.571	0.136
137	-0.611	0.489
138	0.525	0.137
139	0.11	0.217
140	0.089	0.223
141	0.126	0.212
142	0.102	0.219
143	-0.655	0.521
144	0.102	0.219
145	-0.008	0.255
146	0.601	0.136
147	-0.61	0.496
148	0.109	0.217
149	0.03	0.243
150	-0.652	0.503
151	0.105	0.219
152	-0.655	0.521
153	0.09	0.223
154	0.113	0.216
155	-0.756	0.545
156	-0.579	0.485
157	0.059	0.233
158	-0.549	0.474
159	0.857	0.151
160	-0.659	0.513
161	-0.884	0.582
162	0.059	0.233
163	-0.654	0.511
164	0.087	0.224

165	-1.319	0.674	208	0.389	0.15
166	-1.319	0.674	209	0.567	0.136
167	-1.728	0.728	210	0.388	0.149
168	0.137	0.209	211	-0.924	0.647
169	0.086	0.224	212	1.553	0.294
170	-0.689	0.523	213	0.818	0.147
171	0.108	0.217	214	0.341	0.158
172	0.123	0.213	215	-1.277	0.653
173	0.106	0.218	216	-1.344	0.664
174	0.108	0.217	217	-0.955	0.592
175	-1.254	0.663	218	0.196	0.192
176	-1.294	0.67	219	0.606	0.136
177	0.164	0.2	220	-1.428	0.701
178	0.273	0.172	221	-0.714	0.526
179	1.098	0.186	222	-1.019	0.624
180	-0.851	0.616	223	-0.686	0.517
181	0.155	0.202	224	0.458	0.141
182	0.652	0.137	225	-0.55	0.478
183	-1.107	0.622	226	-1.805	0.721
184	0.338	0.158	227	0.321	0.162
185	0.404	0.147	228	-1.179	0.659
186	0.686	0.139	229	-1.409	0.717
187	0.634	0.137	230	-0.904	0.579
188	1.194	0.207	231	-1.054	0.632
189	0.713	0.14	232	-1.095	0.648
190	2.692	0.569	233	-0.831	0.567
191	1.411	0.257	234	-1.199	0.643
192	1.746	0.341	235	-1.316	0.685
193	0.815	0.147	236	-0.391	0.423
194	1.382	0.252	237	0.613	0.136
195	0.367	0.153	238	0.764	0.144
196	0.679	0.139	239	-0.228	0.351
197	-0.366	0.4	240	0.514	0.137
198	0.39	0.149	241	-0.431	0.424
199	-0.915	0.644	242	-0.123	0.303
200	1.668	0.317	243	-0.133	0.305
201	0.666	0.139	244	-0.247	0.365
202	0.595	0.136	245	-0.62	0.492
203	-0.605	0.49	246	-0.356	0.4
204	1.015	0.17	247	-0.384	0.412
205	0.488	0.138	248	-0.625	0.534
206	0.385	0.15	249	1.396	0.256
207	0.648	0.137	250	1.346	0.244

251	-0.601	0.521	294	-1.092	0.67
252	0.735	0.142	295	-1.693	0.724
253	0.137	0.211	296	-1.404	0.729
254	-0.953	0.628	297	-1.494	0.745
255	0.616	0.137	298	-1.882	0.74
256	1.926	0.382	299	-1.726	0.727
257	0.45	0.141	300	-1.897	0.742
258	-0.11	0.297	301	-1.058	0.624
259	1.196	0.207	302	-1.421	0.69
260	0.63	0.137	303	-1.244	0.661
261	1.222	0.214	304	-1.317	0.674
262	1.321	0.239	305	-1.271	0.652
263	1.21	0.211	306	-1.141	0.629
264	1.281	0.229	307	-1.22	0.656
265	0.566	0.136	308	-0.877	0.569
266	1.207	0.21	309	0.64	0.137
267	0.72	0.141	310	0.839	0.149
268	-1.344	0.718	311	1.067	0.18
269	0.748	0.143	312	0.861	0.151
270	1.108	0.188	313	0.938	0.159
271	1.38	0.253	314	0.946	0.16
272	1.227	0.214	315	0.961	0.162
273	0.195	0.193	316	1.466	0.273
274	0.362	0.154	317	-0.46	0.435
275	-0.074	0.281	318	1.246	0.219
276	-1.244	0.661	319	1.212	0.211
277	-0.963	0.602	320	1.252	0.22
278	-1.286	0.668	321	1.373	0.248
279	-0.903	0.65	322	1.239	0.216
280	-1.351	0.719	323	1.749	0.337
281	-1.286	0.706	324	0.177	0.197
282	-0.903	0.65	325	-0.585	0.48
283	-1.434	0.734	326	-0.831	0.557
284	-1.162	0.645	327	-0.522	0.458
285	-1.402	0.729	328	-0.758	0.536
286	-1.816	0.718	329	-0.557	0.471
287	-1.223	0.644	330	-0.524	0.471
288	-1.686	0.723	331	-0.675	0.511
289	-1.344	0.718	332	-0.353	0.393
290	-0.903	0.65	333	-0.571	0.475
291	-1.476	0.698	334	-0.541	0.465
292	-1.316	0.712	335	-0.48	0.442
293	-1.498	0.701	336	-0.624	0.493

337	0.666	0.138
338	-0.584	0.48
339	-0.593	0.483
340	-0.566	0.473
341	0.749	0.143
342	0.826	0.148
343	1.08	0.182
344	1.838	0.366
345	2.244	0.481
346	1.754	0.344
347	0.921	0.157
348	0.779	0.145
349	2.077	0.438
350	1.407	0.256
351	0.696	0.14
352	0.671	0.139
353	0.397	0.148
354	0.727	0.141
355	0.653	0.138
356	0.033	0.242
357	-0.97	0.637
358	0.591	0.136
359	0.086	0.225
360	0.932	0.159
361	0.561	0.136
362	0.657	0.138
363	0.837	0.149
364	0.509	0.137
365	0.439	0.143
366	0.678	0.139
367	0.496	0.138
368	0.763	0.144
369	0.493	0.138
370	1.3	0.232
371	1.007	0.169
372	0.753	0.143
373	0.504	0.137
374	0.53	0.136
375	0.743	0.142
376	0.733	0.142
377	0.07	0.23
378	0.789	0.145
379	0.718	0.141

380	0.736	0.142
381	0.662	0.138
382	0.693	0.14
383	0.179	0.196
384	0.451	0.141
385	0.387	0.15
386	0.444	0.142
387	0.362	0.154
388	0.21	0.188
389	0.28	0.171
390	0.352	0.156
391	0.259	0.176
392	0.36	0.154
393	0.264	0.174
394	0.011	0.249
395	0.283	0.17
396	0.185	0.195
397	0.401	0.147
398	0.393	0.149
399	0.121	0.213
400	0.39	0.149
401	0.387	0.15
402	0.392	0.149
403	-1.047	0.61
404	-0.005	0.255
405	-1.215	0.646
406	-0.765	0.562
407	-1.133	0.627
408	-1.046	0.609
409	-1.191	0.638
410	-0.891	0.583
411	-1.299	0.661
412	-0.922	0.581
413	-0.957	0.589
414	-1.014	0.603
415	-0.982	0.595
416	-0.972	0.593
417	-0.945	0.587
418	-1.38	0.673
419	-0.781	0.543
420	-1.052	0.611
421	-0.625	0.497
422	-0.908	0.595

423	-1.26	0.654
424	-0.981	0.598
425	0.515	0.137
426	-0.821	0.57
427	0.374	0.152
428	-0.593	0.486
429	0.521	0.137
430	0.591	0.136
431	0.842	0.149
432	0.849	0.15
433	1.305	0.234
434	0.845	0.15
435	0.536	0.136
436	0.822	0.148
437	0.883	0.153
438	0.965	0.163
439	0.843	0.15
440	0.764	0.144
441	0.745	0.142
442	0.844	0.15
443	0.843	0.15
444	0.829	0.148
445	0.714	0.141
446	0.874	0.152
447	0.876	0.152
448	1.276	0.227
449	1.101	0.186
450	0.741	0.142
451	0.771	0.144
452	1.21	0.211
453	0.688	0.139
454	0.873	0.152
455	-0.651	0.505
456	0.851	0.15
457	0.775	0.144
458	0.859	0.151
459	0.844	0.15
460	1.292	0.23
461	1.292	0.23
462	0.796	0.146
463	1.3	0.232
464	0.838	0.149
465	0.852	0.15
466	1.165	0.2
467	0.442	0.142
468	0.851	0.15
469	0.858	0.151
470	1.339	0.242
471	1.046	0.176
472	0.846	0.15
473	1.303	0.233
474	1.329	0.24
475	0.772	0.144
476	0.793	0.146
477	0.448	0.142
478	0.858	0.151
479	0.673	0.138
480	1.291	0.229
481	0.838	0.149
482	0.911	0.156
483	1.239	0.216
484	0.873	0.152
485	0.567	0.135
486	0.14	0.207
487	0.121	0.212
488	0.558	0.136
489	0.329	0.16
490	0.142	0.207
491	0.145	0.206
492	-0.868	0.567
493	0.621	0.136
494	0.736	0.142
495	0.029	0.245
496	0.4	0.148
497	0.547	0.136
498	0.568	0.135
499	0.665	0.138
500	0.651	0.137
501	0.001	0.255
502	-0.035	0.266
503	0.639	0.137
504	0.605	0.136
505	0.801	0.146
506	0.788	0.145
507	0.565	0.136
508	0.111	0.216

509	0.141	0.207
510	0.491	0.138
511	-0.525	0.459
512	0.592	0.136
513	0.088	0.223
514	0.639	0.137
515	0.811	0.147
516	1.006	0.169
517	0.793	0.146
518	0.108	0.217
519	0.661	0.138
520	1.045	0.175
521	0.683	0.139
522	0.792	0.146
523	0.685	0.139
524	0.499	0.138
525	0.487	0.138
526	0.489	0.138
527	-0.72	0.539
528	-0.67	0.527
529	1.284	0.231
530	0.066	0.231
531	-0.378	0.405
532	0.976	0.165
533	-0.624	0.496
534	1.246	0.219
535	0.593	0.136
536	-0.745	0.541
537	-0.864	0.623
538	-0.731	0.531
539	-0.903	0.65
540	0.365	0.153
541	-1.044	0.685
542	1.29	0.23
543	0.865	0.152
544	0.16	0.202
545	1.257	0.222
546	0.743	0.142
547	0.621	0.137
548	-0.78	0.546
549	-0.903	0.65
550	-0.895	0.588
551	-1.223	0.644
552	-0.715	0.527
553	-0.951	0.588
554	-0.34	0.392
555	-1.342	0.678
556	-0.226	0.35
557	-0.66	0.529
558	-0.67	0.533
559	-0.623	0.502
560	-1.098	0.679
561	-0.369	0.404
562	-0.134	0.306
563	-0.621	0.5
564	-0.8	0.558
565	-0.759	0.546
566	-0.5	0.456
567	-0.571	0.482
568	-0.532	0.468
569	0.518	0.137
570	1.773	0.342
571	0.63	0.136
572	0.217	0.186
573	-0.238	0.35
574	0.195	0.191
575	0.598	0.136
576	1.508	0.283
577	-0.235	0.345
578	0.144	0.206
579	-0.631	0.499
580	-0.432	0.424
581	-0.747	0.536
582	0.175	0.197
583	0.12	0.215
584	-0.492	0.447
585	-0.731	0.528
586	-0.601	0.486
587	-0.599	0.485
588	0.263	0.174
589	-0.467	0.439
590	-0.489	0.448
591	-0.976	0.597
592	-0.485	0.444
593	0.459	0.14
594	-0.638	0.51

595	-0.641	0.502
596	0.304	0.165
597	0.674	0.138
598	-0.728	0.53
599	-0.669	0.511
600	0.384	0.15
601	0.449	0.142
602	-2.107	0.742
603	-1.012	0.634
604	0.746	0.143
605	1.352	0.246
606	-1.012	0.634
607	-0.753	0.558
608	-0.879	0.602
609	0.637	0.137
610	0.085	0.224
611	-0.96	0.59
612	-0.813	0.552
613	-0.697	0.518
614	-0.974	0.604
615	-0.567	0.476
616	-1.176	0.648
617	-0.959	0.59
618	-1.88	0.727
619	-1.321	0.674
620	-0.43	0.434
621	-0.412	0.416
622	0.535	0.136
623	-0.381	0.403
624	0.132	0.21
625	-0.704	0.52
626	-0.573	0.491
627	-0.398	0.413
628	-0.506	0.454
629	-0.87	0.578
630	-1.076	0.655
631	-0.118	0.305
632	-0.03	0.265
633	-0.309	0.383
634	-0.167	0.325
635	-0.242	0.358
636	0.317	0.163
637	-0.453	0.432
638	-0.214	0.343
639	0.46	0.14
640	-0.39	0.432
641	1.008	0.169
642	0.176	0.197
643	-0.629	0.498
644	-0.329	0.385
645	-1.559	0.693
646	0.768	0.144
647	-0.834	0.574
648	-0.238	0.348
649	-1.134	0.65
650	0.73	0.142
651	0.46	0.14
652	-0.951	0.591
653	-0.951	0.591
654	1.004	0.169
655	0.496	0.138
656	1.227	0.214
657	0.957	0.162
658	-0.749	0.557
659	0.697	0.14
660	0.138	0.208
661	0.852	0.15
662	0.691	0.14
663	0.506	0.138
664	0.394	0.149
665	0.134	0.21
666	0.95	0.161
667	0.759	0.143
668	1.146	0.196
669	0.767	0.144
670	0.971	0.164
671	0.08	0.229
672	-0.445	0.456
673	1.153	0.198
674	1.127	0.192
675	0.639	0.137
676	0.785	0.145
677	1.306	0.234
678	0.865	0.152
679	1.289	0.231
680	0.537	0.136

681	0.178	0.198	724	-0.404	0.415
682	1.629	0.313	725	0.125	0.212
683	1.114	0.189	726	-0.383	0.412
684	0.722	0.141	727	0.145	0.206
685	1.305	0.232	728	0.125	0.212
686	0.717	0.141	729	-0.402	0.417
687	-0.514	0.455	730	-0.353	0.396
688	-0.903	0.65	731	0.867	0.152
689	-0.514	0.455	732	-0.589	0.505
690	0.104	0.218	733	0.185	0.195
691	-0.483	0.444	734	0.182	0.196
692	0.595	0.136	735	-0.965	0.656
693	0.433	0.143	736	-0.523	0.48
694	-0.845	0.561	737	-0.589	0.507
695	0.514	0.137	738	-0.681	0.557
696	-0.565	0.48	739	-0.618	0.519
697	-0.822	0.555	740	1.307	0.237
698	0.522	0.136	741	1.112	0.19
699	0.113	0.215	742	1.112	0.19
700	-0.53	0.461	743	-0.392	0.413
701	-0.53	0.461	744	-0.426	0.427
702	-0.542	0.465	745	0.553	0.136
703	-0.734	0.529	746	0.66	0.138
704	1.276	0.227	747	0.652	0.138
705	0.97	0.164	748	1.058	0.177
706	0.7	0.14	749	0.088	0.223
707	1.019	0.171	750	0.57	0.135
708	0.979	0.165	751	0.667	0.138
709	1.113	0.189	752	0.978	0.164
710	1.082	0.182	753	0.552	0.136
711	-0.483	0.457	754	1.108	0.187
712	0.843	0.149	755	0.583	0.136
713	0.031	0.242	756	1.09	0.183
714	0.005	0.251	757	0.934	0.158
715	0.161	0.201	758	0.413	0.146
716	0.07	0.229	759	-0.363	0.41
717	-1.385	0.716	760	-0.738	0.53
718	-0.597	0.484	761	0.379	0.151
719	0.147	0.205	762	0.421	0.145
720	0.088	0.224	763	-0.581	0.489
721	-1.079	0.616	764	-0.652	0.515
722	-0.378	0.41	765	0.055	0.236
723	0.192	0.193	766	-0.421	0.427

767	-0.939	0.631	810	1.371	0.25
768	-0.737	0.563	811	-1.506	0.731
769	-0.674	0.522	812	0.234	0.182
770	-0.761	0.583	813	-2.398	0.753
771	-0.581	0.489	814	-1.642	0.719
772	-0.995	0.647	815	-1.047	0.643
773	-0.713	0.535	816	-1.085	0.652
774	-0.713	0.535	817	-0.785	0.598
775	-0.883	0.613	818	0.105	0.218
776	-0.757	0.57	819	-0.903	0.65
777	0.815	0.147	820	-0.512	0.457
778	0.957	0.161	821	-0.565	0.48
779	0.835	0.149	822	-0.44	0.43
780	0.945	0.16	823	-0.467	0.44
781	0.981	0.165	824	-0.454	0.438
782	-1.342	0.678	825	-0.446	0.432
783	-1.34	0.664	826	-0.473	0.443
784	-0.958	0.616	827	0.812	0.147
785	-1.386	0.709	828	-0.37	0.404
786	0.099	0.222	829	-0.473	0.443
787	-0.331	0.394	830	1.699	0.332
788	-1.591	0.697	831	1.681	0.327
789	-1.262	0.664	832	1.327	0.24
790	-1.44	0.679	833	1.344	0.244
791	-0.966	0.595	834	1.336	0.242
792	-1.085	0.621	835	1.327	0.24
793	-1.357	0.68	836	1.336	0.242
794	1.349	0.244	837	1.344	0.244
795	1.365	0.247	838	1.346	0.244
796	0.748	0.143	839	1.346	0.244
797	1.078	0.182	840	1.346	0.244
798	1.018	0.17	841	1.327	0.24
799	0.937	0.16	842	1.318	0.237
800	0.667	0.138	843	1.346	0.244
801	0.342	0.158	844	1.334	0.242
802	1.109	0.188	845	1.344	0.244
803	-1.384	0.713	846	1.343	0.244
804	0.84	0.149	847	1.314	0.237
805	1.138	0.194	848	1.314	0.237
806	0.682	0.139	849	1.336	0.242
807	1.464	0.272	850	1.346	0.244
808	0.092	0.221	851	1.083	0.183
809	0.809	0.147	852	0.48	0.139

853	0.951	0.161	896	2.156	0.471
854	0.282	0.17	897	2.156	0.471
855	0.826	0.148	898	2.156	0.471
856	0.277	0.171	899	2.156	0.471
857	0.807	0.147	900	2.156	0.471
858	0.482	0.139	901	2.156	0.471
859	0.455	0.141	902	2.156	0.471
860	0.98	0.166	903	2.156	0.471
861	2.156	0.471	904	1.259	0.225
862	2.156	0.471	905	1.552	0.299
863	1.552	0.299	906	2.156	0.471
864	2.156	0.471	907	2.156	0.471
865	2.156	0.471	908	2.156	0.471
866	2.156	0.471	909	2.156	0.471
867	2.156	0.471	910	2.156	0.471
868	2.156	0.471	911	2.156	0.471
869	2.156	0.471	912	1.552	0.299
870	2.156	0.471	913	2.156	0.471
871	1.552	0.299	914	1.259	0.225
872	2.156	0.471	915	1.259	0.225
873	2.156	0.471	916	2.156	0.471
874	2.156	0.471	917	1.552	0.299
875	2.156	0.471	918	1.08	0.183
876	2.156	0.471	919	2.156	0.471
877	2.156	0.471	920	2.156	0.471
878	2.156	0.471	921	-0.012	0.258
879	1.552	0.299	922	-1.116	0.651
880	2.156	0.471	923	-1.057	0.637
881	2.156	0.471	924	-1.005	0.633
882	2.156	0.471	925	0.057	0.233
883	1.315	0.239	926	0.56	0.136
884	1.552	0.299	927	0.895	0.154
885	1.552	0.299	928	-0.21	0.337
886	2.156	0.471	929	0.856	0.15
887	1.552	0.299	930	0.284	0.17
888	2.156	0.471	931	0.301	0.166
889	2.156	0.471	932	-0.016	0.259
890	2.156	0.471	933	0.306	0.165
891	1.552	0.299	934	0.538	0.136
892	1.963	0.412	935	0.593	0.136
893	1.552	0.299	936	0.059	0.233
894	2.156	0.471	937	0.874	0.152
895	2.156	0.471	938	0.758	0.143

