COMPARISON OF SECONDARY SCHOOL STUDENTS' ABILITY IN MATHEMATICS CONSTRUCTED-RESPONSE ITEMS UNDER CLASSICAL TEST AND ITEM RESPONSE MEASUREMENT THEORIES IN THE IBADAN METROPOLIS, NIGERIA

BY

BABATUNDE KASIM OLADELE

MATRIC NO: 147568

B. ED (EKITI), M. ED (IBADAN)

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR A THESIS IN THE INTERNATIONAL CENTRE FOR EDUCATIONAL EVALUATION (ICEE), SUBMITTED TO THE INSTITUTE OF EDUCATION AWARD DEGREE OF DOCTOR OF PHILOSOPHY OF THE UNIVERSITY OF IBADAN, IBADAN, NIGERIA

SEPTEMBER, 2021

CERTIFICATION

I certify that this thesis was carried out by **Babatunde Kasim OLADELE** at the International Center for Educational Evaluation (ICEE), Institute of Education, University of Ibadan, Ibadan.

Supervisor

Professor Benson Adesina Adegoke
B. Ed, M. Ed., Ph.D. (Ibadan)
Institute of Education
University of Ibadan

DEDICATION

This thesis is dedicated to the Lord Jehovah, my dearest wife; Adenike Saudat Oladele and my wonderful children; Olayinka, Iyanuoluwa and Itunuoluwa for their steadfast love and unflinching support throughout the programme.

ACKNOWLEDGEMENTS

Honour and glory to the Almighty God for this great thing he has done for me. I blessed Jehovah, the creator of heaven and earth for his grace towards the completion of this apex programme on a note of success and good tidings. My indebtedness of gratefulness goes to the following persons for their immerse assistances towards making this programme a huge achievement. My deep appreciation goes to my able, dynamic, God-fearing, energetic and dedicated supervisor, Professor B. A. Adegoke, for his cordiality, concern, guidance, constructive criticism, encouragement, kind advice and for allowing me to tap from his wealth of academic experience during this research project. I pray that God will continue to enlarge your coast and expand your frontier of knowledge in all ramifications. I am highly indebted to you sir.

Also, I express my profound gratitude to all my lecturers starting from, Prof. J. G. Adewale, Emeritus Professor PAI. Obanya, Prof. T. W. Yoloye, Prof. C. O. Onocha, Prof. P. N. Okpala, Prof. E.Adenike Emeke, Prof. Folajogun.V Falaye, Prof. O.A Onuka , Prof. J.A Adegbile, Prof. Euginea A Okwilagwe, Prof. Monica N. Odinko, Prof Modupe M Osokoya, Dr F. Ibode, Dr J. O. Adeleke, Dr Serifat Akorede, Dr Ikmat O. Junaid, Dr J.A. Abijo, Dr E.O Babatunde, Dr M. A. Metibemu, Dr Olutayo T. Omole my respected aunty who serves as my special research assistant and my dear colleague O. M. Desmennu for their impactful knowledge, discipline, dedication and support shown during the programme, your imprints in my life remains indelible. I equally thanked all the non-teaching staff of the Institute of Education, especially Mrs Abimbola P. Oni.

I particularly thank my caring wife, Adenike Oladele and my loving children; Olayinka Oladele, IyanuOluwa Oladele and ItunuOluwa Oladele for their unflinching support, prayers, endurance, perseverance and understanding throughout the programme. Also, I equally express my gratitude to my parents Mr Kasali Oladele (Blessed memory) and Mrs Oluyomi Oladele who objectively helped me in laying a solid education foundation and made my choice to become an academic a dream come true with prayers. To all my siblings Adebambo, Bukola, Bidemi, Oluwaseun, and Busayo the daughter of my beloved late Sister Bolanle, I appreciate you for all your prayers and immense show of love. In addition, my appreciation goes to my father in law Alhaji Bosun Adeniran (Blessed memory) and mother in law, Mrs Adeniran. I will not forget to thank my sis-in-laws and family, The Ibikunles, Omotoshos, Sanusi's and the rest of my immediate and extended family for their prayers and immense support.

It is also important for me to acknowledge and appreciate the entire congregation of Celestial Church of Christ, Oluwaseyi Parish, Idi Ori, Sasa, Ibadan. My special appreciation goes to the following PhD colleagues who have in one way or the other supported me in this programme. I will like to mention a few of them; Astua Godwin, Nkem Adeyemi-Adewoyin, Azeez Fatai, Gideon Oloyede, Kunle Ayanwale, Paul Adegoke and the others. This acknowledgement will not be complete if I did not recognise all members of staff of the Deputy Vice-Chancellor (Administration) Office, University of Ibadan, whom I worked with during the commencement of this programme. I appreciate Professors Arionla Sanya, Aiyelari and Adebowale, the former and present Deputy Vice-Chancellors (Administration) for their support and advice. I equally appreciate the Oyo State Government, the School Principals, all the Secondary Schools Head of Departments of Mathematics and the students who participated in this study. It is my prayer that the Almighty God shall reward you all richly.

ABSTRACT

Test items of public examinations are expected to be valid and reliable when good frameworks are used. Besides, the scoring of the items must be guided by an objective theoretical framework. Studies have shown that Mathematics Constructed-Response Test Items (MCRTI) of the West African Examinations Council (WAEC) and National Examinations Council (NECO) are scored using Classical Test Theories (CTT) framework which have been adjudged subjective. This study was, therefore, designed to score and compare students' ability in WAEC and NECO MCRTI under CTT and Item Response Theory (IRT). The dimensionality, equivalence of scores and differential item functioning (between males and females) of WAEC and NECO MCRTI under CTT and IRT were also examined.