939	0.291	0.168
940	1.071	0.18
941	0.783	0.145
942	0.678	0.139
943	0.61	0.136
944	0.777	0.145
945	0.491	0.138
946	0.302	0.166
947	0.449	0.141
948	0.285	0.17
949	0.29	0.168
950	0.317	0.163
951	0.291	0.168
952	0.291	0.168
953	-0.254	0.355
954	-0.062	0.277
955	-0.25	0.353
956	-0.322	0.383
957	0.311	0.164
958	0.163	0.201
959	0.241	0.18
960	0.573	0.136
961	0.843	0.15
962	0.314	0.163
963	0.293	0.168
964	0.313	0.163
965	0.315	0.163
966	0.459	0.141
967	0.497	0.138
968	0.492	0.138
969	0.49	0.138
970	0.822	0.148
971	0.749	0.143
972	1.187	0.206
973	1.184	0.205
974	1.238	0.217
975	1.163	0.2
976	1.168	0.201
977	1.104	0.187
978	1.156	0.198
979	1.169	0.202
980	1.126	0.192
981	1.154	0.198
982	1.149	0.197
983	1.158	0.2
984	1.169	0.202
985	1.141	0.195
986	1.158	0.199
987	1.121	0.192
988	1.132	0.193
989	1.169	0.202
990	1.164	0.201
991	0.45	0.142
992	0.601	0.136
993	0.746	0.142
994	0.222	0.185
995	0.747	0.142
996	0.636	0.137
997	0.692	0.14
998	0.583	0.136
999	0.224	0.184
1000	-0.675	0.529
1001	0.464	0.14
1002	0.364	0.154
1003	-0.627	0.512
1004	0.691	0.139
1005	0.739	0.142
1006	0.732	0.142
1007	0.237	0.181
1008	0.588	0.136
1009	0.061	0.233
1010	0.498	0.138
1011	-1.259	0.701
1012	-0.812	0.557
1013	-0.725	0.53
1014	-1.205	0.647
1015	-1.377	0.676
1016	-1.205	0.647
1017	-0.756	0.54
1018	-0.725	0.53
1019	-0.559	0.478
1020	-1.205	0.647
1021	-1.296	0.67
1022	-0.797	0.557
1023	-0.637	0.498
1024	0.751	0.143

1025	1.231	0.214	1068	-0.999	0.602
1026	1.046	0.176	1069	-1.332	0.666
1027	1.293	0.231	1070	-1.112	0.627
1028	0.921	0.157	1071	-1.171	0.657
1029	1.135	0.193	1072	-1.069	0.635
1030	1.88	0.378	1073	-1.298	0.678
1031	1.418	0.261	1074	-0.952	0.591
1032	1.449	0.269	1075	-0.853	0.566
1033	1.21	0.211	1076	-0.79	0.548
1034	1.381	0.253	1077	0.909	0.156
1035	1.457	0.271	1078	0.861	0.151
1036	1.099	0.186	1079	0.642	0.137
1037	1.469	0.275	1080	0.804	0.146
1038	1.274	0.225	1081	1.059	0.178
1039	1.457	0.271	1082	-0.846	0.564
1040	1.448	0.266	1083	1.032	0.173
1041	-1.123	0.652	1084	-0.51	0.456
1042	0.152	0.204	1085	0.283	0.17
1043	0.151	0.204	1086	-0.325	0.383
1044	-1.027	0.609	1087	0.242	0.179
1045	-0.961	0.594	1088	1.029	0.173
1046	-1.121	0.628	1089	0.81	0.147
1047	-1.014	0.627	1090	0.918	0.157
1048	-1.469	0.728	1091	-0.522	0.46
1049	-1.221	0.682	1092	0.806	0.147
1050	-1.812	0.754	1093	0.394	0.149
1051	-0.886	0.572	1094	0.735	0.142
1052	-1.145	0.633	1095	0.34	0.158
1053	-0.956	0.608	1096	0.391	0.149
1054	-1.5	0.712	1097	0.577	0.136
1055	-1.19	0.686	1098	0.456	0.141
1056	-1.357	0.693	1099	-0.473	0.44
1057	-1.101	0.642	1100	-0.47	0.439
1058	-1.101	0.642	1101	-0.515	0.456
1059	-1.189	0.657	1102	-0.493	0.447
1060	-1.203	0.674	1103	-0.462	0.436
1061	-0.766	0.542	1104	-0.462	0.436
1062	-1.055	0.632	1105	-0.517	0.456
1063	-1.407	0.701	1106	-0.441	0.428
1064	-1.434	0.706	1107	0.58	0.136
1065	-0.333	0.393	1108	-0.696	0.53
1066	-1.121	0.643	1109	-0.367	0.398
1067	-0.82	0.557	1110	0.121	0.213

1111	-0.903	0.65	1154	-0.472	0.445
1112	-0.576	0.487	1155	-0.006	0.256
1113	-0.506	0.452	1156	0.066	0.231
1114	-0.462	0.436	1157	0.136	0.209
1115	-0.441	0.428	1158	-0.759	0.546
1116	0.124	0.212	1159	-1.342	0.678
1117	-0.325	0.382	1160	0.563	0.136
1118	-0.474	0.441	1161	-0.515	0.462
1119	-0.485	0.445	1162	-0.451	0.437
1120	-0.567	0.474	1163	0.032	0.243
1121	-0.696	0.53	1164	0.891	0.154
1122	0.121	0.213	1165	0.413	0.146
1123	-0.43	0.429	1166	0.914	0.157
1124	-0.415	0.418	1167	-0.469	0.439
1125	-0.518	0.467	1168	0.442	0.142
1126	-0.435	0.428	1169	0.47	0.14
1127	-0.405	0.414	1170	-0.506	0.454
1128	-0.585	0.492	1171	-0.518	0.457
1129	-0.615	0.498	1172	0.865	0.151
1130	0.101	0.22	1173	-0.631	0.499
1131	-0.021	0.262	1174	0.928	0.158
1132	0.14	0.208	1175	0.843	0.15
1133	0.083	0.225	1176	0.962	0.162
1134	0.074	0.228	1177	0.905	0.156
1135	0.36	0.154	1178	0.951	0.161
1136	-0.02	0.262	1179	0.938	0.159
1137	0.487	0.139	1180	0.943	0.16
1138	-0.887	0.582	1181	0.854	0.15
1139	0.491	0.138	1182	0.858	0.151
1140	-0.531	0.467	1183	-0.598	0.485
1141	-0.893	0.574	1184	0.883	0.153
1142	0.131	0.21	1185	0.381	0.151
1143	0.518	0.137	1186	0.363	0.154
1144	0.484	0.139	1187	-0.779	0.546
1145	0.834	0.149	1188	-0.06	0.275
1146	0.495	0.138	1189	0.952	0.161
1147	-0.827	0.556	1190	0.749	0.143
1148	-0.61	0.496	1191	0.096	0.221
1149	-0.002	0.255	1192	-0.605	0.49
1150	-0.902	0.586	1193	0.1	0.22
1151	-1.311	0.673	1194	0.462	0.14
1152	-0.672	0.51	1195	0.679	0.139
1153	-0.519	0.463	1196	0.033	0.242

1197	-1.691	0.711
1198	-0.47	0.442
1199	-0.329	0.383
1200	-0.523	0.468
1201	-0.856	0.614
1202	-0.856	0.614
1203	-0.335	0.392
1204	-0.702	0.555
1205	-0.415	0.434
1206	-0.468	0.44
1207	-0.375	0.419
1208	-0.375	0.419
1209	-0.246	0.35
1210	-0.931	0.604
1211	-0.375	0.419
1212	-0.357	0.409
1213	-0.166	0.318
1214	-0.357	0.409
1215	-0.375	0.419
1216	-0.584	0.483
1217	-0.375	0.419
1218	-0.246	0.35
1219	-0.49	0.452
1220	-0.607	0.51
1221	-0.196	0.336
1222	-0.875	0.633
1223	-0.382	0.415
1224	-0.256	0.359
1225	-0.442	0.428
1226	-0.461	0.435
1227	-0.659	0.526
1228	-0.434	0.433
1229	-0.484	0.444
1230	-0.415	0.417
1231	-0.604	0.489
1232	-0.474	0.44
1233	-0.456	0.44
1234	-0.747	0.557
1235	-0.459	0.441
1236	-0.548	0.467
1237	-0.947	0.609
1238	-0.74	0.55
1239	-0.765	0.555
1240	-0.496	0.449
1241	-0.63	0.498
1242	0.442	0.143
1243	-0.775	0.559
1244	-0.475	0.44
1245	0.427	0.144
1246	-0.375	0.419
1247	-0.357	0.409
1248	-0.357	0.409
1249	-0.584	0.483
1250	-0.166	0.318
1251	-0.357	0.409
1252	-0.375	0.419
1253	-0.194	0.329
1254	-0.375	0.419
1255	-0.77	0.559
1256	-0.607	0.51
1257	-0.848	0.624
1258	-0.49	0.452
1259	-0.277	0.362
1260	-0.256	0.359
1261	0.58	0.136
1262	0.893	0.155
1263	-0.607	0.51
1264	-0.23	0.35
1265	-0.325	0.382
1266	-0.445	0.439
1267	-0.303	0.377
1268	-0.472	0.445
1269	-0.507	0.466
1270	-0.35	0.392
1271	0.641	0.137
1272	-0.407	0.42
1273	-0.551	0.482
1274	-0.322	0.38
1275	-0.322	0.38
1276	-0.375	0.402
1277	-0.356	0.394
1278	-0.563	0.49
1279	1.368	0.251
1280	-0.134	0.304
1281	-0.459	0.445
1282	-0.298	0.371

1283	-0.356	0.394
1284	-0.375	0.419
1285	-0.357	0.409
1286	-0.584	0.483
1287	-0.344	0.405
1288	-0.607	0.51
1289	-0.49	0.452
1290	-0.77	0.559
1291	-0.461	0.438
1292	-0.835	0.619
1293	-0.166	0.318
1294	-0.375	0.419
1295	-0.375	0.419
1296	-0.196	0.336
1297	-0.271	0.36
1298	-0.223	0.34
1299	-0.246	0.35
1300	-0.415	0.434
1301	-0.357	0.409
1302	-0.375	0.419
1303	-0.77	0.559
1304	-0.375	0.419
1305	-0.87	0.585
1306	-0.246	0.35
1307	-0.297	0.388
1308	-0.921	0.604
1309	0.509	0.137
1310	-0.375	0.419
1311	0.499	0.138
1312	-0.77	0.559
1313	-0.246	0.35
1314	-0.31	0.384
1315	-0.246	0.35
1316	-0.246	0.35
1317	-0.375	0.419
1318	-0.77	0.559
1319	-1.045	0.636
1320	-0.295	0.37
1321	-0.375	0.419
1322	-0.375	0.419
1323	-0.986	0.599
1324	-0.994	0.658
1325	-1.48	0.699
1326	-0.814	0.565
1327	-0.458	0.452
1328	-0.649	0.521
1329	-0.555	0.49
1330	-1.451	0.689
1331	-0.953	0.592
1332	-0.738	0.548
1333	-0.903	0.65
1334	-1.628	0.701
1335	-1.515	0.722
1336	-1.316	0.66
1337	-0.742	0.531
1338	-0.535	0.463
1339	-0.994	0.658
1340	-0.87	0.571
1341	-1.32	0.692
1342	-1.12	0.628
1343	-0.745	0.547
1344	-0.424	0.421
1345	-0.997	0.602
1346	-0.226	0.341
1347	-0.575	0.513
1348	-0.767	0.548
1349	-0.557	0.48
1350	-0.273	0.364
1351	-0.708	0.521
1352	-0.574	0.49
1353	-0.273	0.364
1354	-0.491	0.452
1355	-0.472	0.446
1356	-0.457	0.462
1357	-0.196	0.336
1358	-0.238	0.351
1359	-0.336	0.399
1360	-0.245	0.349
1361	-0.363	0.406
1362	-0.788	0.601
1363	-0.351	0.408
1364	-0.706	0.521
1365	-0.607	0.51
1366	-0.452	0.437
1367	-0.77	0.559
1368	-0.848	0.624

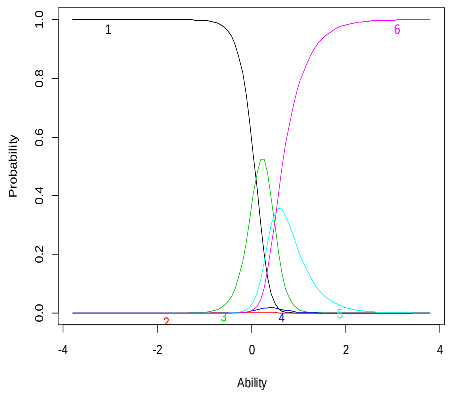
1369	-0.468	0.44	1412	-0.64	0.522
1370	-0.607	0.51	1413	-0.935	0.61
1371	-0.29	0.379	1414	-1.236	0.681
1372	-1.049	0.67	1415	-1.232	0.666
1373	-0.239	0.356	1416	-0.569	0.517
1374	-0.825	0.556	1417	-0.903	0.65
1375	-0.227	0.345	1418	-1.536	0.691
1376	-0.825	0.556	1419	-1.258	0.663
1377	-0.468	0.438	1420	-0.932	0.601
1378	-0.663	0.515	1421	-0.903	0.65
1379	-0.595	0.491	1422	-0.883	0.629
1380	-0.623	0.493	1423	-0.937	0.611
1381	-0.897	0.575	1424	-0.904	0.605
1382	-0.623	0.493	1425	-1.166	0.666
1383	-0.751	0.558	1426	-1.088	0.697
1384	-0.732	0.574	1427	-1.045	0.621
1385	-0.903	0.65	1428	-0.401	0.44
1386	-0.454	0.465	1429	-0.88	0.619
1387	-0.603	0.51	1430	-1.508	0.717
1388	-0.24	0.361	1431	-0.282	0.379
1389	0.117	0.217	1432	-0.418	0.435
1390	-0.188	0.328	1433	-1.277	0.689
1391	-0.405	0.413	1434	-1.021	0.646
1392	-0.252	0.352	1435	-0.431	0.45
1393	-0.264	0.357	1436	-0.454	0.465
1394	-0.264	0.357	1437	-0.454	0.465
1395	-0.252	0.352	1438	-1.923	0.726
1396	-0.698	0.536	1439	-1.459	0.722
1397	-0.264	0.357	1440	-1.11	0.654
1398	-0.236	0.351	1441	-0.945	0.654
1399	-0.319	0.384	1442	-0.903	0.65
1400	-0.319	0.384	1443	-0.903	0.65
1401	0.961	0.162	1444	-0.597	0.519
1402	-0.252	0.352	1445	-1.226	0.683
1403	-0.218	0.341	1446	-1.414	0.705
1404	-0.218	0.341	1447	-1.027	0.675
1405	-1.154	0.635	1448	-1.11	0.654
1406	-0.688	0.518	1449	-0.903	0.65
1407	-0.605	0.495	1450	-0.277	0.382
1408	-1.002	0.611	1451	-1.349	0.691
1409	-0.434	0.432	1452	-0.277	0.382
1410	-0.721	0.575	1453	-0.945	0.654
1411	-0.471	0.445	1454	-1.11	0.654

1455	-1.223	0.714
1456	-1.414	0.705
1457	-1.166	0.666
1458	-1.166	0.666
1459	-1.166	0.666
1460	-1.262	0.686
1461	-1.209	0.675
1462	-1.393	0.672
1463	-0.938	0.615
1464	-1.286	0.68
1465	-1.316	0.697
1466	-1.229	0.669
1467	-1.386	0.671
1468	-1.597	0.702
1469	-2.053	0.738
1470	-0.342	0.402
1471	-0.967	0.611
1472	-0.408	0.44
1473	-1.317	0.66
1474	-0.782	0.568
1475	-1.059	0.624
1476	-0.718	0.543
1477	-1.331	0.676
1478	-0.987	0.628
1479	0.08	0.23
1480	-0.561	0.486
1481	-0.638	0.516
1482	-1.196	0.664
1483	-0.969	0.612
1484	-0.971	0.612
1485	0.15	0.207
1486	0.19	0.195
1487	-0.979	0.622
1488	-1.029	0.639
1489	-1.688	0.717
1490	-1.237	0.676
1491	-0.903	0.65
1492	-0.642	0.548
1493	-0.429	0.449
1494	-0.277	0.382
1495	-1.572	0.722
1496	-0.868	0.603
1497	-1.351	0.719
1498	-1.386	0.68
1499	-0.783	0.547
1500	-2.004	0.735
1501	-1.48	0.731
1502	-1.32	0.704
1503	-1.19	0.71
1504	-1.19	0.71
1505	-1.324	0.665
1506	-1.17	0.641
1507	-0.956	0.589
1508	0.19	0.195
1509	-1.332	0.695
1510	-0.903	0.65
1511	0.19	0.195
1512	0.129	0.214
1513	-0.454	0.465
1514	0.15	0.207
1515	0.15	0.207
1516	0.15	0.207
1517	0.19	0.195
1518	0.15	0.207
1519	0.953	0.162
1520	0.08	0.23
1521	-0.454	0.465
1522	-1.332	0.695
1523	-1.088	0.697
1524	-0.277	0.382
1525	-1.234	0.716
1526	-0.401	0.44
1527	-1.332	0.695
1528	-0.903	0.65
1529	0.078	0.229
1530	-1.147	0.668
1531	-1.088	0.697
1532	0.08	0.23
1533	0.08	0.23
1534	-1.635	0.743
1535	-1.514	0.739
1536	-1.719	0.754
1537	-1.388	0.705
1538	-0.632	0.521
1539	0.08	0.23
1540	-1.719	0.754

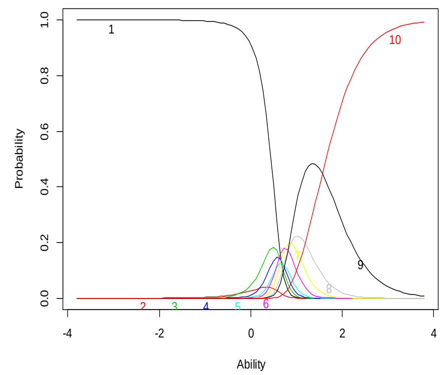
1541	-0.632	0.521	1544	-1.088	0.697
1542	-0.723	0.557	1545	0.081	0.226
1543	-1.388	0.705	1546	0.08	0.23

APPENDIX VI

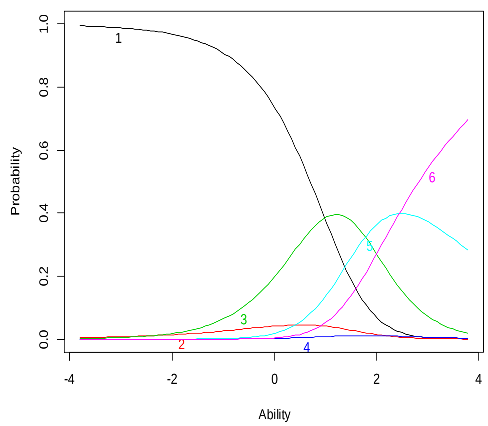
CATEGORY RESPONSE CURVES OF PHYSICAL GEOGRAPHY ACHIEVEMENT TEST



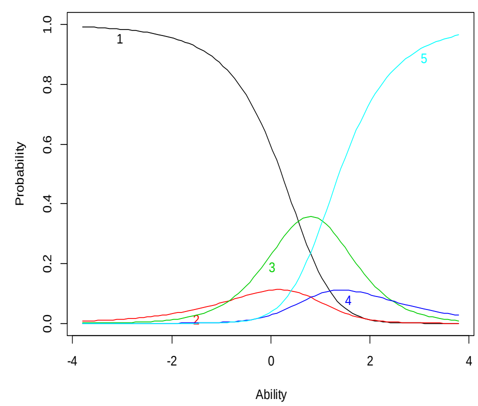
Category response curves of item 1A



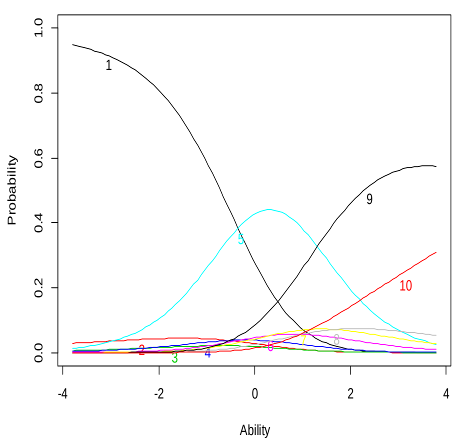
Category response curves of item 1B



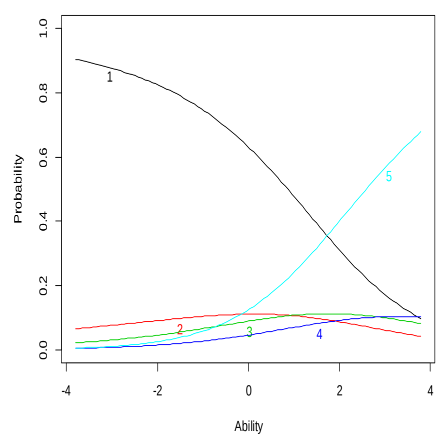
Category response curves of item 1C



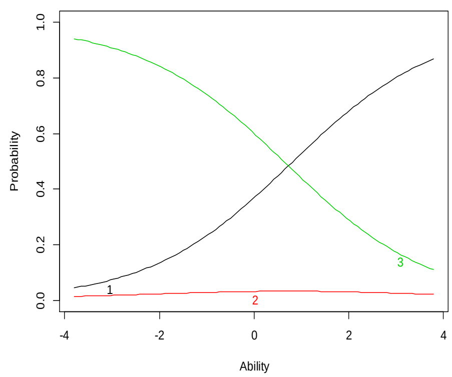
Category response curves of item 1D



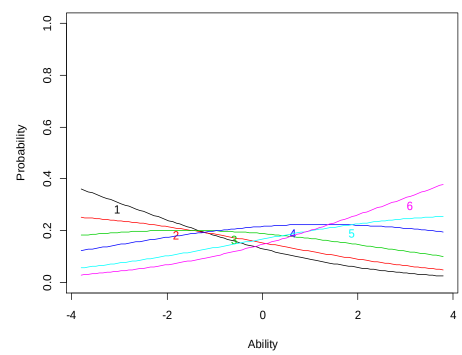
Category response curves of item 2A



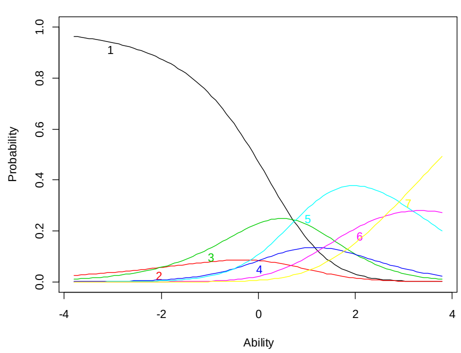
Category response curves of item 2B



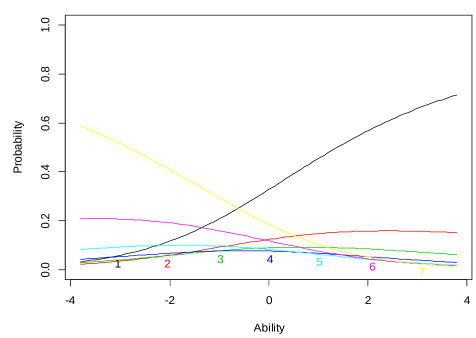
Category response curves of item 2C



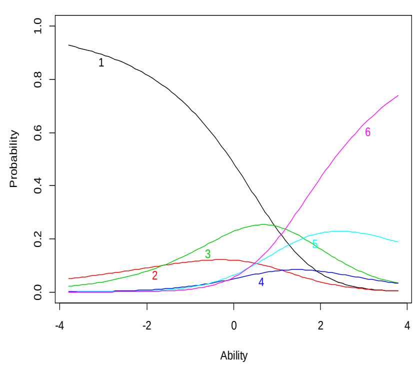
Category response curves of item 3A



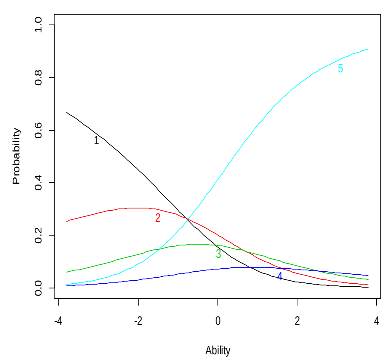
Category response curves of item 3B



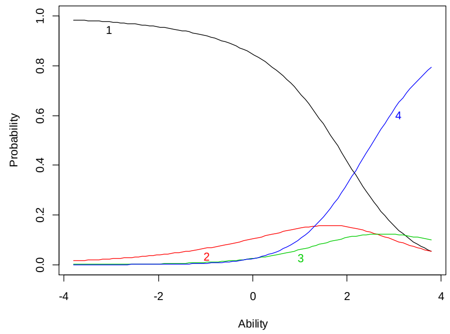
Category response curves of item 4A

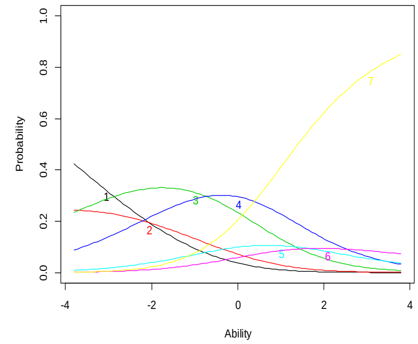


Category response curves of item 4B



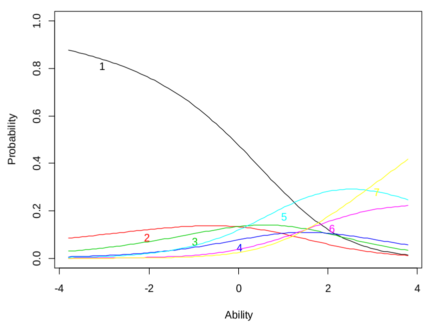
Category response curves of item 4C

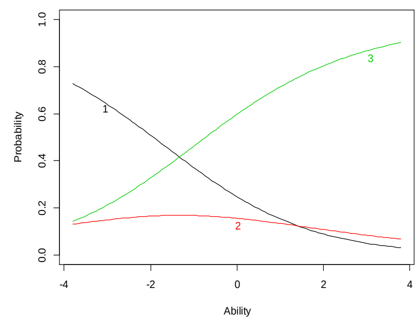




Category response curves of item 5AI

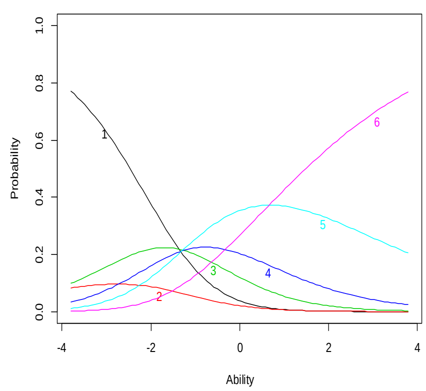
Category response curves of item 5AII

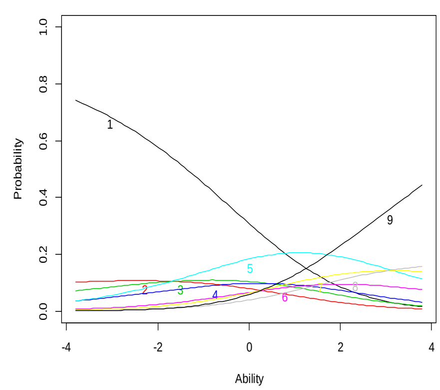




Category response curves of item 5B

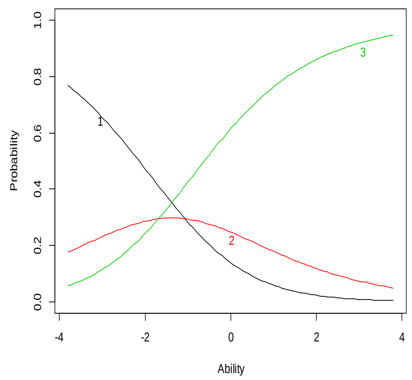
Category response curves of item 6A



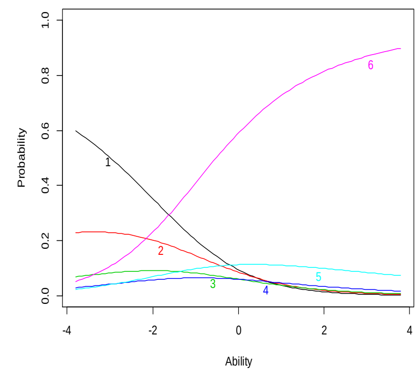


Category response curves of item 6B

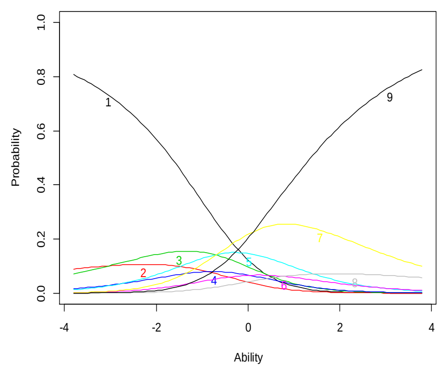
Category response curves of item 6C



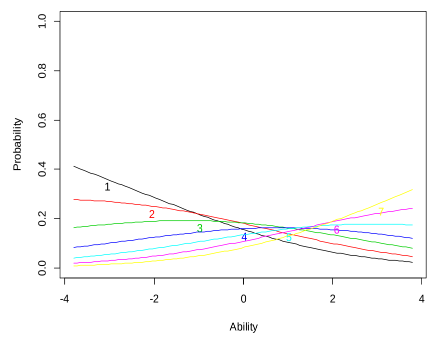
Category response curves of item 7AI



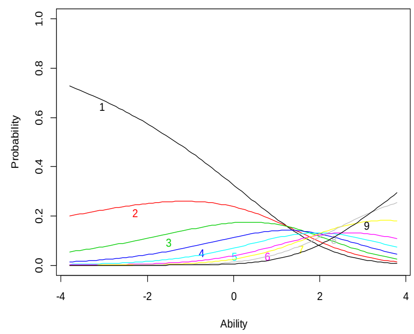
Category response curves of item 7AII



Category response curves of item 7B



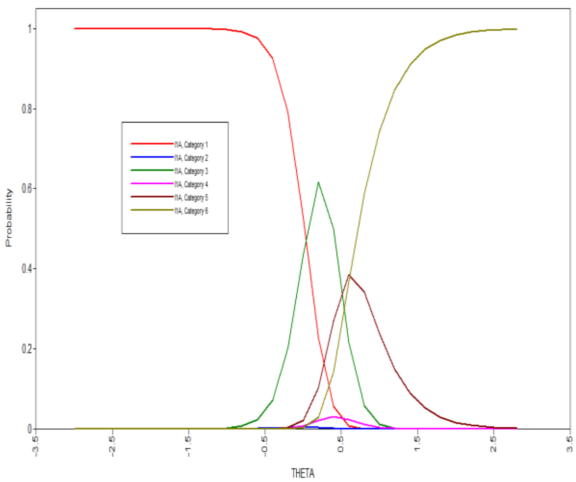
Category response curves of item 8A

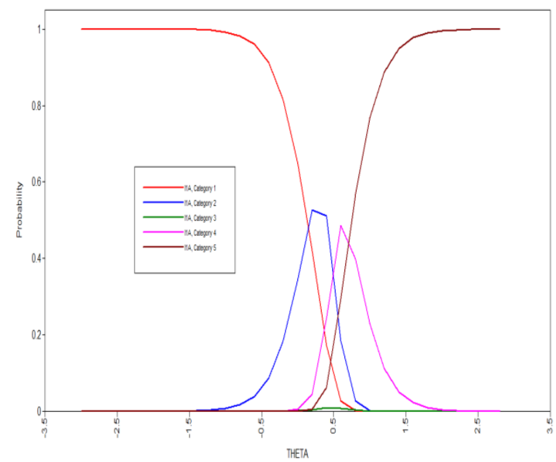


Category response curves of item 8B

APPENDIX VII: MALE AND FEMALE ITEM CHARACTERISTIC CURVES

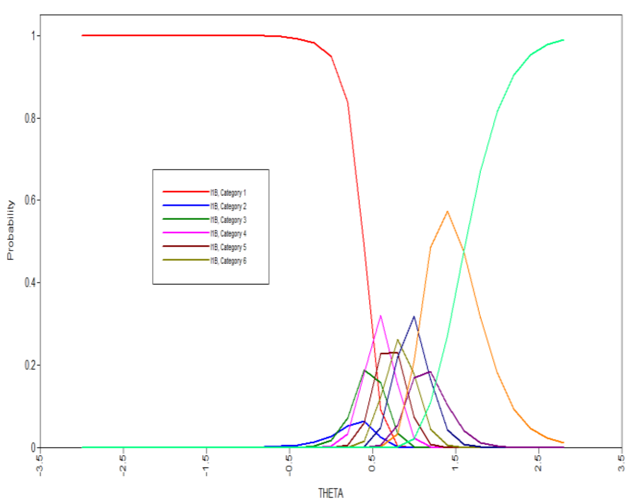
The item characteristic curve of the items for male and female students is presented in Figures i1a male and female – i8b male and female i8b respectively.

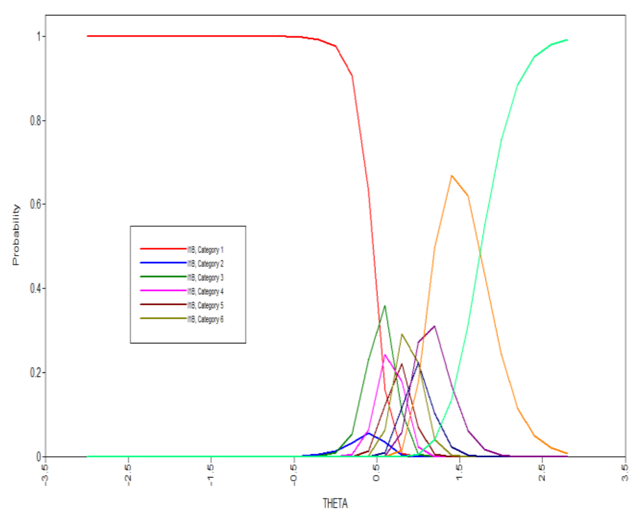




i1a Male   i1a Female

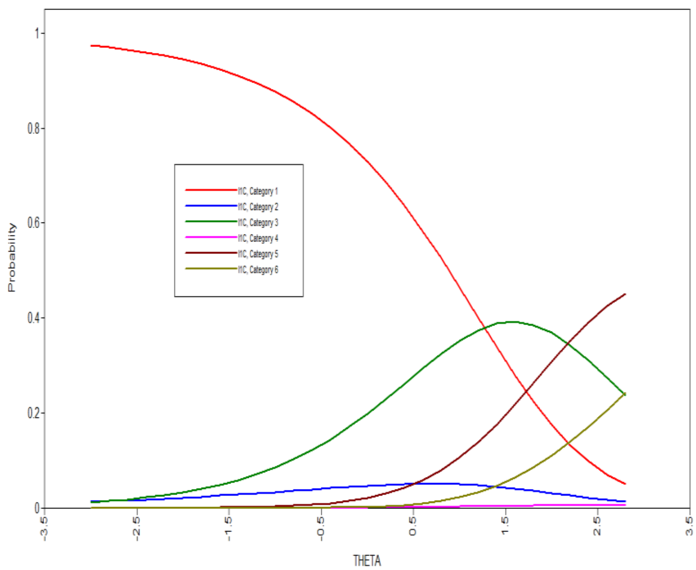
The diagram above shows the Item characteristic curve (ICC) for item 1a for male and female students respectively.To answer research question 4, the ICC above were observed to make comparison.it was seen that the two ICC difers fom one another so i1a is not invariant.

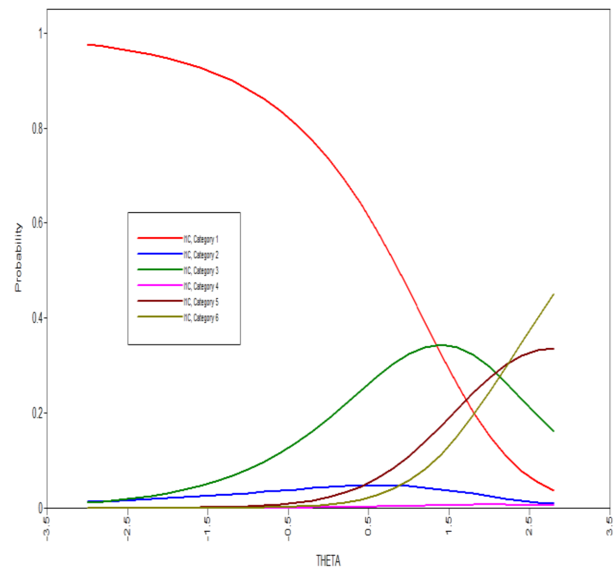




Male i1b   Female i1b

The diagram above shows the Item characteristic curve (ICC) for item 1b for male and female students respectively.To answer research question 4, the ICC above were observed to make comparison.it was seen that the two ICC are similar to one another so i1b is invariant.





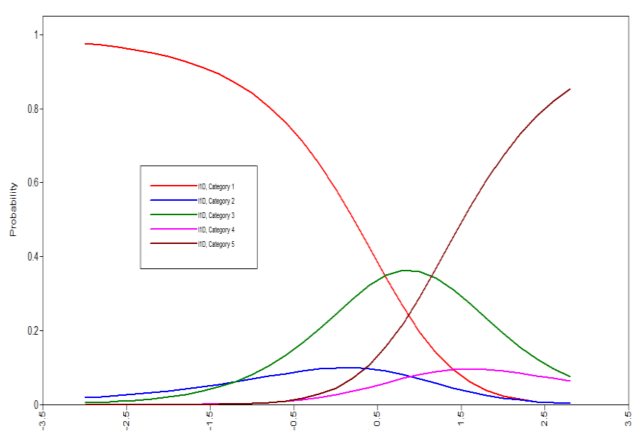
Male i1c

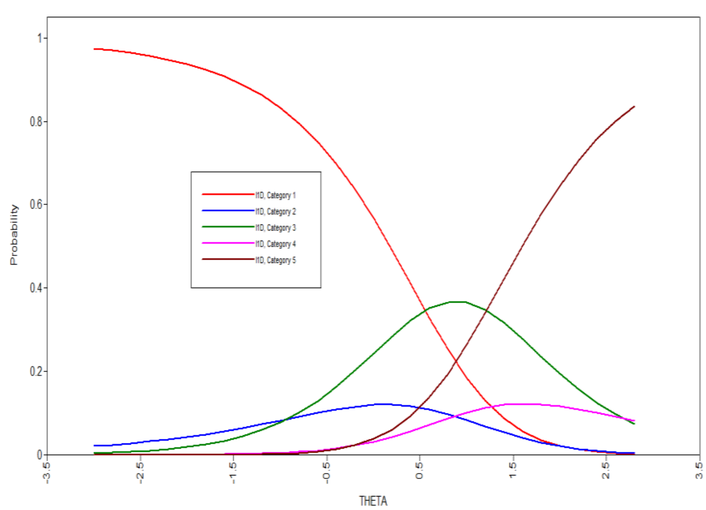
Female i1c

The diagram above shows the Item characteristic curve (ICC) for item 1c for male and female students respectively.To answer research question 4, the ICC above were observed to make comparison.it was seen that the two ICC difers fom one another so i1a is not invariant

Male i1d

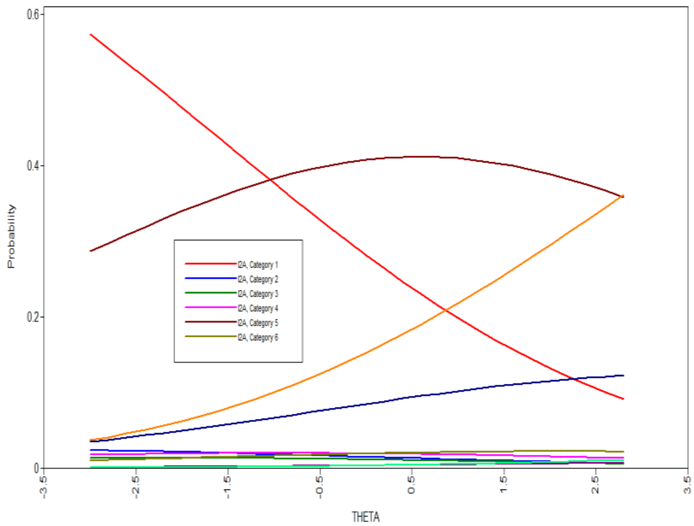
Female i1d

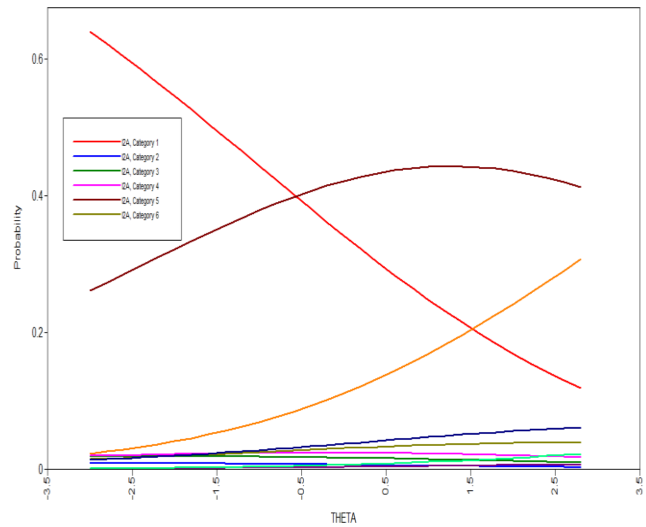




The diagram above shows the Item characteristic curve (ICC) for item 1d for male and female students respectively.To answer research question 4, the ICC above were observed to make comparison.it was seen that the two ICC are similar to one another so i1d is invariant.