The study was anchored to the Classical Test and Item Response Measurement Theories, while descriptive survey design was adopted. Counterbalance procedure was employed in the tests administration. One educational zone out of the two in Ibadan was sampled. All the five Local Government Areas (LGAs) in the educational zone were used. Twenty-four co-educational public Senior Secondary Schools (SSS) were randomly selected from the LGAs using proportionate to size sampling technique. Two intact classes of Senior Secondary School 3 (SSS3) in each school were used. In all, 1151 (565 males, 586 females) SSS3 students were sampled. The WAEC and NECO MCRTI (2013-2015) were used for data collection. The reliability coefficient established were WAEC MCRTI ($r = 0.72$) and NECO MCRTI ($r = 0.71$). Data were subjected to Exploratory Factor Analysis, Parallel Analysis (CTT models), IRT-Generalized Partial Credit Model (GPCM) and Graded Response Model (GRM), and correlated samples t-test at 0.05 significance level.

The two MCRTI were multi-dimensional. Under CTT, WAEC MCRTI had four dimensions, while NECO MCRTI had three dimensions. Under IRT, WAEC MCRTI and NECO MCRTI had three dimensions each. The WAEC MCRTI (GPCM=0.24, GRM=-3.16) were easier than NECO MCRTI (GPCM=2.10, GRM=4.95). Students' mean score in WAEC MCRTI under CTT was lower ($\bar{x}$=35.88, SD=10.02) than under IRT (GPCM) ($\bar{x}$=41.70, SD=7.0). The mean score in NECO MCRTI under CTT was lower ($\bar{x}$=33.49, SD= 12.39) than under IRT (GPCM) ($\bar{x}$=41.67, SD=6.98). The mean differences were significant $t_{(1150)} = 34.83$ (WAEC) and $t_{(1150)} = 33.32$ (NECO). Students' mean score in WAEC MCRTI under CTT was lower ($\bar{x}$ = 35.88, SD = 10.02) than under IRT (GRM) ($\bar{x}$=41.70, SD=7.04). The mean score in NECO MCRTI under CTT was lower ($\bar{x}$=33.49, SD=12.39) than under IRT (GRM) ($\bar{x}$=41.67, SD=7.01). The mean differences were significant $t_{(1150)} = 34.86$ (WAEC) and $t_{(1150)} = 35.04$ (NECO). The adjusted scores under CTT and IRT models were equal. Three items out of 15 WAEC MCRTI exhibited DIF under CTT, while 14 exhibited DIF under IRT in favour of males. None of the NECO MCRTI items exhibited DIF under CTT, while nine exhibited DIF under IRT in favour of males.

Item Response Theory models were more effective than Classical Test Theory in scoring constructed-response tests, equating and detecting differential item functioning. Public examining bodies should score constructed-response test items using Item Response Theory models.

Keywords: WAEC and NECO constructed-response items, CTT and IRT scoring and equating, Items dimensionality, Differential item functioning

Word count: 495

TABLE OF CONTENTS

CHAPTER THREE: METHODOLOGY

LIST OF TABLES

TABLE **PAGE**

LIST OF FIGURES

LIST OF APPENDIXES

CHAPTER ONE

INTRODUCTION

1.1 Background to the Problem

One of the core subjects at the senior secondary school level of education in Nigeria is Mathematics. The Nigerian Educational Research and Development Council (NERDC, 2013) structured the Mathematics curriculum at this level of education around the following four themes; algebraic processes, number and numeration, geometry and statistics. These concepts affect people's way of being in one way or another. For instance, geometry helps learners to appreciate the shapes of objects and it is very useful in architectural designs; number and numeration are used in the process of counting, additions, subtraction, multiplication and other arithmetic processes, algebraic processes gives a sense of logical reasoning and precise thinking towards word problem solving, while statistics helps human being in developing abilities in ranking and classifying of an object, entities and phenomena.

The significance of mathematics to national development is great and this probably explains why a pass at a minimum of credit (C6) in senior secondary school certificate examination (SSSCE) is a prerequisite for all candidates who wish to study all courses in core sciences, engineering, medicine, nursing and agriculture in Nigeria tertiary institutions (Joint Admissions and Matriculation Board, 2007). The results of candidates in Mathematics in SSSCE are used to gauge their ability to cope with courses in higher institutions of learning that require a fair knowledge of Mathematics such as engineering, medicine, mathematics education, social sciences and the core sciences.

Regardless of the significance of Mathematics to the socio-economic growth of human beings as well as being a prerequisite for admission into science and science-based courses in higher institutions, reports from (Ariyo, 2017; Omotayo, 2016) show that students' level of performance in both teacher-made and standardized tests is not quite impressive.