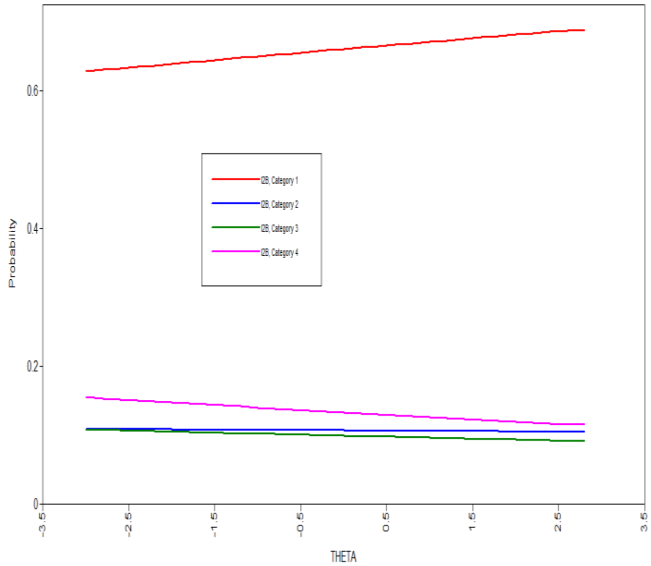
Male i2a Female i2a

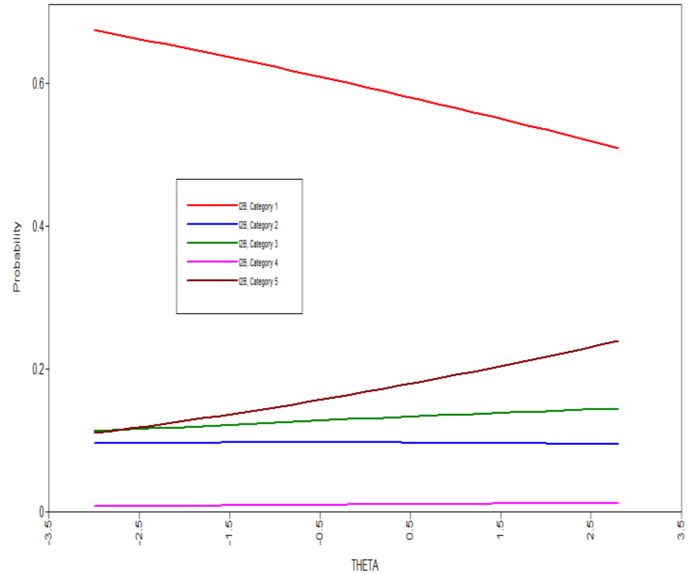




The diagram above shows the Item characteristic curve (ICC) for item 2a for male and female students respectively.To answer research question 4, the ICC above were observed to make comparison.it was seen that the two ICC are similar to one another so i2a is not invariant.

Male i2b Female i2b

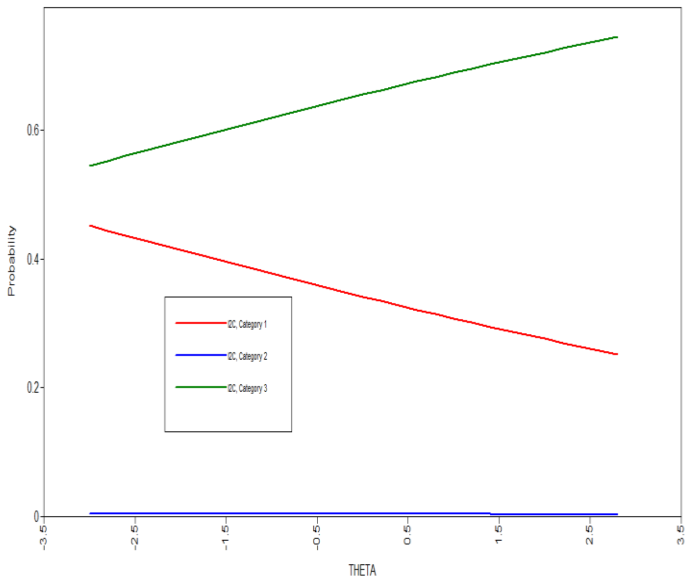


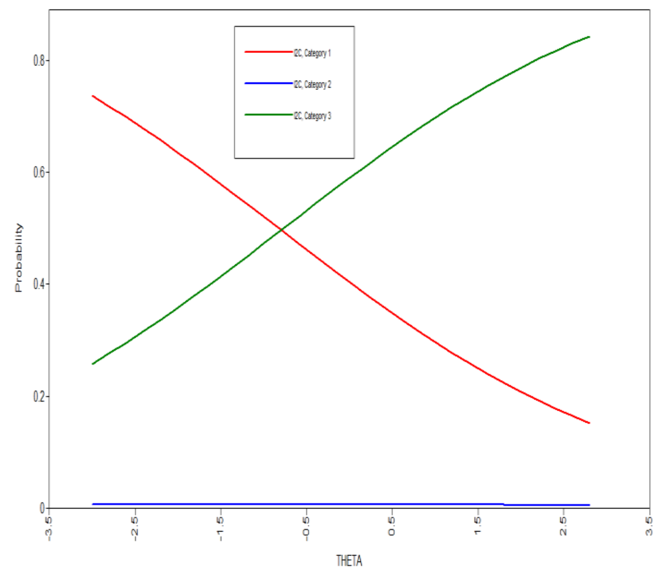


The diagram above shows the Item characteristic curve (ICC) for item 2b for male and female students respectively.To answer research question 4, the ICC above were observed to make comparison.it was seen that the two ICC differs from one another so i2b is not invariant.

Male i2c

Female i2

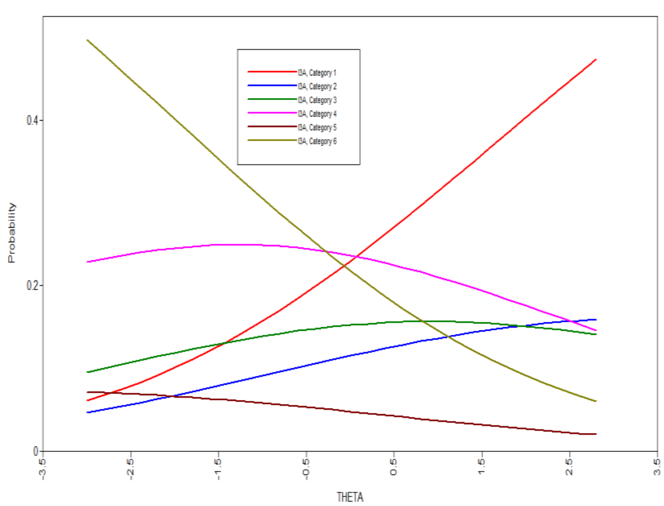


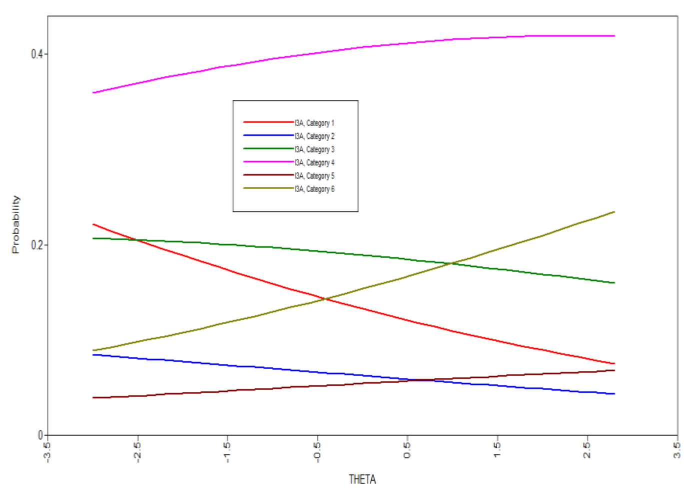


The diagram above shows the Item characteristic curve (ICC) for item 2c for male and female students respectively.To answer research question 4, the ICC above were observed to make comparison.it was seen that the two ICC differs from one another so i2c is not invariant.

Male i3a

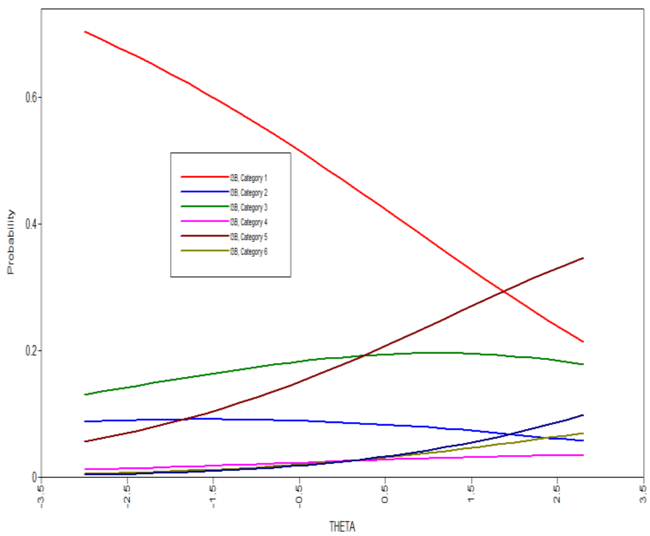
Female i3a



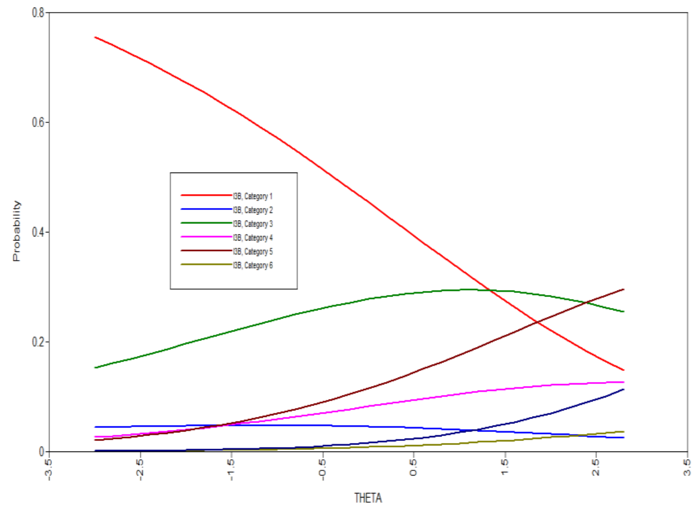


The diagram above shows the Item characteristic curve (ICC) for item 3a for male and female students respectively.To answer research question 4, the ICC above were observed to make comparison.it was seen that the two ICC differs from one another so i3a is not invariant.

Male i3b

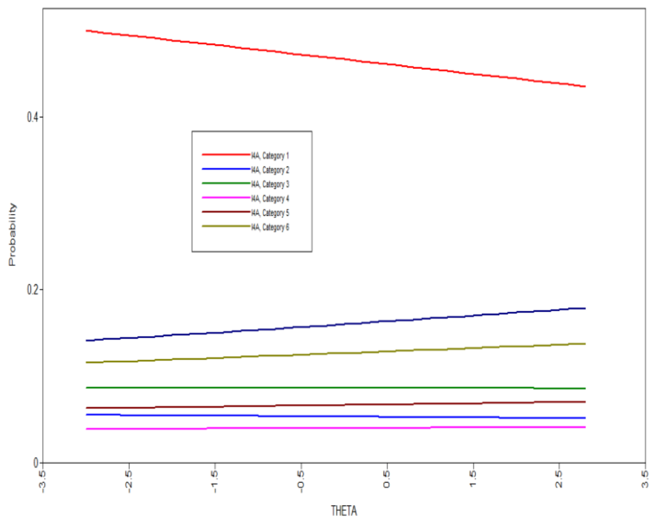


Female i3b

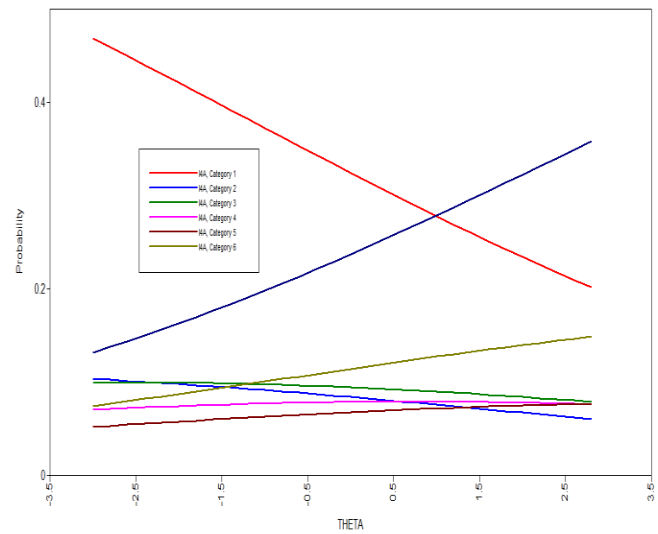


The diagram above shows the Item characteristic curve (ICC) for item 3b for male and female students respectively.To answer research question 4, the ICC above were observed to make comparison.it was seen that the two ICC differs from one another so i3b is not invariant.

Male i4a

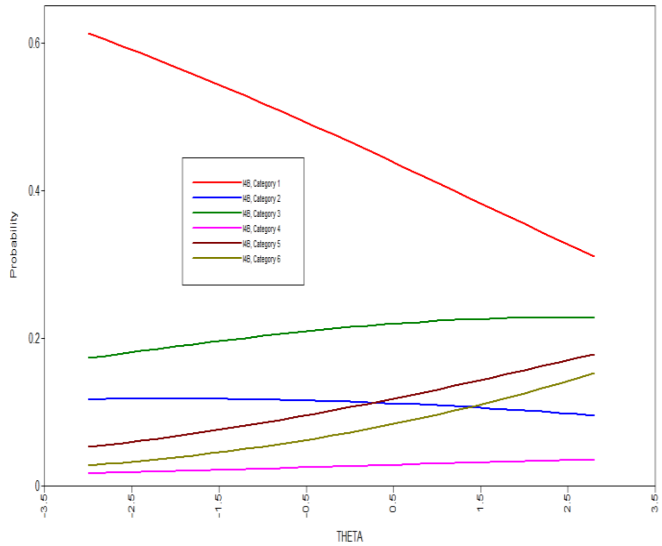


Female i4a

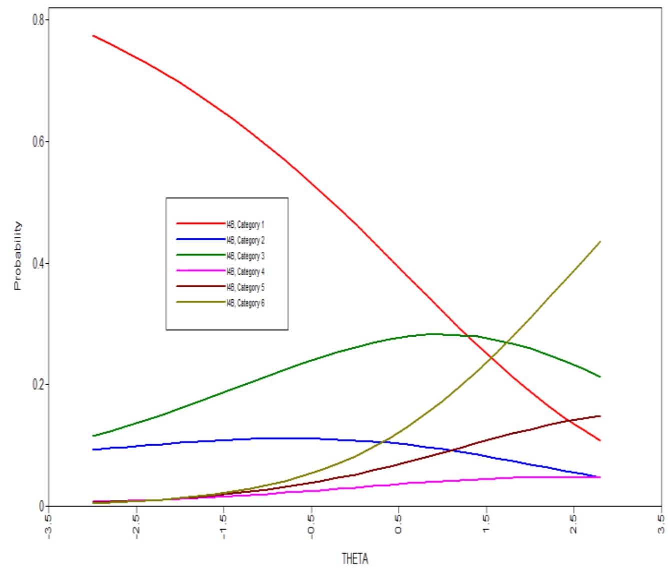


The diagram above shows the Item characteristic curve (ICC) for item 4a for male and female students respectively.To answer research question 4, the ICC above were observed to make comparison.it was seen that the two ICC differs from one another so i4a is not invariant.

Male i4b



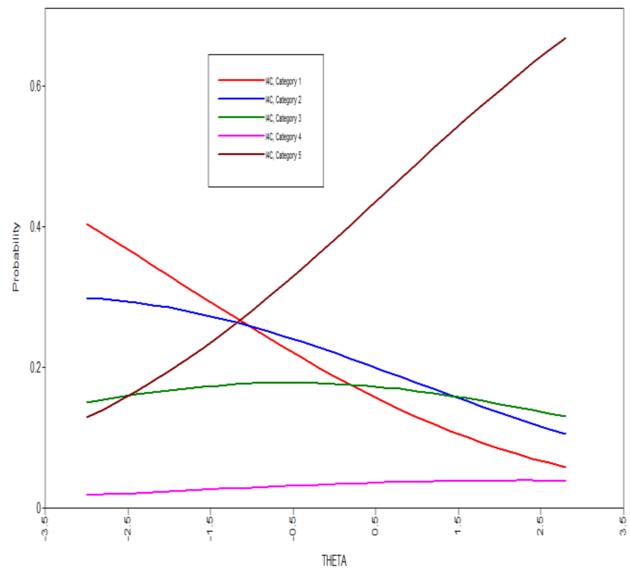
Female i4b

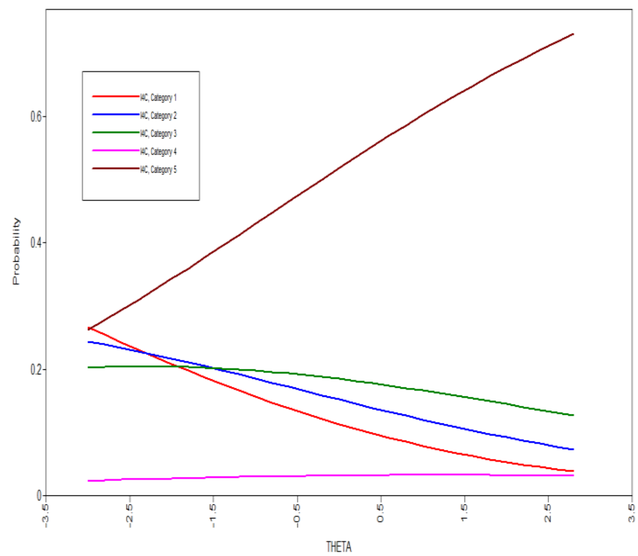


The diagram above shows the Item characteristic curve (ICC) for item 4b for male and female students respectively.To answer research question 4, the ICC above were observed to make comparison.it was seen that the two ICC differs from one another so i4b is not invariant.

Male i4c

Female i4c

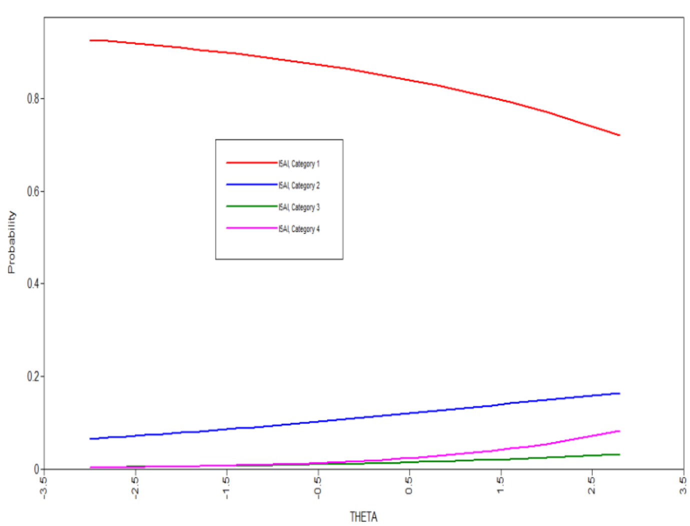


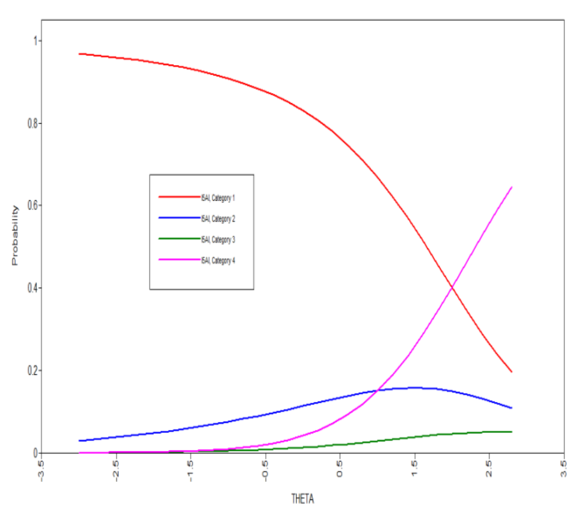


The diagram above shows the Item characteristic curve (ICC) for item 4c for male and female students respectively.To answer research question 4, the ICC above were observed to make comparison.it was seen that the two ICC differs from one another so i4c is not invariant.

Male i5a1

Femalei5al

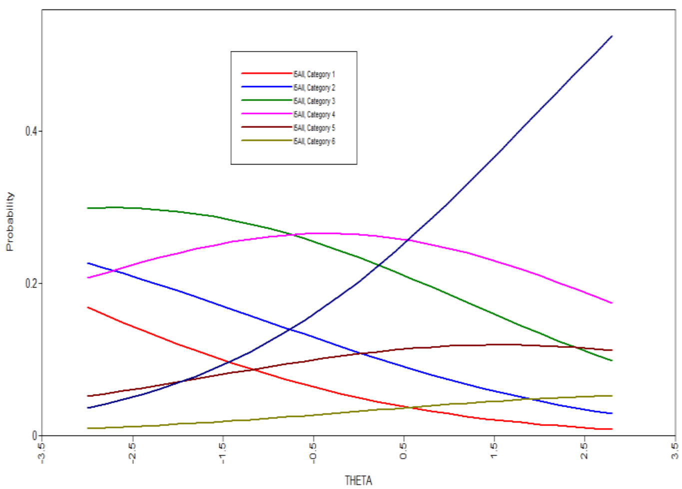


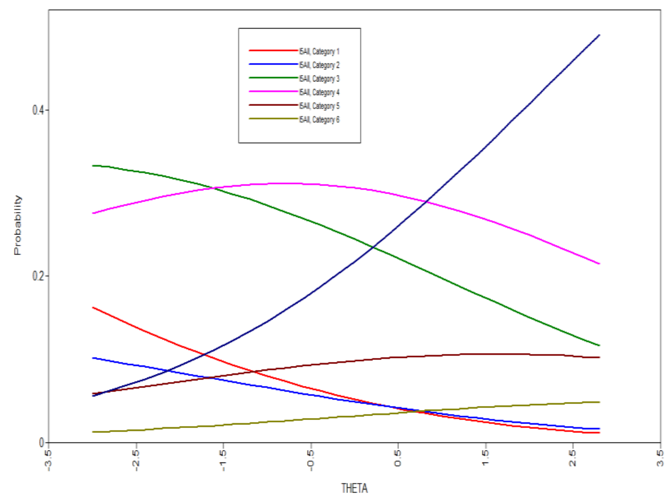


The diagram above shows the Item characteristic curve (ICC) for item i5aI for male and female students respectively.To answer research question 4, the ICC above were observed to make comparison.it was seen that the two ICC differs from one another so i5aI is not invariant.

Male i5aII

Female i5aII

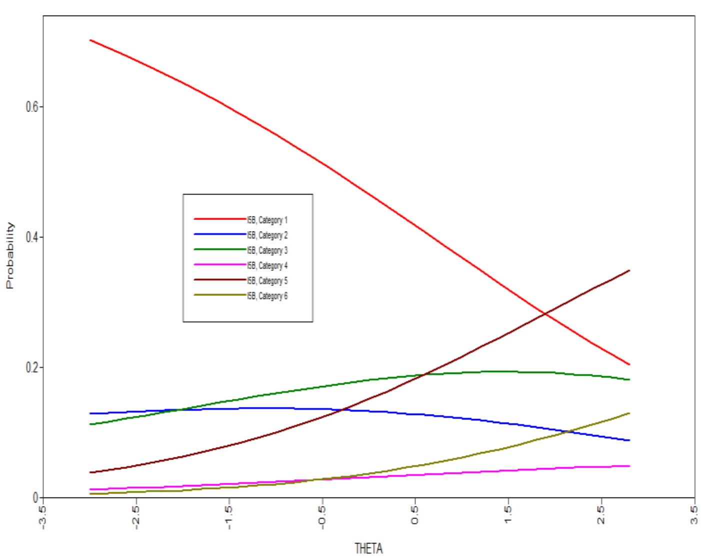


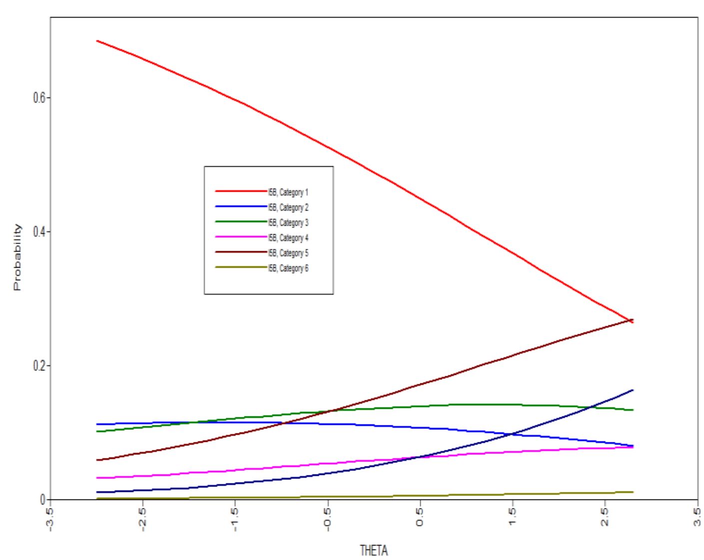


The diagram above shows the Item characteristic curve (ICC) for item i5aII for male and female students respectively.To answer research question 4, the ICC above were observed to make comparison.it was seen that the two ICC are closely similar in shape to one another so i5aII is invariant.

Male i5b

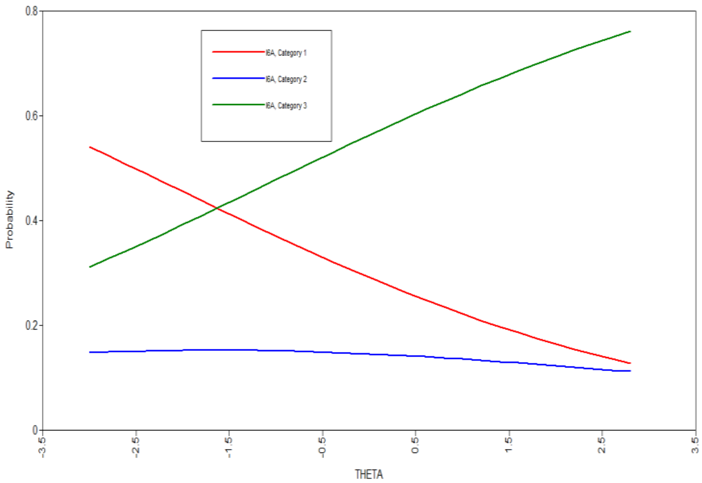
Female i5b



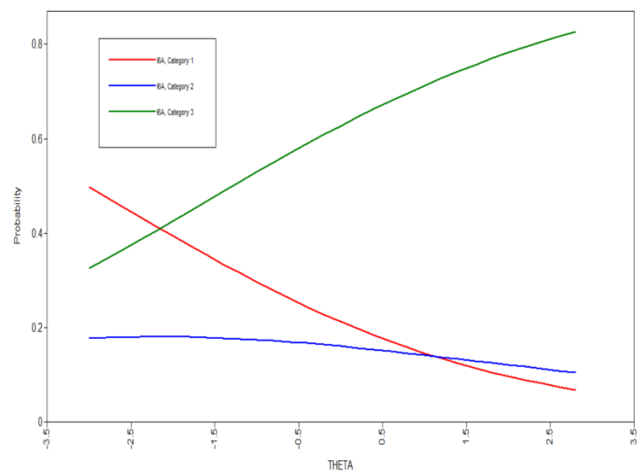


The diagram above shows the Item characteristic curve (ICC) for item i5b for male and female students respectively.To answer research question 4, the ICC above were observed to make comparison.it was seen that the two ICC differs from one another so i5b is not invariant.

Male i6a

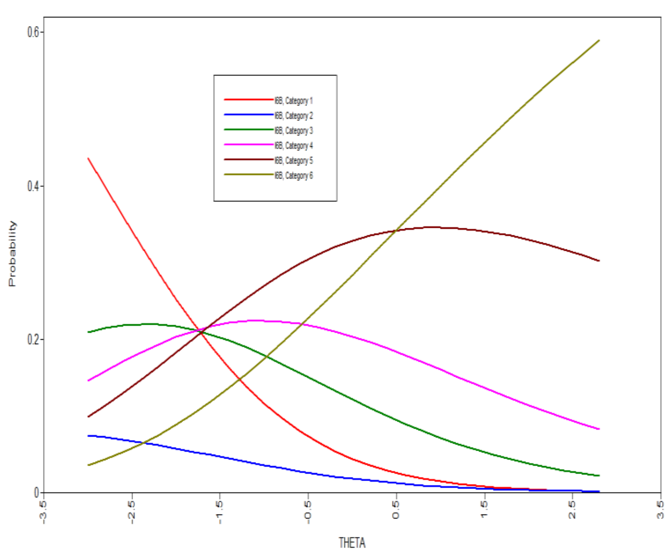


Female i6a

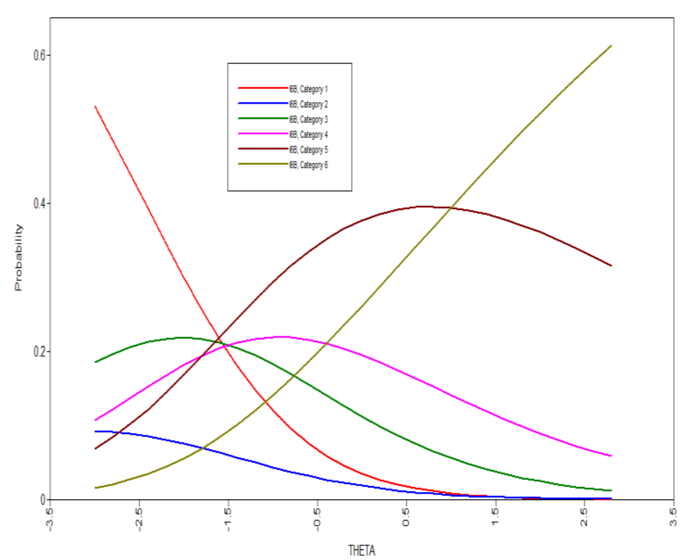


The diagram above shows the Item characteristic curve (ICC) for item i6a for male and female students respectively.To answer research question 4, the ICC above were observed to make comparison.it was seen that the two ICCs are similar to each other so i6a is invariant.

Male i6b

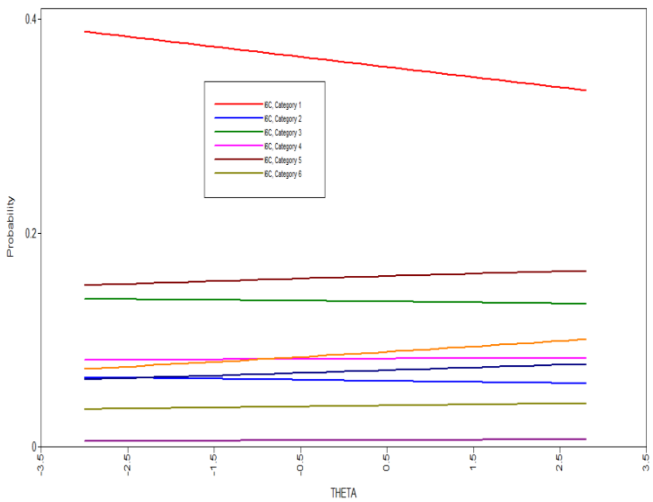


Female i6b

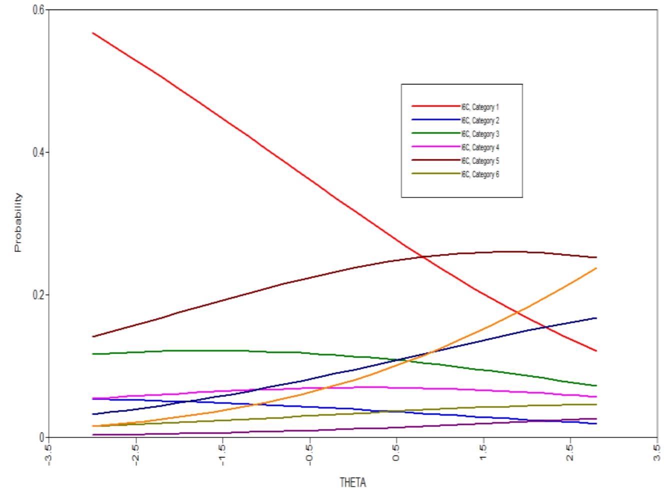


The diagram above shows the Item characteristic curve (ICC) for item i6b for male and female students respectively.To answer research question 4, the ICC above were observed to make comparison.it was seen that the two ICCs are similar to each other so i6b is invariant.

Male i6c

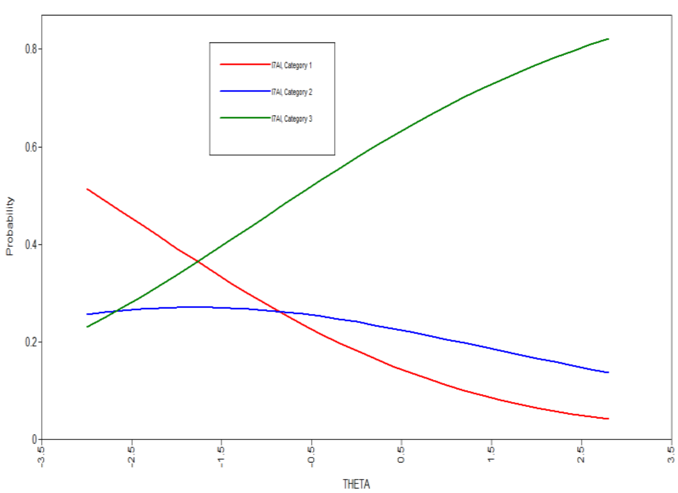


Female i6c

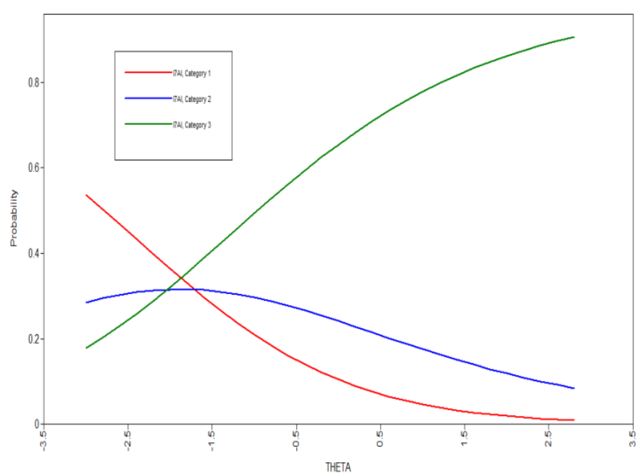


The diagram above shows the Item characteristic curve (ICC) for item i6c for male and female students respectively.To answer research question 4, the ICC above were observed to make comparison.it was seen that the two ICCs differs to each other so i6c is not invariant.

Male i7ai

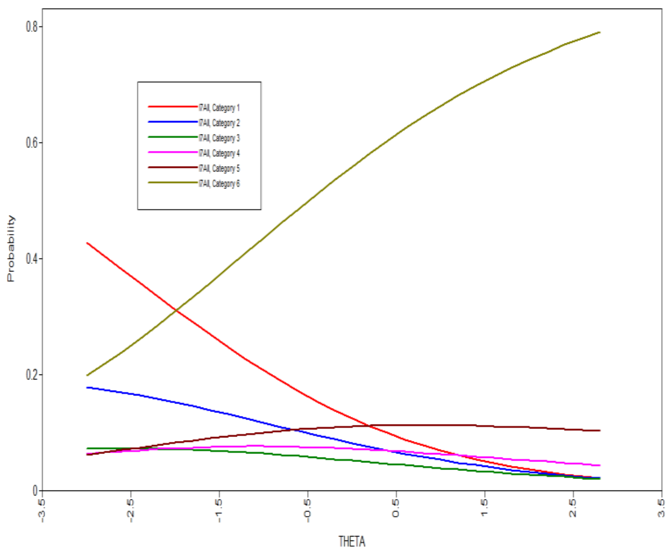


Female i7ai

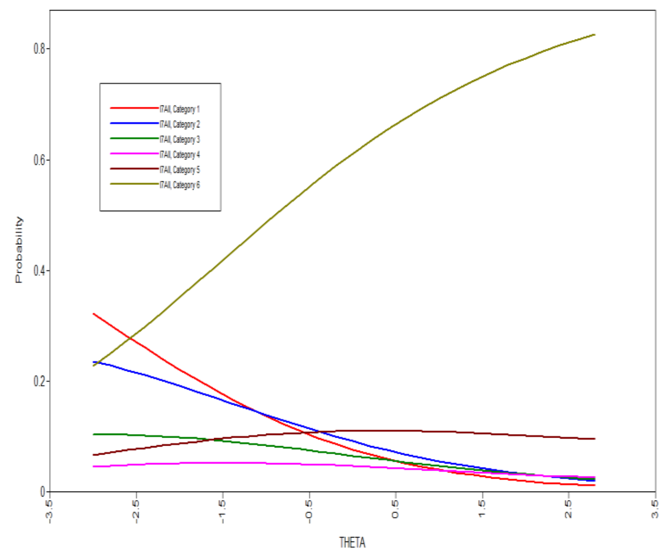


The diagram above shows the Item characteristic curve (ICC) for item i7a for male and female students respectively.To answer research question 4, the ICC above were observed to make comparison.it was seen that the two ICCs are similar to each other so i7a is invariant.

Male i7aii

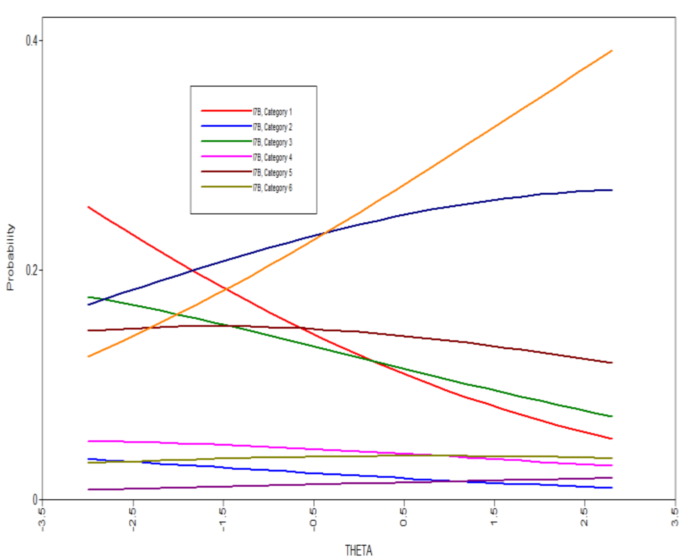


Female i7aii

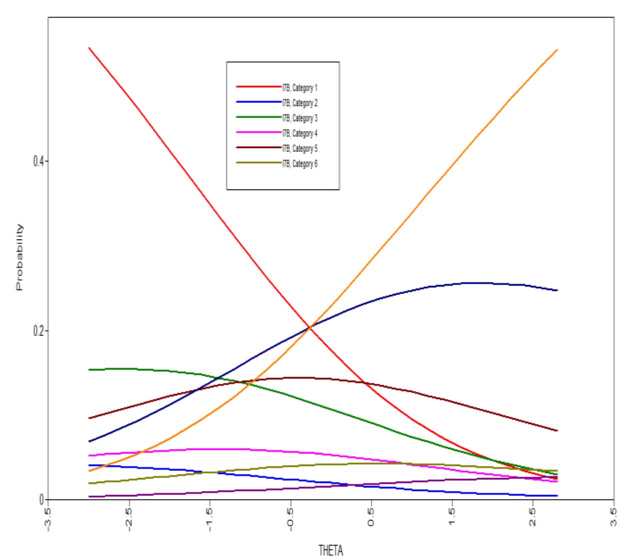


The diagram above shows the Item characteristic curve (ICC) for item i7aii for male and female students respectively.To answer research question 4, the ICC above were observed to make comparison.it was seen that the two ICCs are similar to each other so i7aii is invariant.