Data from these public examination organisations saddled with the conduct of SSSCE show that candidates' performance in Mathematics in SSSCE is slightly above average level

Table 1.1: Statistics of Students Performance in WAEC and NECO Mathematics (2007 – 2018)

	WAEC	**NECO**	**WAEC**	**NECO**	**WAEC**	**NECO**
Year	**(A1-C6) % Higher Passes**	**(A1-C6) % Higher Passes**	**(D7-E8) % Passes**	**(D7-E8) % Passes**	**(F9) % Failure**	**(F9) % Failure**
2007	46.75	49.50	26.72	42.60	24.24	7.90
2008	52.27	47.31	23.83	36.69	17.23	16.00
2009	47.04	49.20	25.56	26.61	23.41	24.19
2010	41.95	42.33	26.85	32.73	27.20	24.94
2011	40.40	35.86	31.50	33.32	27.90	30.82
2012	46.64	48.51	28.72	25.89	24.54	25.60
2013	44.24	46.25	26.53	47.28	29.03	6.47
2014	40.20	52.29	30.53	38.42	29.17	9.29
2015	57.02	49.50	26.91	26.30	16.06	24.20
2016	70.23	80.16	19.37	15.04	12.84	4.80
2017	59.22	70.85	25.59	24.28	10.41	4.87
2018	76.84	71.48	24.35	19.22	18.40	9.30

Source: The National Examinations Council, Minna, Niger State and West African Examinations Council, Yaba, Lagos (2018)

Table 1.1 reveals that the proportion of students who made at least credit pass in Mathematics, in a review of twelve years (2007 to 2018) is averagely below 55% of the total number of candidates who sat for the examinations. The implication of this is that about 50% of Nigerian students who might have wished to read courses in Mathematics education, sciences and science-based courses in higher institutions could not get the opportunity to secure admission based on not having a minimum of C6 in Mathematics. However, there were slightly appreciable improvements in the achievement of candidates in WAEC Mathematics in 2015, 2016, 2017 and 2018 and WAEC Mathematics in 2016, 2017 and 2018. Also, students' performance in NECO Mathematics has not been remarkable for almost a decade (2007-2015). Although there was an impressive performance in 2016, it was not sustained in 2017; the level of candidates' performance nosedived in 2017. The ultimate is yet to be attained. The ultimate is for students and schools to achieve 100% success in SSCE. To explain why candidates' performance was poor, the WAEC and NECO Chief Examiners' reports Mathematics in 2013, 2014, and 2015 stated that candidates usually failed mostly in constructed-response (CR) test items (or theory test items).

Specifically, the reports showed that some candidates demonstrated a lack of computation abilities, could not apply formulae to solve problems and on most occasions, some candidates presented their responses to CR in haphazard manners. Past studies (Akinoso, 2011; Moyinoluwa, 2015, Omotayo, 2016; Ariyo, 2017; Adegbuyi, 2018) have tried to isolate the causes of students' failure in Mathematics in public examinations. Factors such as teacher teaching method, type of testing (Adegoke, 2016), lack of consideration for lower-achieving students, lack of students' engagement, especially the use of lecture method (Adegoke, 2013), students' poor attitude to Mathematics and anxiety (Adeleke, 2007; Goolsby, 2013), students' low level of reading habits (Idigo, 2010), have been known in the literature as causes of some students' failure in Mathematics. Some solutions have been proffered; among which is the adoption of the student-centred method of teaching Mathematics (Adegoke, 2003) and the use of computer-assisted instruction (Ogunyomi, 2010; Adegoke, 2016). Despite these, the performance of students remains slightly above average and the ultimate solution is yet to be found.

An aspect that has not been adequately researched is the assessment procedures being adopted by the NECO and WAEC. Part of these factors that explain the poor performance might be established on the assessment processes used by the public examining bodies in test construction, administration, marking methods and measurement simulations (scoring models). The most key concern to verify is the fitness of scoring frameworks being used by these external examinations organisations. This concern can be linked to the scope of psychometrics (model and practice of psychosomatic assessment). One of the reasons why student's level of performance over the years is slightly above average in Mathematics could be as a consequence of assessment practices employed by the public examining bodies. It is therefore necessary that searchlight be beamed on the testing procedures of the public examining bodies.

There are two common types of questions usually used by WAEC and NECO for the assessment of students' performance in SSSCE Mathematics are multiple-choice (selected-response) and essay questions (constructed-response). For selected-response questions, the WAEC use 50 items with four response modes A, B, C and D each, whereas NECO uses 60 items with the items positioned under five responses modes of A, B, C, D and E. For the constructed-response questions named Paper two, the WAEC practice 13 items while NECO practices 12 items. Appendix IA and IB show the format of CR test items of WAEC and NECO respectively.

Constructed-response items are questions with implicit open-ended or different forms of items that entail students to supply responses or answers, the answers are later read and marked by the subject examiner. The requirement of constructed-response items from the examinee is as different as the estimated abilities. It is used to measure a wide variety of abilities. For instance, the examinee can be requested to answer certain items using a mathematical equation and as well provide an explanation, support response with reasons or numerical evidence or display data. Some cognitive processes like simplifying, apply, find, solve and analyse are better assessed with CR items because such abilities seem difficult to estimate successfully with multiple-choice items. The CR items are used to create probably most likely higher construct validity and are assumed to be suitable for determining assured abilities which entail dissimilar stages of the psychological feature procedure. Conversely, there are difficult tasks associated with constructed-response items.

A major problem associated with CR test items is the inherent difficulty in scoring the items accurately and reliably. Moreover, the time involves in scoring is overwhelming and also expensive. The content-related validity is frequently conceded or delimited, specifically for huge content domains. Scoring in constructed-response is fundamentally subjective and involves consideration for reducing the adverse influence of bias on the scoring validity (Livingston and Ruppy, 2004). One of such issues is the model for scoring students' responses. Another issue is the test theory upon which the examiner wishes to base the scoring. In testing, the two major test theories are; the Classical Test Theory and Item Response Theory (Hambleton and Jones 1993: Oladele and Adegoke, 2020).