Male i7b



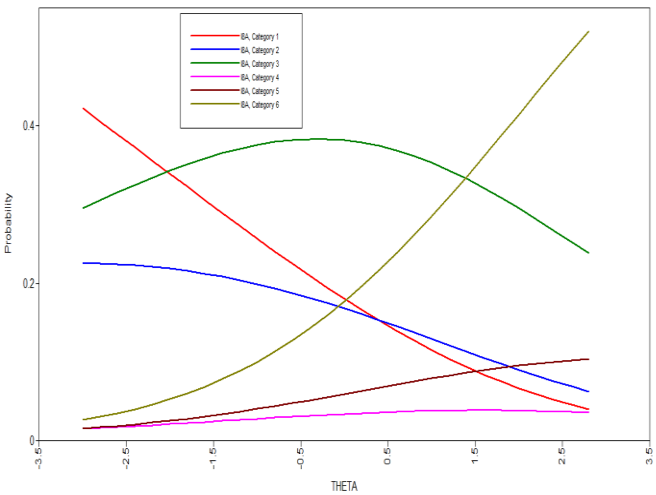
Female i7b

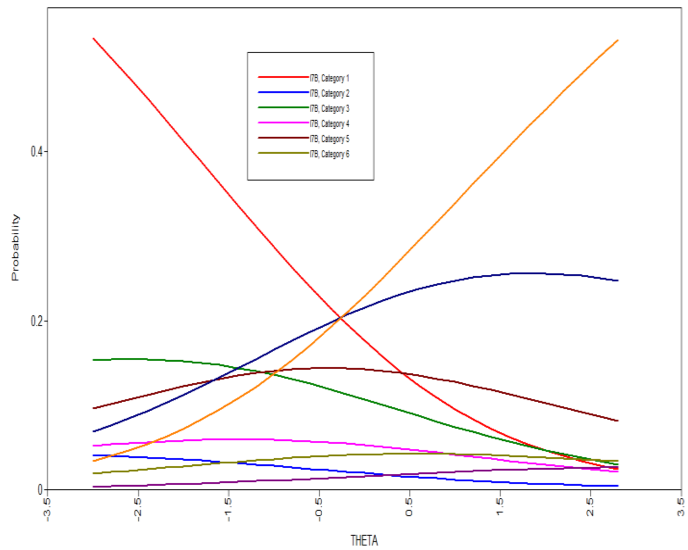


The diagram above shows the Item characteristic curve (ICC) for item i7b for male and female students respectively.To answer research question 4, the ICC above were observed to make comparison.it was seen that the two ICCs differ form each other so i7b is not invariant.

Male i8a

Female i8a

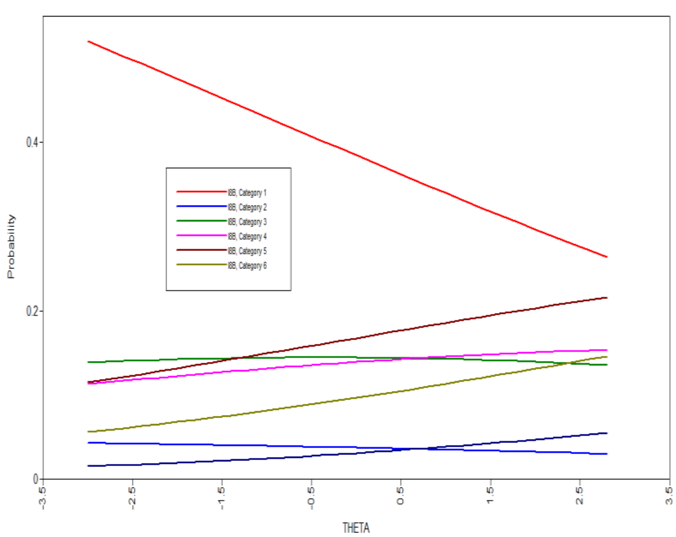


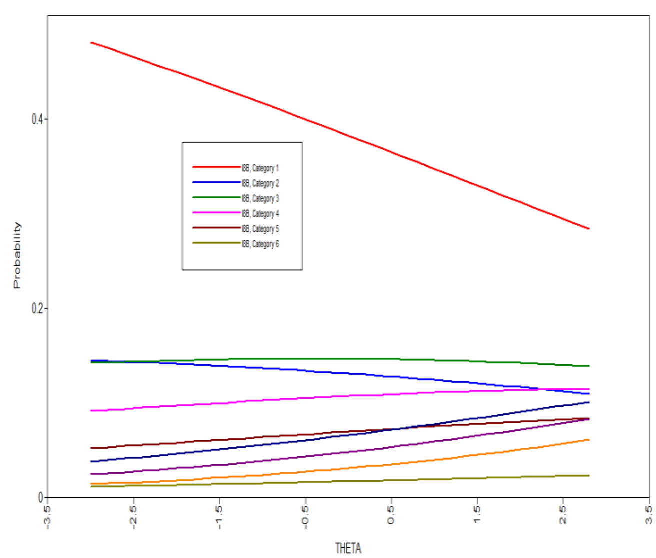


The diagram above shows the Item characteristic curve (ICC) for item i8a for male and female students respectively.To answer research question 4, the ICC above were observed to make comparison.it was seen that the two ICCs differ form each other so i8a is not invariant.

Male i8b

Female i8b

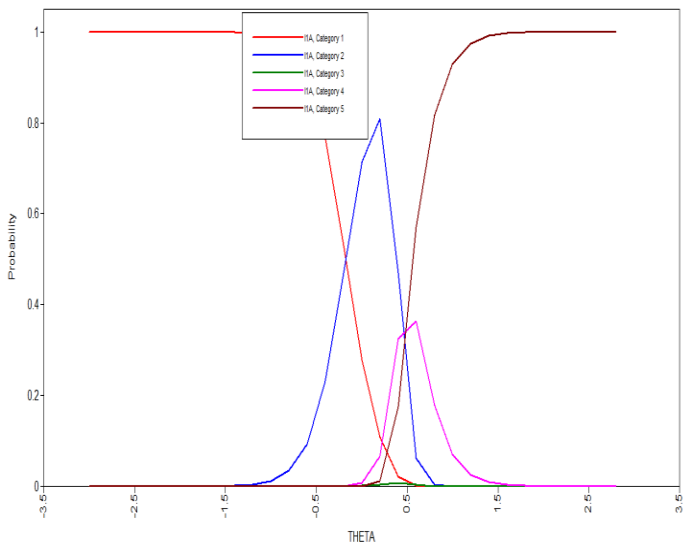




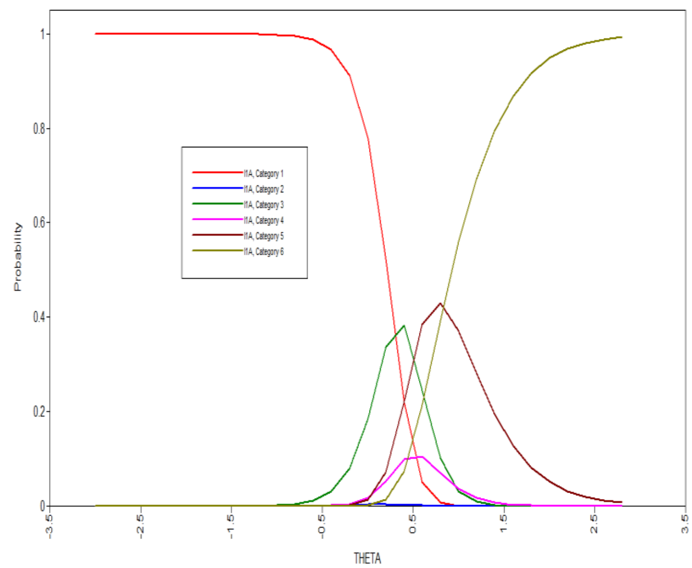
The diagram above shows the Item characteristic curve (ICC) for item i8b for male and female students respectively.To answer research question 4, the ICC above were observed to make comparison.it was seen that the two ICCs differ form each other so i8b is not invariant.

APPENDIX VIII: URBAN AND RURAL ITEM CHARACTERISTIC CURVES

The CRC of the items on the test were compared. The results are presented in Figure rural and urban i1a- rural and urban i8b



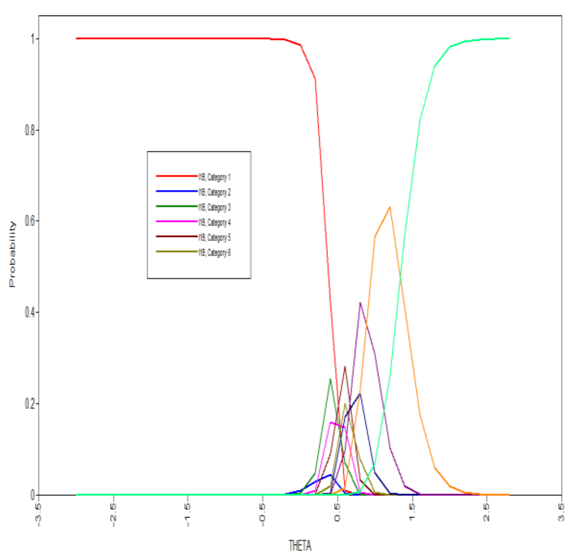
Rural i1a



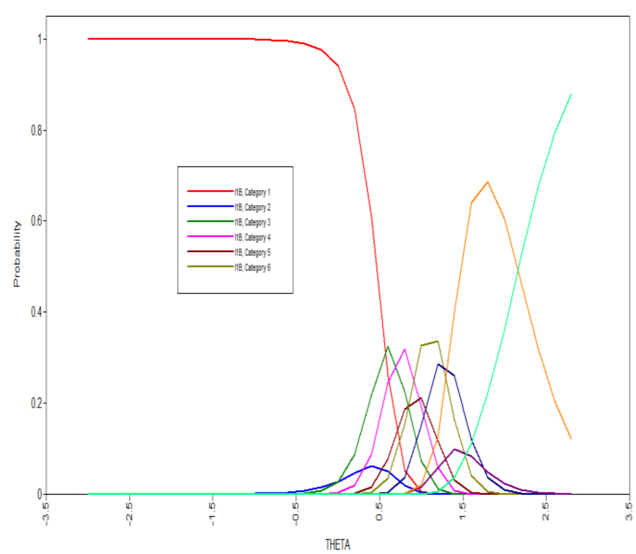
Urban i1a

The diagram above shows the Item characteristic curve (ICC) for item i1a for rural and urban students respectively.To answer research question 5, the ICC above were observed to make comparison.it was seen that the two ICCs differ form each other so i1a is not invariant.

Rural i1b

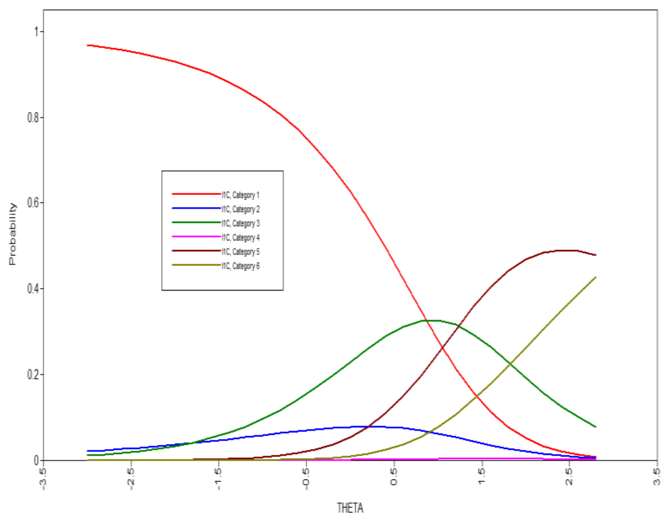


Urban i1b

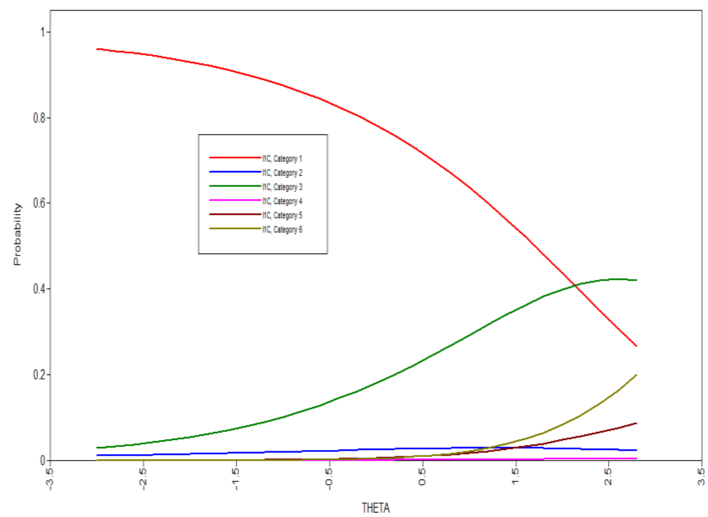


The diagram above shows the Item characteristic curve (ICC) for item i1b for rural and urban students respectively.To answer research question 5, the ICC above were observed to make comparison.it was seen that the two ICCs differ form each other so i1b is not invariant.

Rural i1c

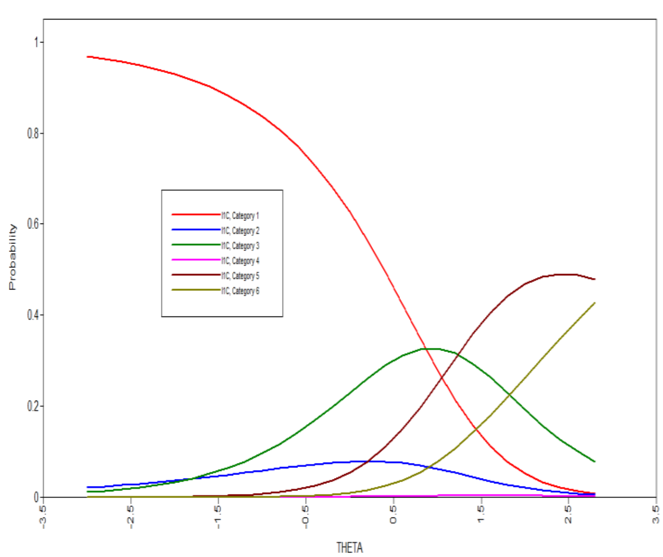


Urban i1c

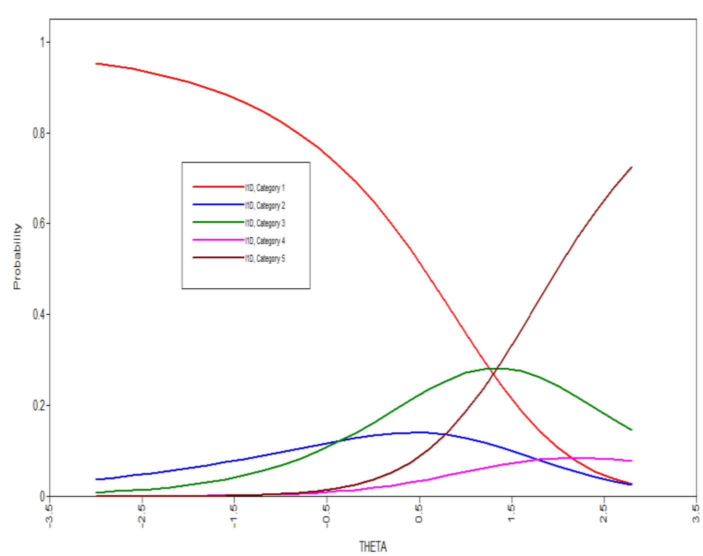


The diagram above shows the Item characteristic curve (ICC) for item i1c for rural and urban students respectively.To answer research question 5, the ICC above were observed to make comparison.it was seen that the two ICCs differ form each other so i1c is not invariant.

Rural i1d



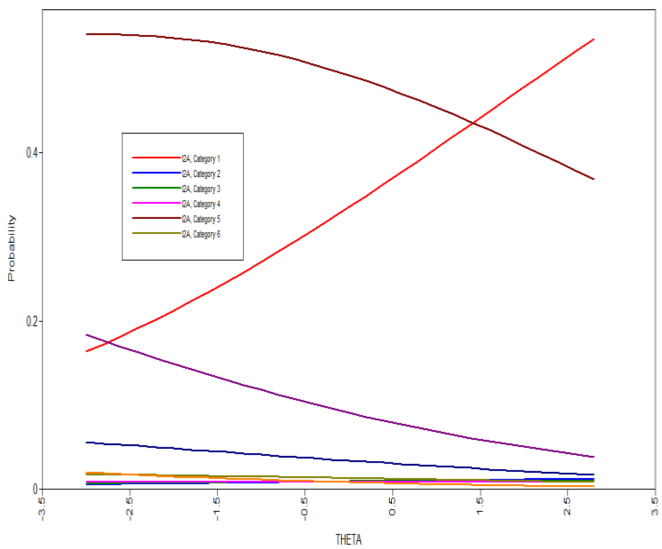
Urban i1d

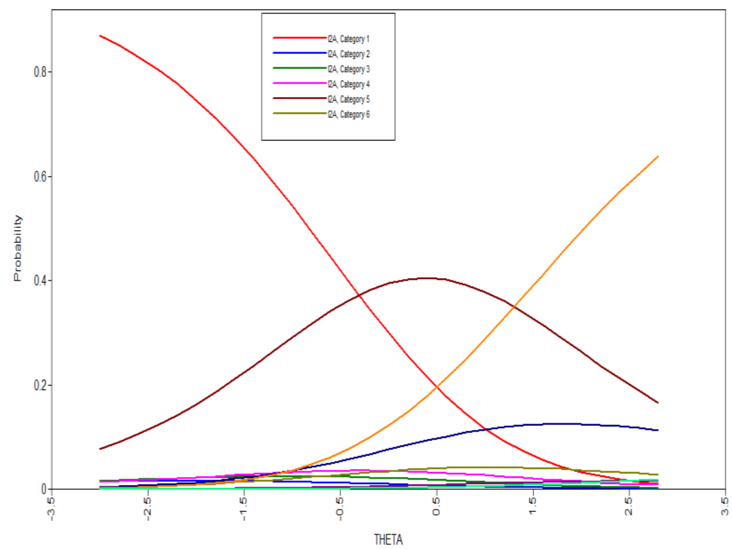


The diagram above shows the Item characteristic curve (ICC) for item i1d for rural and urban students respectively.To answer research question 5, the ICC above were observed to make comparison.it was seen that the two ICCs differ form each other so i1d is not invariant.

Rural i2a

Urban i2a

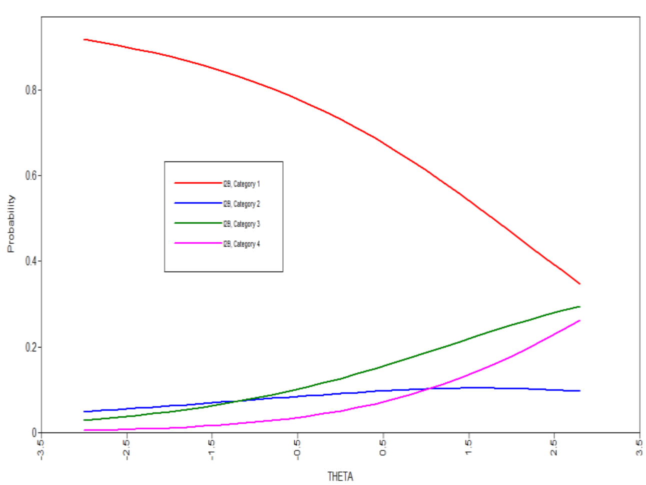


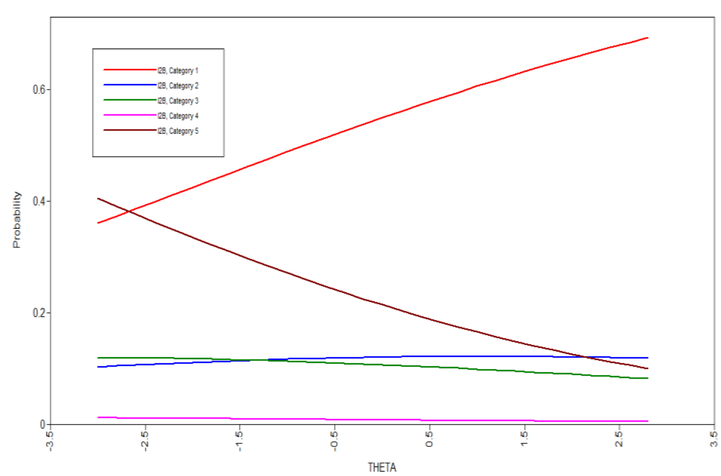


The diagram above shows the Item characteristic curve (ICC) for item i2a for rural and urban students respectively.To answer research question 5, the ICC above were observed to make comparison.it was seen that the two ICCs differ form each other so i2a is not invariant.