The basis of measurement theory has been rooted in Classical Test Theory for over ten years. CTT has permitted the establishment of various outstanding measurements and ample scales in the professional assessment practices in Africa. This is due to the acceptance of interpretation which can be practically used for students' achievement and aptitude test outcomes. Classical test theory, also identified as true-score theory, accepts that each person has a true score, T that will be attained if there were no errors in measurement (Cappelleri, Lundry and Hays, 2014). It is stipulated that an individual's true mark is described as the expected score over an infinite number of independent administrations of the scale. Scale users on no occasion observe a person's true mark, but only an observed score, X. It is presumed that an observed mark (X) = true mark (T) + some error (E). In CTT, the total raw score of the student's in a test is the score obtained on the test items. The CTT and its models have been investigated and used constantly and effectively over six decades, and various testing platforms today remain resolutely ingrained in classical test theory and approaches. At present, WAEC and NECO still use CTT in scoring. Regardless of the acceptance of classical test analysis as an essential measure of standardised test and measurement tools, some deficiencies are encountered (Hambleton and Jones, 1993; Ojerinde, 2013).

Hambleton and Jones (1993) stated the major limits of CTT to include: (a) the person statistics (that is, observed score) is item sample-dependent, and (b) item statistics (that is, item difficulty and item discrimination) are students sample-dependent. Consequently, the assessments of CTT are not generalisable through individuals and this is one of its weaknesses. The Item response theory was developed by psychometricians to cater for the deficiency of CTT. Both test theories use a different procedure to allocate numbers to features of students using different procedures and on both person answers and on the properties of items that were administered.

Item Response Theory (IRT) is a broad statistical theory around the students, item and test performance and how performance correlates to the abilities that are measured by the items in the test. IRT is grounded on the IRT which integrates assessment rules around a student, item and test performance, in what way performance correlates to understanding as assessed by the item on a test (Schumacker, Si and Mount, 2003). According to Ayanwale (2017), IRT is primarily interested in whether a student responses to a question correctly and not in the raw scores in a test which are stated as the item-pattern scoring procedure (Ayanwale, 2017). The measurement outcomes under IRT refer to a measure that placed both students and items on a common scale. More importantly, a measurement scale must be available. Meanwhile, the true picture of the latent variable cannot be established because the task involved in scaling is challenging.

However, two suppositions must be set to identifying IRT models. The first is the dimensional arrangement of the data, and the other relays the mathematical method of the item characteristic curve (symbolised by ICC). Responses to items could be discrete or continuous and also dichotomously or polytomous estimated; the item group scores could be ordered or unordered. Ostini and Nering (2006), stated that there are various traits underlying test performance; and there are various means (that is, models) in which the relationship between item responses and the underlying abilities can be stated. Inside the overall IRT context, several frameworks have been established. The prominent ones among them include rating scale, binomial, poison, facet, dichotomous, multinominal logit or polytomous. These frameworks are used for the scoring of discrete or continuous, dichotomous and polytomous item responses.

Two methods can be employed in measuring the constructed-response test items; these are the point score marking and holistic marking. In point score marking, the students are given a particular score for each point correctly stated in line with the prepared marking guide, while in holistic marking, the point(s) is given by the general impression of the examiner about the points or responses that the students give. In both the WAEC and NECO public examinations, the point score marking system is usually adopted in line with the marking guide. The usual practice is to add the scores which the candidates get in each item. This gives the candidate's overall performance in the essay test. This method falls within the purview of classical test theory (CTT). Under the IRT framework, constructed-response test items are scored polytomous.

Polytomous and dichotomous items are categorical items but polytomous items answer categories are more than one. This involves ordered and categorical data by the number of categories into which the data can be positioned and as well distinct by verges that differentiate the categories (Adegoke. 2013). Reasonably, there is constantly one less boundary than there are categories. An example is in scoring a dichotomous item that entails only one boundary to distinct the two probable answer categories (correct or incorrect). This one boundary is called difficulty level. For dichotomous items, only one difficulty parameter is assessed for each item. In polytomous test items like the Likert 4-point (with responses: Agree; Strongly Agree; Disagree and Strongly Disagree) item entails three boundaries to distinct the four probable categories. Therefore, in a four-point Likert scale, three difficulty levels (b1, b2, and b3) are estimated. In an essay item that is recorded over five; likely categories include 0, 1, 2, 3, 4, and 5. That is, a candidate can score either zero, or 1, or 2, or 3 or 4 or 5. This gives six categories. Conversely, when the polytomous framework is used, there are five categories, that is, the six categories minus one. Therefore, five difficulty parameters (b1, b2, b3, b4 and b5) are estimated.

Several measurement models have been established for polytomous items. These include the Graded Response Model (GRM), Nominal Response Model (NRM) Partial Credit Model (PCM) and Generalised Partial Credit Model (GPCM). In this study, emphasis was placed on GRM and GPCM only because these models can score students ability and estimate parameters of items effectively. In both the GPCM and GRM, the difficulty and discrimination indices are included unlike in the PCM where only the difficulty parameter is indicated. In both the GPCM and GRM, the slope parameter (discrimination index) remains continuous through the categories in each question but could be dissimilar among. questions

One of the major issues to be considered in the use of the item response theory frameworks is the dimensionality of the test items. The dimensionality of a test is the amount of latent variables that are dignified by the test. All unidimensional tests estimate mainly single latent variable and the multidimensional tests estimate more than one latent variable. Usually, selected-response test items are unidimensional while constructed-response test items are inherently multidimensional. This is because, in answering a Mathematics CR test item the examinee is expected to demonstrate some traits, like computational skills, the skill to decode problems into mathematical language and demonstrate logical or abstract thinking. As stated in both WAEC and NECO syllabi, Mathematics Examination aims to examine candidates: (i) mathematical competency and computational skills (ii) understanding of mathematical concepts and their relationships to the acquisition of entrepreneurial skills for everyday living in the global world (iii) ability to translate problems into mathematical language and solve them using appropriate methods (iv) ability to be accurate to a degree relevant to the problem at hand and (v) logical, abstract and precise thinking. In this study, one of the foci was to determine the number of dimensions of the CR test items of WAEC and NECO.

Studies (Adegoke, 2016; Livingston and Ruppy, 2004) have shown that, in addition to the use of CTT, many public examining bodies in the United Kingdom, United States of America, Germany and China and are using polytomous models such as the graded response and generalized partial credit models to estimate candidates' scores in CR test items. Differences in students' scores have been observed when CTT and IRT models were compared. The IRT model appears to be fairer to candidates when scores emanating from CTT and IRT models are compared. This has, therefore, prompted many of these public examining bodies to depend majorly on the IRT model. In Nigeria and indeed in most countries in Africa, this is not the case. What obtains, for instance, in Nigeria is that the two major public examining bodies conducting SSSCE (WAEC and NECO) mainly adopt the CTT model in scoring candidates CR test items. In Africa, there is a dearth of literature on this aspect of the assessment. Thus, this study was designed to compare the scores of students in Mathematics within the frameworks of CTT and IRT.

In Nigeria, there are arguments for and against having two parallel examining bodies for the conduct of the same SSSCE. More importantly, according to studies (Adegoke and Obot, 2020; ; Obot, 2019; Anigbo, 2018 Ojerinde 2013), there are arguments about the equivalence or otherwise of the senior secondary school certificates being awarded by the two major public examining bodies. This is because in most subjects, including mathematics, the distribution of test items of the two public examining bodies are not equal in terms of number and spread through the numerous levels of the cognitive domain. Hence, the study was therefore designed to compare students' scores in Mathematics CR test items of WAEC and NECO under the frameworks of CTT and IRT. The scores of the students emanating from the use of CTT and IRT frameworks are then equated to further determine the extent of equivalence of scores emanating from the two examining bodies. Also, the differential functioning of the CR test items, under the two models, of WAEC and NECO were compared to determine which of the two examining bodies is fairer or otherwise to both male and female candidates.

Test equating conventionally relates to the statistical procedure of defining similar scores along with diverse methods of an examination. This could be processed with either a classical test model or an item response model. Under IRT, equating refers to the procedure of engaging scores from two or more equal test procedures onto a common score measure. The result of scores from two separate test methods can be equated unswervingly, or preserved as though they came from the same test procedure. When the tests are not equivalent, the general process is referred to as linking. This is the process of equating the units and origins of two scales on which the abilities of students have been obtained from results on separate tests. The procedure is equivalent to equating degrees Fahrenheit with degrees Celsius by changing measurements from one scale to the other. The purpose of comparing scores is a by-product of equating those outcomes from equating the scales acquired from test results.

The Differential Item Functioning (DIF) is broadly used for detecting bias and to state if a test is fair or not. Ogbebor and Onuka (2013) described test fairness as any test given to a set of examinees with an equal chance to prove their acquired knowledge. When a test item has the element of DIF, it shows that the items are not fair to some of the examinees or a subgroup. The subgroup can be in terms of the location of examinees, gender, ethnicity or social and economic background of the learners. This study, therefore, considered the subgroup based on the gender of the students. This is because of the generally held view that mathematical issues are more inclined to males. Most illustrations in Mathematics textbooks at the senior secondary school level reflect more of activities of the men than that of the women.

1.2 . Statement of the Problem

Constructed-response (CR) test items seem difficult to score objectively, accurately and reliably because of their required contents compared to the selected-response items. The scoring of CR items is naturally subjective and therefore, needs attention on several problems to lessen the undesirable influence of subjectivity on the scoring validity. One of such problems is the framework for scoring students' responses. Another issue is the test theory upon which the examiner desires to employ to score the items.

In West Africa, especially in Nigeria, there is a paucity of literature on the effectiveness of Item Response Theory (IRT) in the development and scoring of CR test items. A literature search revealed that in the developed countries like the USA, Britain and Japan, public examining bodies are adopting the IRT framework in scoring candidates' responses to CR test items. Developments in test theory have revealed shortfalls in the Classical Test Theory (CTT) framework. The responsiveness of the Joint Admissions and Matriculation Board to the inadequacies of CTT and the probable benefits presented by IRT has driven it to adopt the IRT framework in the testing processes. This is limited to selected-response test items.

From reports, the WAEC and NECO still employed the CTT framework in scoring constructed-response items. This may be one of the reasons why student's performance in WAEC and NECO Mathematics has not improved appreciably. This is because studies have shown that CTT tends to underestimate the score of students in CR test items. It is on this basis that this study examined how scoring under IRT and CTT affects the abilities of students in Mathematics CR test items. The CTT model and two ordinal polytomous models of the IRT framework (generalised partial credit model and graded response model) were used. More importantly, the estimated students' abilities through the models were equated to determine their equivalence. Also, the differential functioning of the items with regards to gender was examined under CTT and IRT frameworks.