Rural i2b

Urban i2b

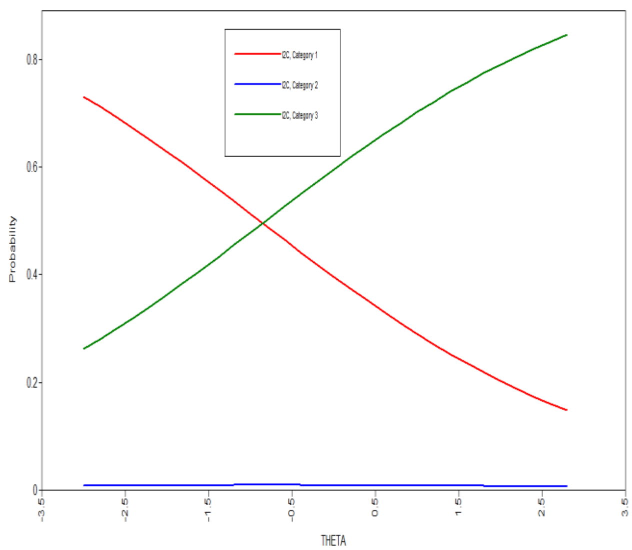


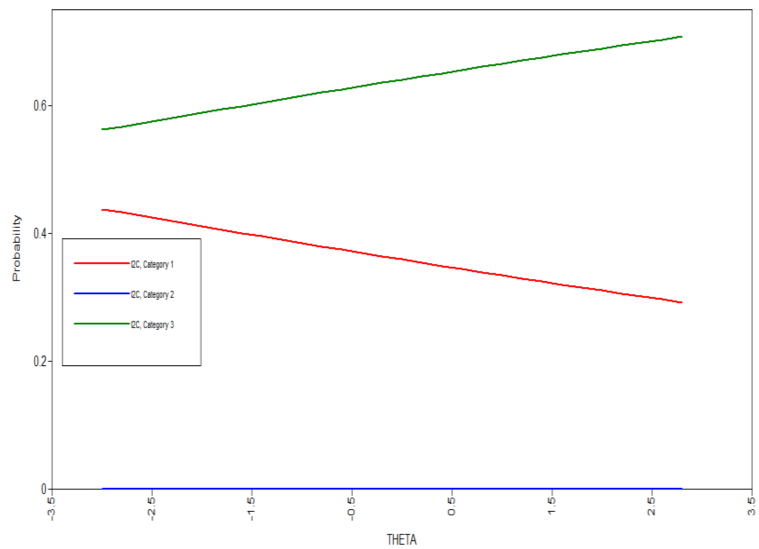


The diagram above shows the Item characteristic curve (ICC) for item i2b for rural and urban students respectively.To answer research question 5, the ICC above were observed to make comparison.it was seen that the two ICCs differ form each other so i2b is not invariant.

Rural i2c

Urban i2c

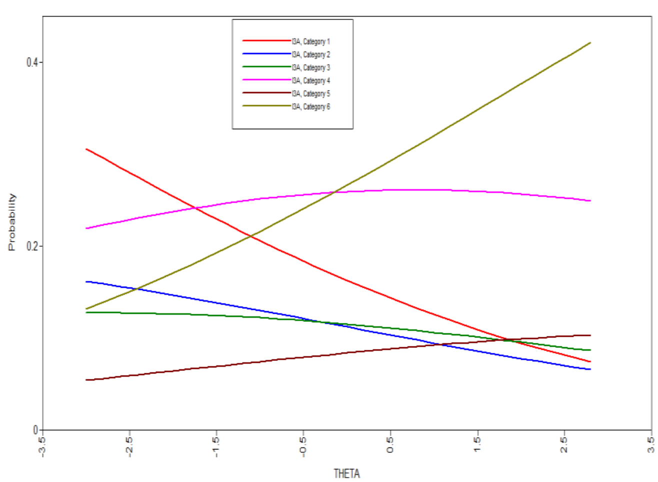


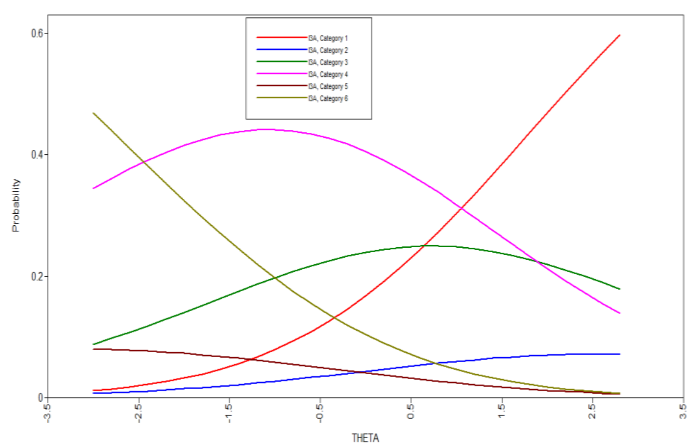


The diagram above shows the Item characteristic curve (ICC) for item i2c for rural and urban students respectively.To answer research question 5, the ICC above were observed to make comparison.it was seen that the two ICCs differ form each other so i2c is not invariant.

Rural i3a

Urban i3a

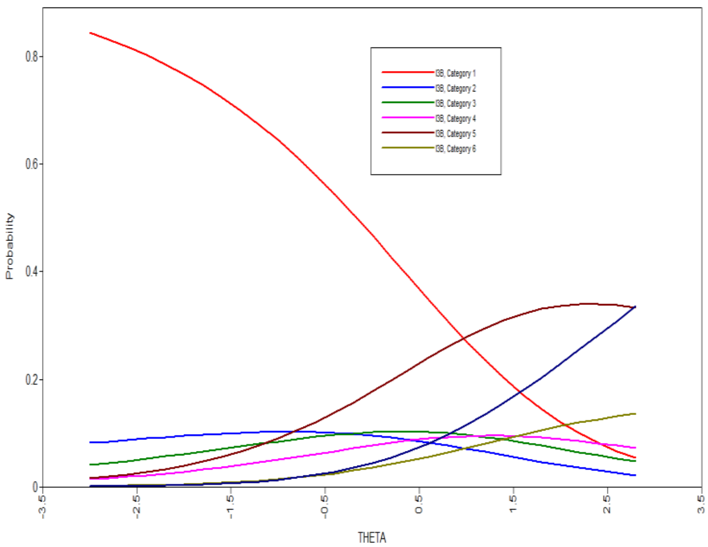


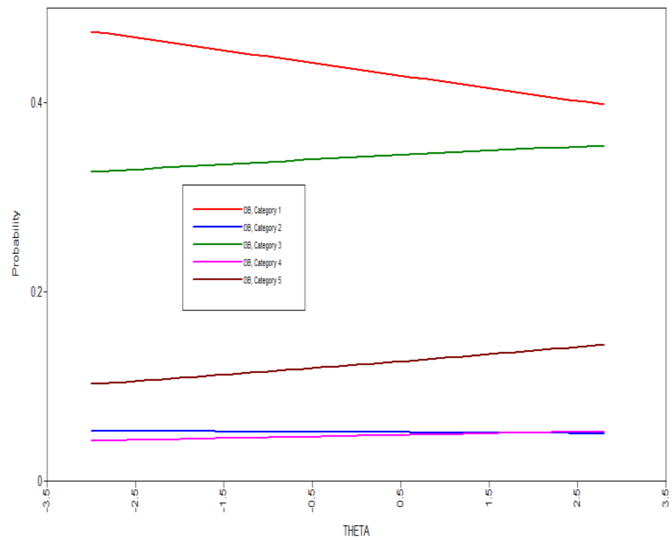


The diagram above shows the Item characteristic curve (ICC) for item i3a for rural and urban students respectively.To answer research question 5, the ICC above were observed to make comparison.it was seen that the two ICCs differ form each other so i3a is not invariant.

Rural i3b

Urban i3b

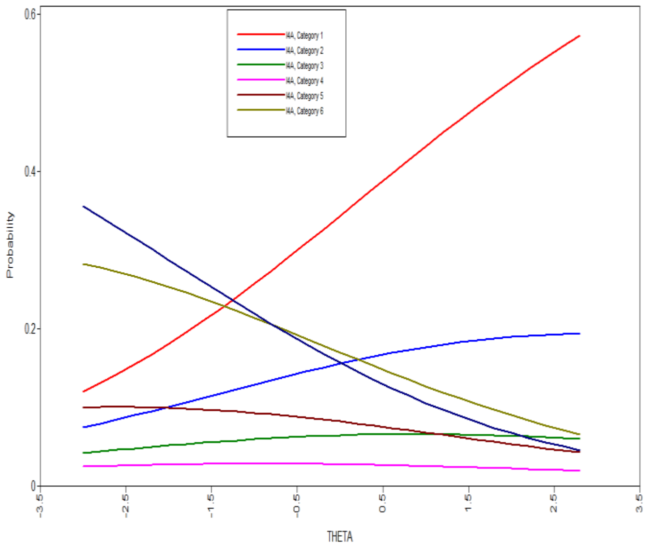


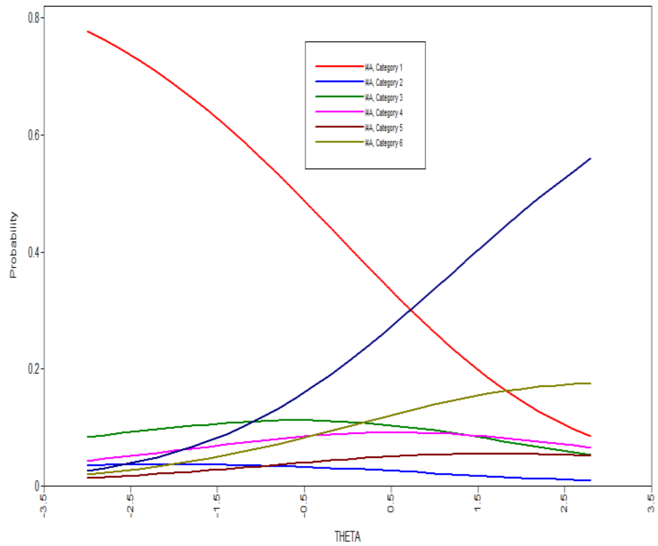


The diagram above shows the Item characteristic curve (ICC) for item i3b for rural and urban students respectively.To answer research question 5, the ICC above were observed to make comparison.it was seen that the two ICCs differ form each other so i3b is not invariant.

Rural i4a

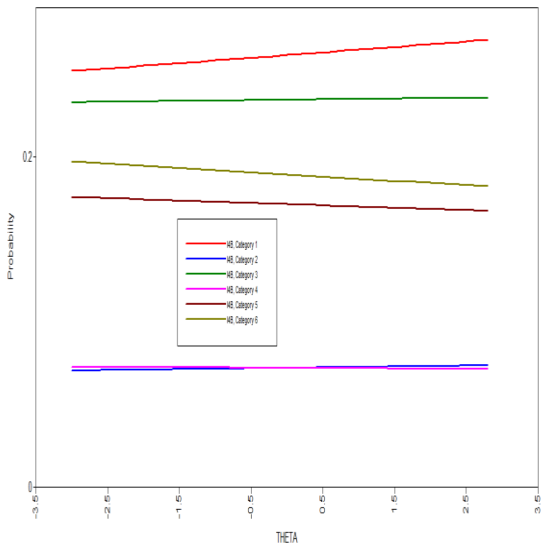
Urban i4a



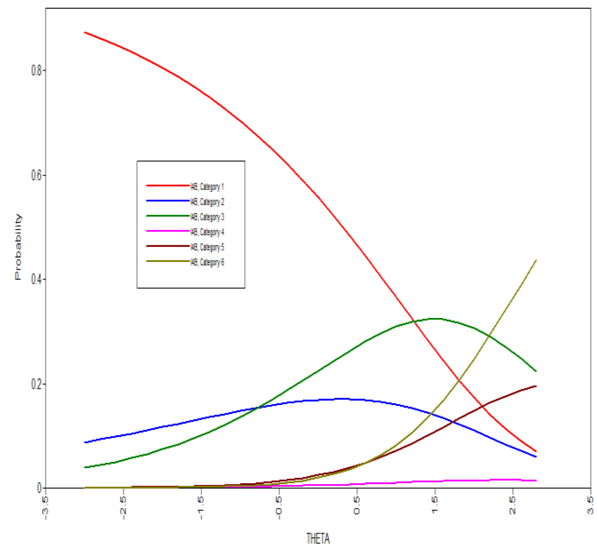


The diagram above shows the Item characteristic curve (ICC) for item i4a for rural and urban students respectively.To answer research question 5, the ICC above were observed to make comparison.it was seen that the two ICCs differ form each other so i4a is not invariant.

Rural i4b

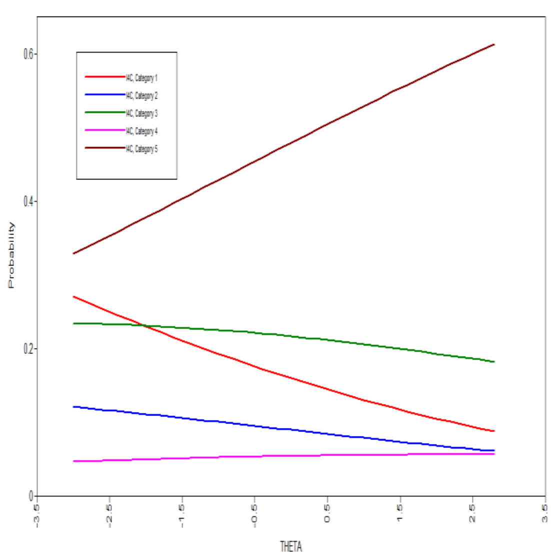


Urban i4b

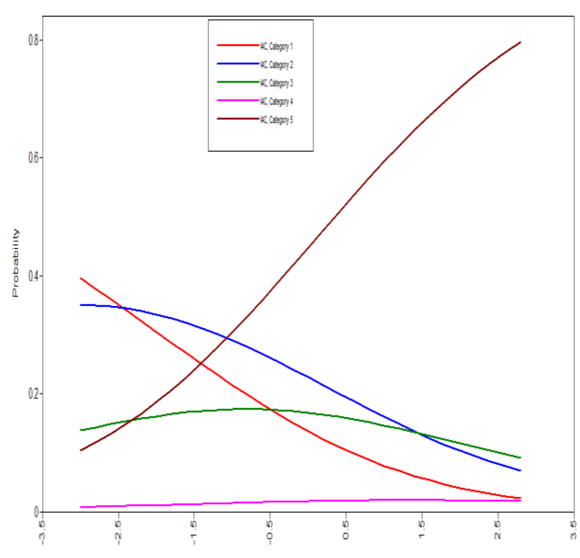


The diagram above shows the Item characteristic curve (ICC) for item i4b for rural and urban students respectively.To answer research question 5, the ICC above were observed to make comparison.it was seen that the two ICCs differ form each other so i4b is not invariant.

Rural i4c

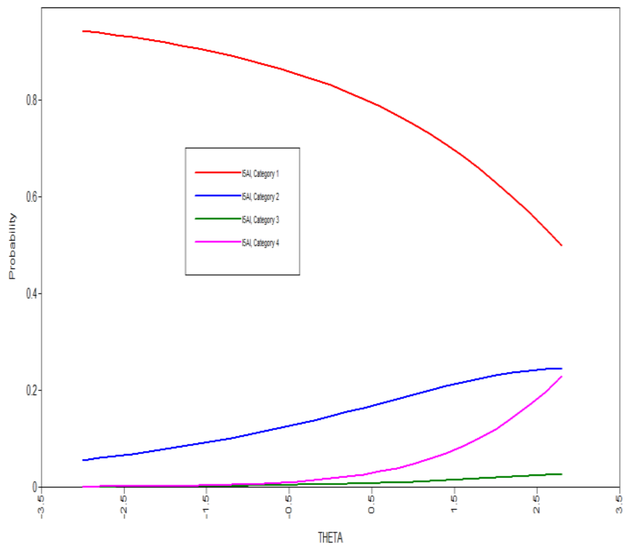


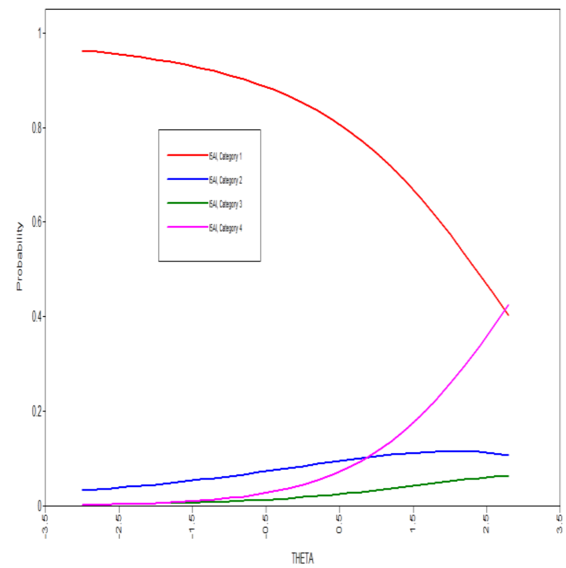
Urban i4c



The diagram above shows the Item characteristic curve (ICC) for item i4c for rural and urban students respectively.To answer research question 5, the ICC above were observed to make comparison.it was seen that the two ICCs differ form each other so i4c is not invariant.

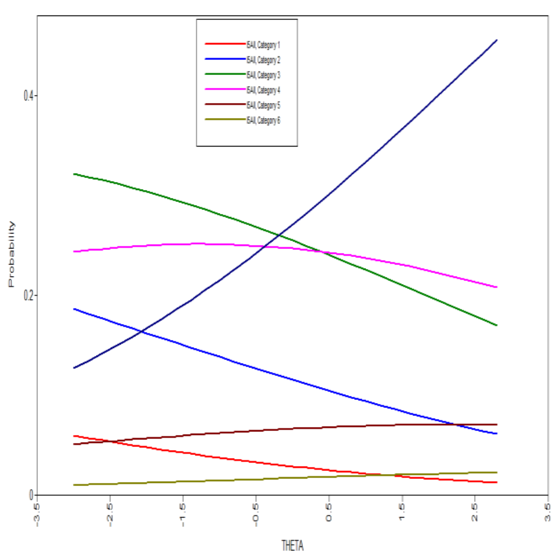
Rural i5a1 Urban i5a1

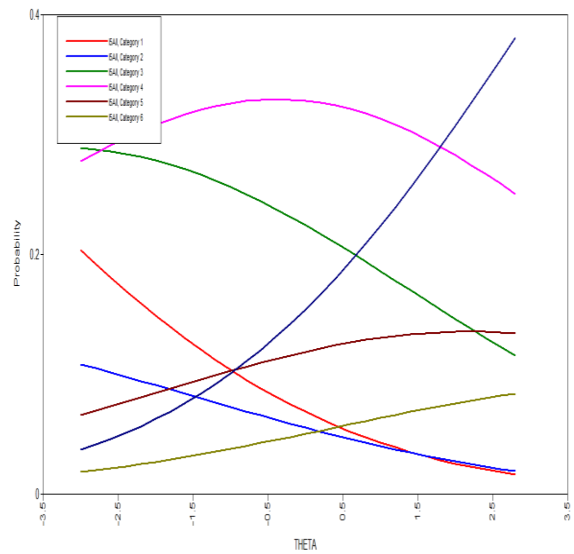




The diagram above shows the Item characteristic curve (ICC) for item i5a1 for rural and urban students respectively.To answer research question 5, the ICC above were observed to make comparison.it was seen that the two ICCs differ form each other so i5a1 is not invariant.

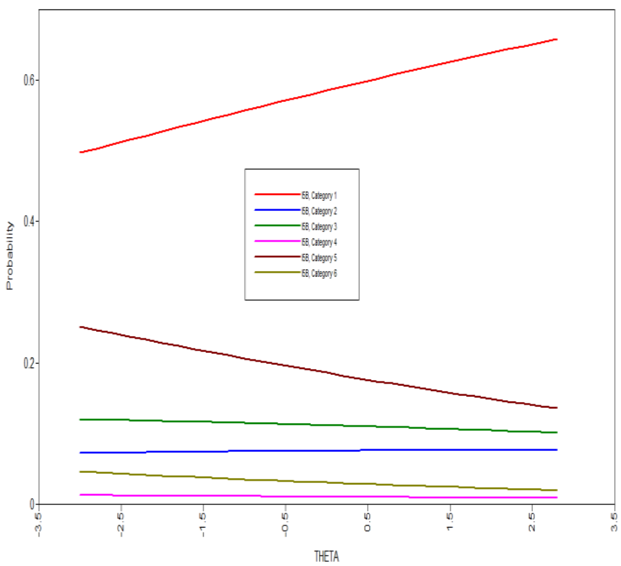
Rural i5aii Urban i5aii



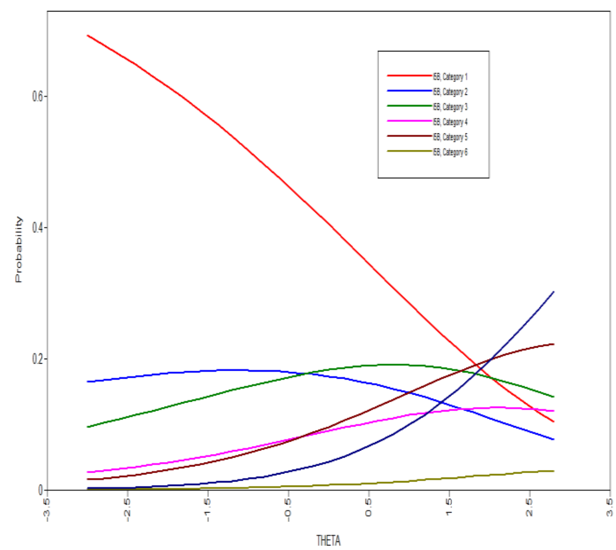


The diagram above shows the Item characteristic curve (ICC) for item i5aii for rural and urban students respectively.To answer research question 5, the ICC above were observed to make comparison.it was seen that the two ICCs differ form each other so i5aii is not invariant.