1.3 **Research Questions**

The following questions were answered in this study:

1. How many dimensions do the 15-item of WAEC and NECO Mathematics constructed-response tests have under:
   (a) Classical Test Theory (CTT) framework?
   (b) Item Response Theory (IRT) framework:
      i.Generalised Partial Credit Model (GPCM)?
      ii. Graded Response Model (GRM)?
2. What are the parameter estimates of 15-item of WAEC and NECO Mathematics constructed-responses tests under the framework of?
   (a) Item Response Theory (IRT):
      i.Generalised Partial Credit Model (GPCM)?
      ii. Graded Response Model (GRM)?
3. Is there any statistically significant difference between the ability scale scores of students in 15-item WAEC and NECO Mathematics constructed-response tests along?
   (a) Classical Test Theory (CTT) framework?
   (b) Item Response Theory (IRT) framework:
      i.Generalised Partial Credit Model (GPCM)?
      ii. Graded Response Model (GRM)?
4. How equivalent are the scores of the students in WAEC and NECO Mathematics constructed-response tests under the framework of?
   (a) Classical Test Theory (CTT)?
   (b) Item Response Theory (IRT)?

5. Which of the 15-item of WAEC and NECO Mathematics constructed-response test function differentially between male and female students using the framework of:
    (a) Classical Test Theory (CTT)?
    (b) Item Response Theory (IRT)?

1.4. Scope of the Study

This study focused on the entire senior secondary student's III (SSS3) in Ibadan Metropolis. The emphasis was on constructed-response Mathematics test items as being presently constituted in both the WAEC and NECO senior secondary school certificated examinations. Also. assessments of item and person parameters of ordinal polytomous IRT models (The graded response and generalised partial credit models) were equated. Conversely, ability estimates of students, given different combinations of the framework of polytomous IRT models and CTT models, were also estimated.

1.5. Significance of the Study

The study outcomes offered empirical suggestions to education stakeholders such as the examiners, examinees, evaluators, public examining bodies, educational planners, Government, ministries of education and Mathematics teacher's on the need to incorporate both classical test and item response frameworks in assessing the psychometric properties of constructed-response items and score test items.

Firstly, the findings of this study would enable Mathematics teachers in secondary schools to examine students, based on the content of the curriculum. It would enable teachers in Mathematics education, public examining bodies and examiners have empirical information on the effective use of test models in estimating item parameters, establishing dimensionality, detecting differential item functioning and equating of test items. Secondly, the results of this study would help students to have confidence in that the test being used to measure the ability are reliable and valid.. For the public examining bodies, the finding has serve as an eye opener to the possible application of item response theory (IRT) framework in estimating item parameters of tests used for assessing candidate's ability in Mathematics.

Similarly, it would also provide clues on comparability of both the CTT and IRT parameters. The findings have communicated the effectiveness of scoring and estimation methods of test measurement frameworks on item parameter estimates to Mathematics teachers, examiners, and public examining bodies. More importantly, the findings revealed which of the two frameworks produces better assessment of examinees' scores.

Nevertheless, results from this study have added to the literature on constructed-response tests and how to use polytomous models in the scoring Mathematics test items. Moreover, this study provided empirical evidence on the equivalence or otherwise of WAEC and NECO scoring and graded system. It gives an insight into the need for public examinations bodies and test experts to conduct differential item functioning with test models to identify and remove test items that functions differentially between male and female examinees or other subgroups. More so, the results would be of benefit to the Government and ministry of education in terms of establishing an informed policy concerning the quality of test items being administered by public examining bodies in secondary schools.

1.6 Conceptual Definition of Terms

For this study the following terms were defined thus:

Classical Test Theory: This refers to a test theory that described the observed mark on a test as the totality of the true mark and measurement error.

Constructed-response Item: This refers to a type of open-ended *essay* question that demonstrates cognitive reasoning and knowledge.

Differential Item Functioning: This is a method used to detect if test items favoured a subgroup of examinees than the other.

Item analysis: This refers to the process of estimating the difficulty index and discriminating power of test items.

Item parameters: This refers to the measure of difficulty, discrimination and guessing. These are denoted by a, b, c in the three-parameter model of IRT.

Item Response Theory: This refers to a psychometric theory and family of associated mathematical models that relate latent trait(s) of interest to the probability of responses to items on the assessment.

Model fit: This refers to the extent to which the estimated trait of an examinee agrees with the test items.

Polytomous Models: This refers to frameworks that are used to score items with more than two options or categories.

Public Examining Bodies: This refers to bodies that are saddled with the responsibility of conducting certification examinations for students to transit from one level of education to another.

Students' Ability: Scores of the students in the constructed-response test that denoted by (***Θ) in*** IRT.

Test Equating: This refers to a numerical procedure use for the changing of scores of test forms and the test forms are used interchangeably.

1.7 Abbreviations

1PL - The one-parameter logistic (model)

2PL **-** The two-parameter logistic (model)

3PL **-** The three-parameter logistic (model)

CR - Constructed-response

CTT **-** Classical Test Theory

DIF - Differential Item Functioning

GPCM - Generalised Partial Model

GRM - Graded Response Model

IRF **-** Item Response Function

IRT **-** Item Response Theory

MIRT - Multidimensional Item Response Theory

NABTEB - National Business and Technical Examination Board

NECO - National Examinations Council

NRM **-** Nominal Response Model

PCM - Partial Credit Model

SR - Selected-Response

SSCE - Senior Secondary School Examination

TCC - Test Characteristic Curve

UIRT - Uni-dimensional Item Response Theory

WAEC - West African Examinations Council

WASSCE - West African Senior School Certificate Examination

Θ - Students ability parameter

CHAPTER TWO

LITERATURE REVIEW

This chapter presents the review of relevant literature as follows:

2.1. Theoretical Background

In the practice of educational and psychological measurement, the test theories and their related models are significant because they offer a framework for considering issues and addressing technical problems. The two prominent test theories are the Classical Test Theory (CTT) and Item Response Theory (IRT). This present study is predicated on these theories because of their relevance and contributions to the field of educational testing and measurement over the years. Classical Test Theory is concerned with test scores of students that capture the core of three principles. These are test scores (often called the observed score), true score, and an error score. The CTT assumes a simple linear model relating the observable test score (x) to the totality of two unobservable (or usually refers to latent variables), true score (T) and error (E), that is, X= T+E (Hambleton and Jones, 1993).

Item Response Theory (IRT) is a model-based measurement in which trait level estimations depend on both individual responses and on the features of items that were tested. The procedures of measurement in IRT give better strength, efficiency, flexibility and reliability in trait measurement than the classical test framework which has been in practice for more than the 20th Century. The underlying principle employed in IRT models for testing is that individual and item parameters can be fully separated and this is brought to bear on assessing students traits and test features with greater exactness and flexibility. IRT now triggers several major tests. Apart from educational testing to measure student’s ability, IRT has also been applied to personality trait measurements, as well as attitude measurements and behavioural ratings. IRT principles are convoluted in both selecting the most fitting items for a student and equating across different subsets of items. The IRT serves as a substitute to CTT as a foundation for observing the association between item responses and the ability of a student being examined by the test or the scale (Hambleton and Swaminathan, 1985).

Several varied IRT models are obtainable for use in an extensive collection of education psychological specialisations. Although early IRT models emphasised dichotomous item formats extensions to other formats had its allowed submissions in various areas. IRT frameworks have been developed for rating scales, partial credit scoring and multiple categories scoring of items. In IRT, the ability of a student and the likelihood of responding to a question rightly centered on the form of responses to the question that establish a test is modelled. Item Response Theory is used to evaluate the parameters of an item independent of the characteristics of examinees to which it is administered and other items that institute the test. The three noticeable equations termed that is currently used for predictions are 1parameter logistic, 2parameter logistic and 3parameter logistic models.

Although it is one parameter that is usually attributed to the trait level of a person, the item is frequently characterised by the 3PL parameter logistic model. The separate ability level is repeatedly selected through theta, which denotes the number of characteristics, attributes or abilities that an individual has. The three parameters associated with the item are discrimination power (a), the difficulty parameter (b), and the guessing parameter (c). In a cognitive task, the a-parameter shows the degree to which students' answer to an item differs with, or transmits to the ability level (Nenty, 2000). The b-parameter is the measure of trait characteristics in an item. This denotes the cognitive opposition of the item. This can also be described as the measure of attributes below the assessment required to overwhelm the assignment. The c-parameter is the likelihood that an individual deficient in the trait responds to the item correctly. Hambleton and Swaminathan (1985) summarized the features of Item models as follows: first, an IRT model must stipulate the association between the observed answer and underlying unobservable construct. Second, the model must offer a way to measure scores on the ability. Third, the student's scores will be the foundation for the estimation of the underlying construct. Lastly, and IRT model assumes that the performance of students can be entirely projected or described by one or more skills.

IRT comprises a set of models that describe the interactions between a person and the test items. Persons may have different traits and instruments may be planned to measure more than one trait. There are two main forms of IRT, These are the unidimensional and multidimensional IRT. The Unidimensional IRT (UIRT) assumed that all items in an instrument stake a mutual feature while the multidimensional IRT (MIRT) items do not share a mutual feature. Most of the study on IRT has focused on multidimensional frameworks that are proficient in defining many inherent people's abilities and numerous tests dimensions. Some of the notable scholars in this field include Reckase (1985), DeMars (2002) and Chalmers (2012).

2.1.1 Classical Test Theory and Its Related Models

This is one of the models developed to evaluate the scores of examinees. It refers to a concept around students marks that reveals the nucleus of three conceptions. These are test scores (perceived score), true score, and an error score. The CTT assumes a simple linear framework involving the noticeable test score (x) to the amount of two unobservable (latent variables), true score (T) and error (E), that is, X= T+E (Hambleton and Jones, 1993). Nenty (2000) also provided the central formula of CTT to be Xo=X+Xe. This means that the assessment of any behavioural feature is the true score (X), while that which results from the measurement is the observed score (Xo) and (Xe) is the error of measurement.

The assumptions in the classical test model are that (a) true score and error scores are uncorrelated, (b) the average error score in the population of students is zero, and (c) error scores equivalent tests are uncorrelated. In this formulation, where error scores are defined, true score is the difference between test score and error score. The classical test model does not entreat a multifaceted theoretic model to correlate a student's trait to succeed a specific question (Nenty, 2000). As an alternative, CTT mutually examines a group of students and observes the achievement rate on an item of the students (supposing estimated in dichotomous). This accomplishment degree of a certain group of students on a question, often identified as the p-value of the question, is employed as the index for the item difficulty (essentially, it is an opposite pointer of item difficulty, with a higher value signifying an easier item). The ability of an item to discriminate between higher ability students and lower ability students is identified as item discrimination, which is regularly stated statistically as the Pearson product-moment correlation coefficient between the scores on the item (for instance, 0 and 1 on an item scored right-wrong) and the scores on the aggregate test.