Rural i5b

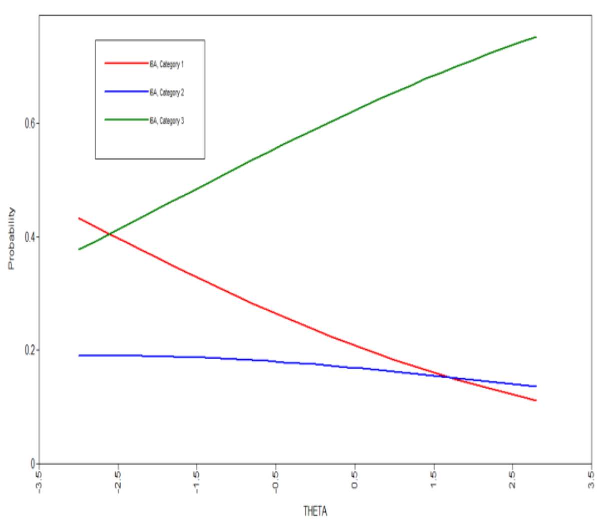


Urban i5b

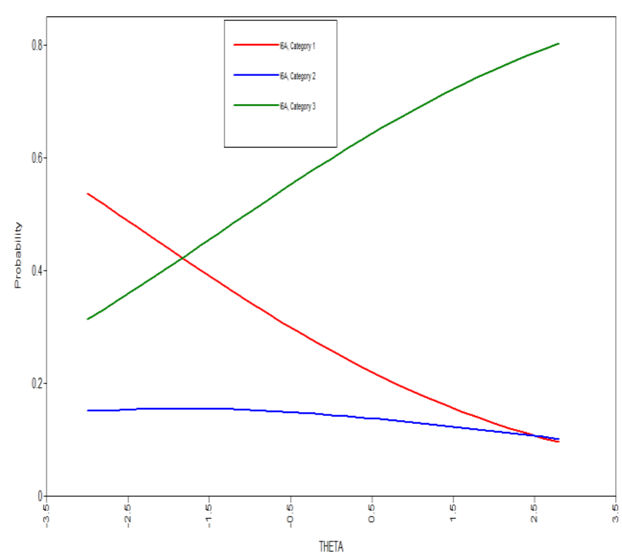


The diagram above shows the Item characteristic curve (ICC) for item i5b for rural and urban students respectively.To answer research question 5, the ICC above were observed to make comparison.it was seen that the two ICCs differ form each other so i5b is not invariant.

Rural i6a

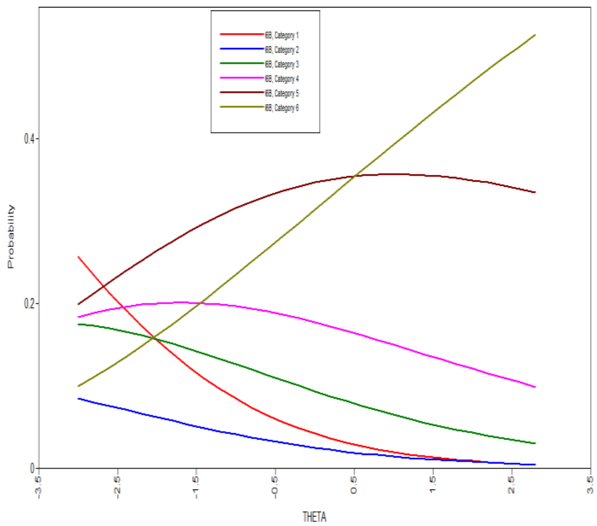


Urban i6a

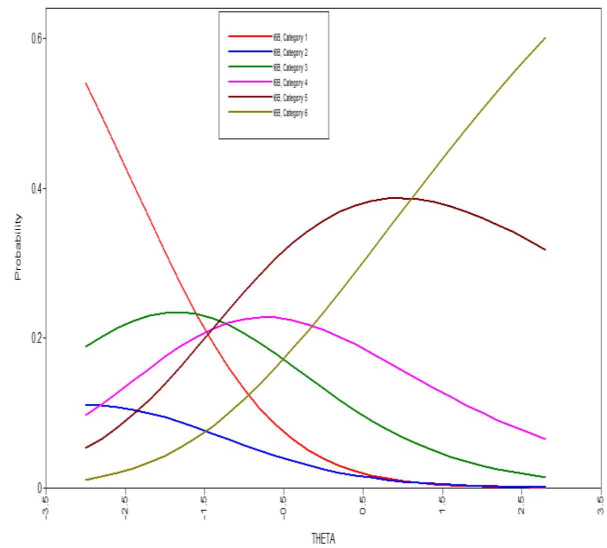


The diagram above shows the Item characteristic curve (ICC) for item i6a for rural and urban students respectively.To answer research question 5, the ICC above were observed to make comparison.it was seen that the two ICCs are similar to each other so i6a invariant.

Rural i6b

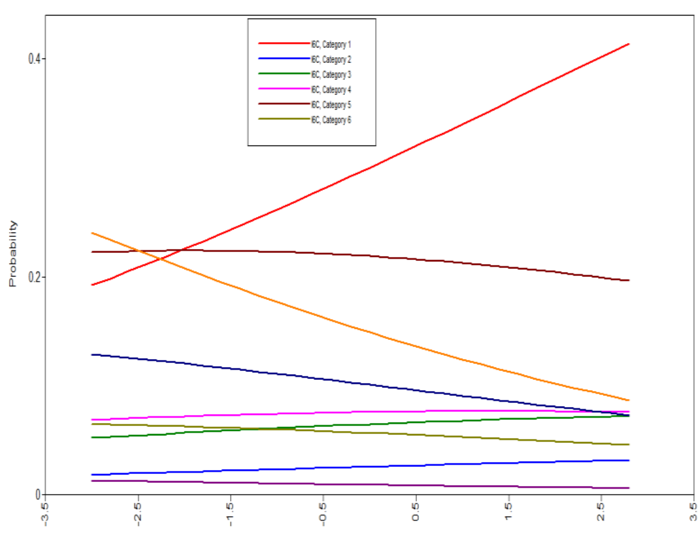


Urban i6b

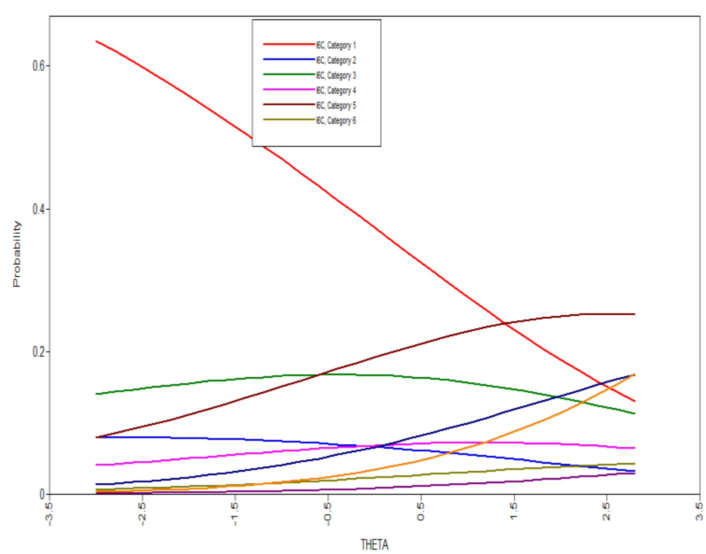


The diagram above shows the Item characteristic curve (ICC) for item i6b for rural and urban students respectively.To answer research question 5, the ICC above were observed to make comparison.it was seen that the two ICCs differ from each other so i6b is not invariant.

rural i6c



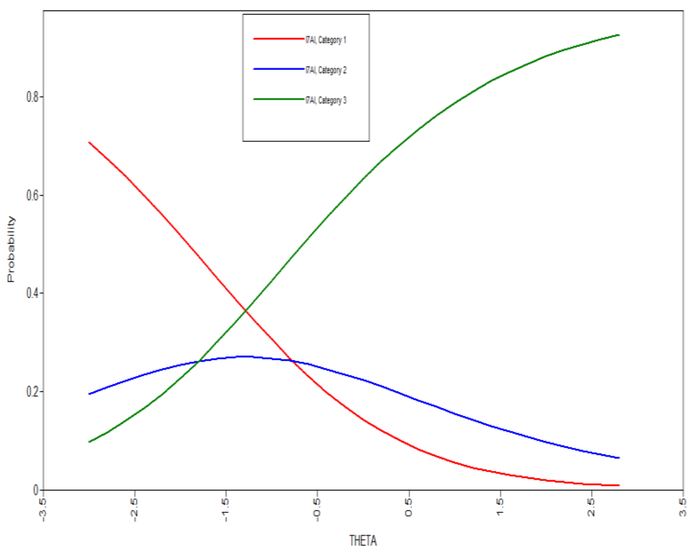
Urban i6c

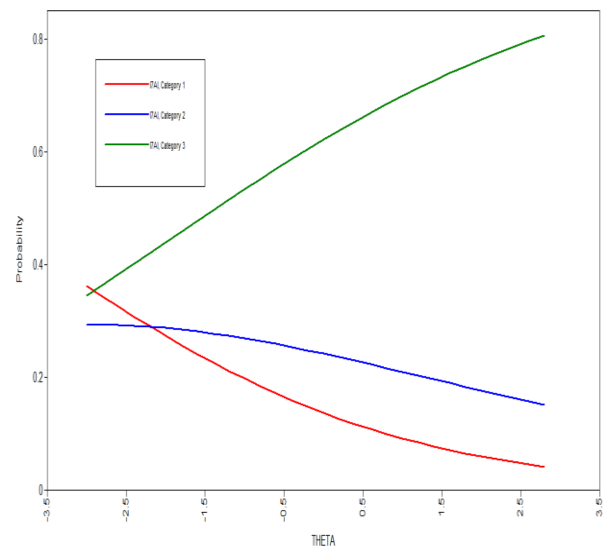


The diagram above shows the Item characteristic curve (ICC) for item i6c for rural and urban students respectively.To answer research question 5, the ICC above were observed to make comparison.it was seen that the two ICCs differ from each other so i6c is not invariant.

Rural i7ai

Urban i7ai

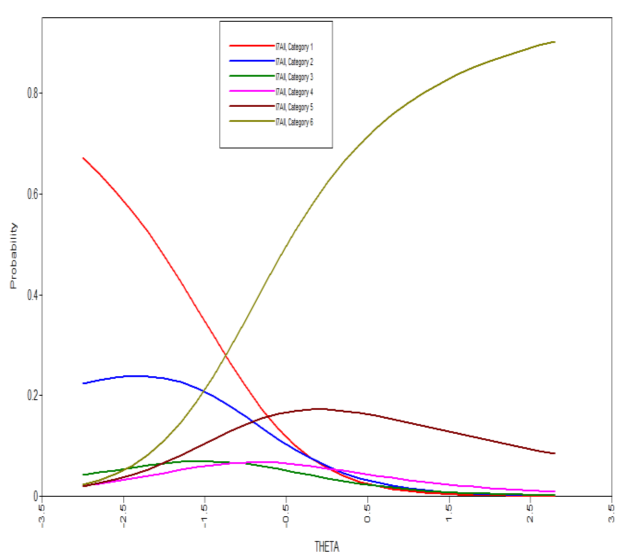


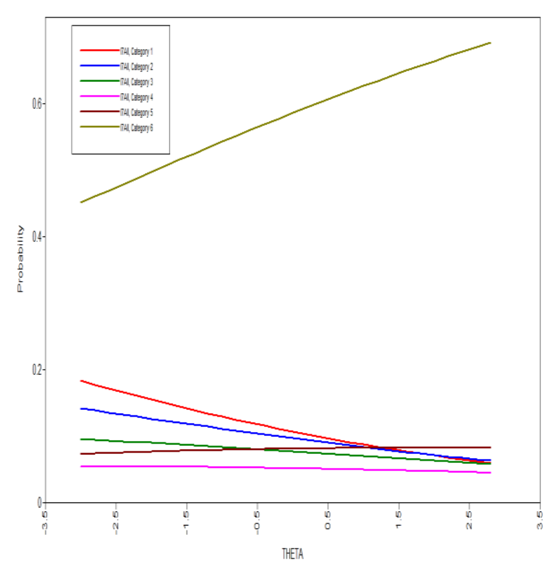


The diagram above shows the Item characteristic curve (ICC) for item i7ai for rural and urban students respectively.To answer research question 5, the ICC above were observed to make comparison.it was seen that the two ICCs differ from each other so i7ai is not invariant.

Rural i7aii

Urban i7aii

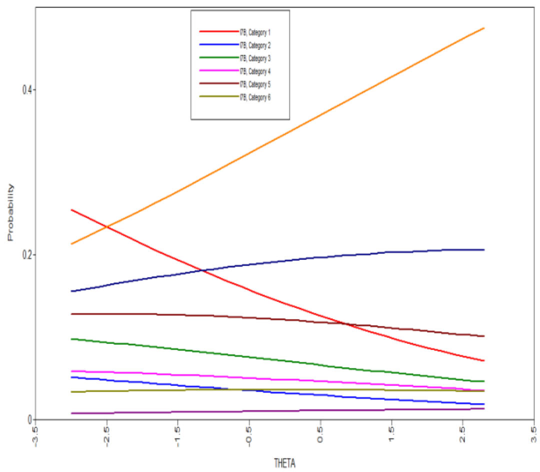


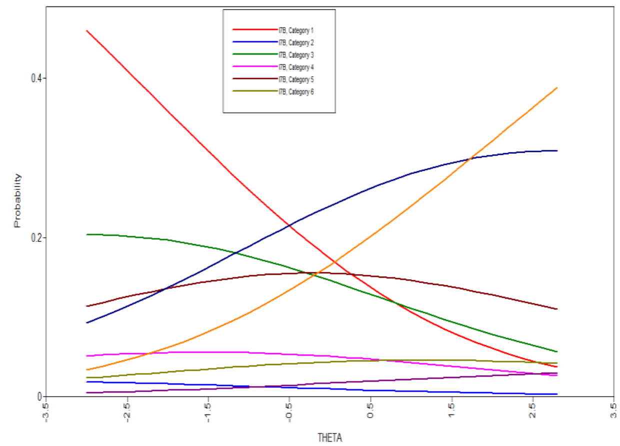


The diagram above shows the Item characteristic curve (ICC) for item i7aii for rural and urban students respectively.To answer research question 5, the ICC above were observed to make comparison.it was seen that the two ICCs differ from each other so i7aii is not invariant.

Rural i7b

Urban i7b

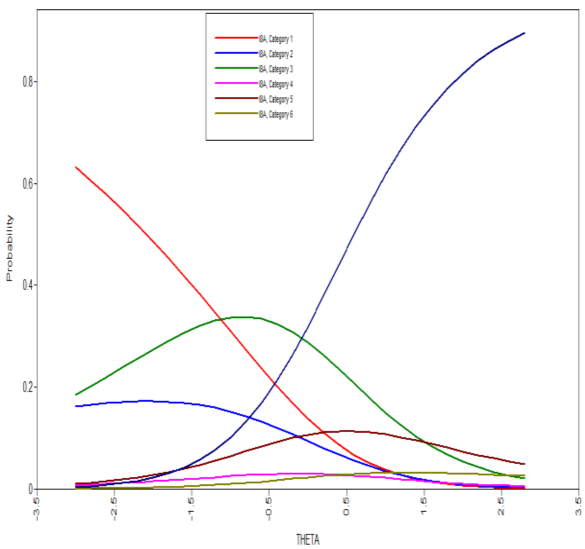


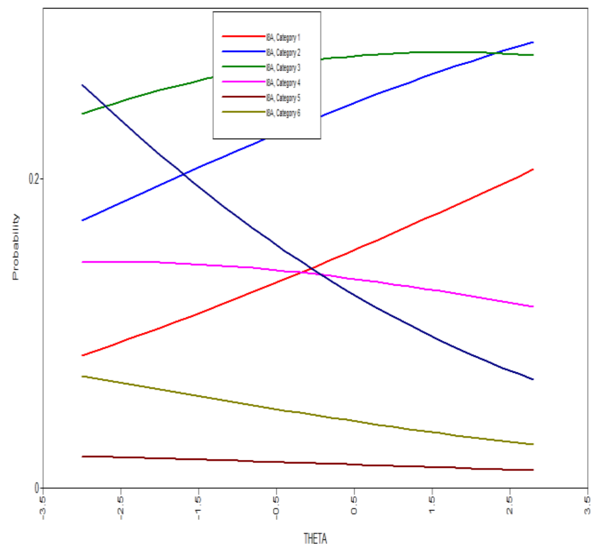


The diagram above shows the Item characteristic curve (ICC) for item i7b for rural and urban students respectively.To answer research question 5, the ICC above were observed to make comparison.it was seen that the two ICCs differ from each other so i7b is not invariant.

Rural i8a

Urban i8a

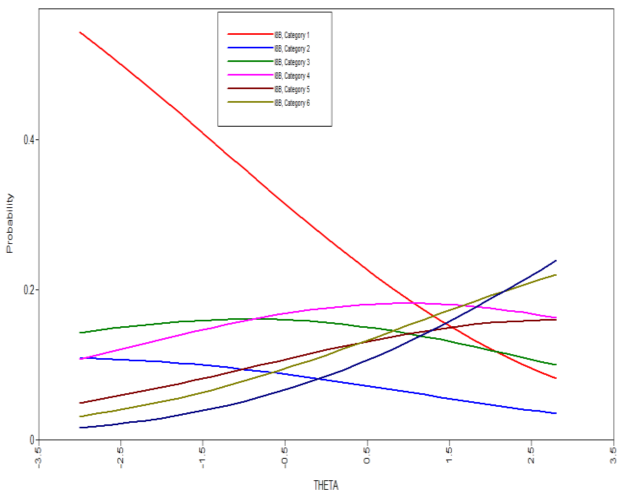


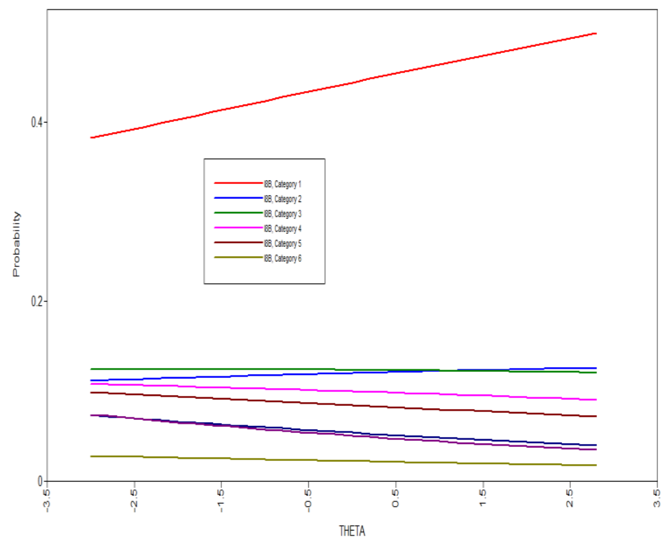


The diagram above shows the Item characteristic curve (ICC) for item i8a for rural and urban students respectively.To answer research question 5, the ICC above were observed to make comparison.it was seen that the two ICCs differ from each other so i8a is not invariant.

Rural i8b

Urban i8b





The diagram above shows the Item characteristic curve (ICC) for item i8b for rural and urban students respectively. To answer research question 5, the ICC above were observed to make comparison.it was seen that the two ICCs differ from each other so i8b is not invariant.