Once an item is dichotomously scored, this estimation is frequently computed as a point-biserial correlation coefficient. The main drawback of CTT can be summarised as a circular dependency: (a) The individual statistic (observed score) is (item) sample dependent, and (b) the item statistics (item difficulty and item discrimination) are (students) sample dependent. This circular dependence postures some theoretical difficulties in CTT's application in some measurement situations (like test equating and computerised adaptive testing). Despite the theoretical weakness of CTT in terms of its round dependence of item and person statistics, measurement experts have worked out practical solutions within the framework of CTT for some otherwise difficult measurement problems Nenty (2000). For instance, test equating can be accomplished empirically within the CTT framework (equipercentile equating). Likewise, experiential methods have been planned to achieve item-invariant measurement (Thurstone absolute scaling). It is fair to say that, to a pronounced degree, even though certain concerns may not have been addressed hypothetically within the CTT outline, many have been solved through ad hoc empirical processes (Fan 1998).

2.1.2 Item Response Theory and Its Related Models

The IRT models offer rich statistical tools aimed at exploring the achievement test and cognitive scale data. Under the simple situation, these data encompassed a model of persons answering test or scale items which are in two categories. The concern could be defined in the assessment of items and personal characteristics. All the approaches were mostly established between 1960 and 1980. However, there was a brief historical review of IRT summarised by Bock (1997), but Thurstone during the 1920s sow the seeds for the IRT models. There was a book written by Lord and Novick (1968) in which the ideas of

IRT were presented and Birnbaum chapter was one of the seeds. Some of the texts and collection by researchers include Embretson and Reise (2000), Hambleton, Swaminathan, and Rogers (1991), (Lord, 1980; Hulin, Drasgow, and Parsons, 1983) Before IRT was established, CTT was used to assess a person's score on a test (Novick, 1968). Item response theory models overwhelmed some restrictions of CTT. Under CTT, the measurement focus is the test, while in item response theory, test items are measured.

Originally, IRT was established for dichotomous questions while polytomous models for ordinal and nominal questions was introduced by Samejima (1969) and Bock (1972). In addition, numerous taxonomy of ordinal and nominal item response models was presented by Thissen and Steinberg (1984) and briefly unfolding different methods showing the relationship between these models. Mellenbergh (1995) and, van der Linden and Hambleton (1997) in their study created another series of item response models for polytomous questions which covers some articles unfolding IRT models for polytomous questions. Specifically, the item response model states that a person's mark on the test is a unidimensional underlying ability variable, commonly represented as $\boldsymbol{\Theta}$**.** In cognitive measures, this underlying variable might be well categorised as rigorousness subject to what the measure is planned to estimate. Several instances of using the item response model with cognitive measures are in Thissen, Steinberg and Schaeffer (1984). This underlying variable is similar to a trait in a factor analysis framework for constant variables, and hence, item response and factor analysis frameworks are likewise considerably correlated. Since test and scale can be used to estimate more than one underlying factor, multidimensional item response models were also established with some developments.

Usually, a model of N persons answers to *ni* questions on one case, and yet the quantity of definite period that one case signifies could differ. Under certain conditions, examinees are measured many times, possibly with a similar form or a correlated form of questions. Some studies have defined submissions and improvements of item response modeling to such longitudinal conditions in psychology (Embretson and Reise, 2000). However, the representation is done considering *Yij* which is the dichotomous reaction of person i (i = 1, 2 . . . N persons) to question j (j = 1, 2 . . . ni items). know that examinee i is dignified on n items, thus, it is essentially assumed altogether that examinees are assessed on all items. Similarly, the symbolisation might infer that questions are nested within subjects. Typically, this is mutual for persons and questions to be intersected because all N persons are given similar n questions. The only exemption to this is in computerised adaptive assessment where the questions that a person takes are sampled from an enormous group of possible questions centered on the consecutive question answers.

2.1.3 Comparison of Classical Test Theory and Item Response Theory

Classical test theory has been the basis for measurement theory for over 90 years, yet it is confronted with the difficulties of non-correlation of true and error scores, group dependence item statistics (items difficulty and discrimination), assumption of equal error of measurement among all students (Akindele, 2004; Enu, 2014). This gave rise to the expansion of modern method that uses Item Response Theory. It is a set of models which, by relating the likelihood of a particular reaction by an individual with a specified ability level to the characteristics of question designed to elicit the level to which the person possesses that trait. Item response model attempts to model the association between a student's underlying traits and the likelihood of the student answering a particular item rightly. Nenty (2000) stated that IRT attempt to explain the process, estimate the parameters of concerns and predict the results of the task. This implies that an unobserved latent trait in a test enables such task amid separate students and the item to be possible during the assessment process; hence IRT is regarded as latent trait theory which focuses on test items.

However, the item response theory is proficient to evaluate the parameters of an item independent of the characteristics of both examinees to which it is exposed and other questions that constitute the test. The 1 parameter logistic, 2 parameters logistic and 3 parameter logistic models are presently being used to make predictions. Whereas only one parameter is attributed to a personality trait, the item is always characterised by the 3PL. The level of a personal attribute is commonly labeled as theta (θ), representing the sum of ability and level of ability possessed by a person (Baker, 2001).

The parameters related to the item are discrimination power (a), the difficulty parameter (b), and the guessing parameter (c). In a mental task, the a-parameter specifies the degree to which students' response to an item varies with, or relates to their trait level or ability (Nenty, 2000). The b-parameter is the amount of trait essential in an item. This represents the cognitive resistance of the item or task. In other words, this is the quantity of a trait in measurement objective essential to overwhelm the task or item. The c-parameter is the probability that a person completely lacking in the trait will overcome or answer the item correctly. The IRT, when compared to the CTT is considered as the suitable model in examining psychometric assessments of old and new methods. Though IRT has been studied for over five decades but CTT is presently being investigated and used always. Even today, several testing programmes still use CTT in the development and scoring of tests. This is because of the inherited benefits of classical test theory above item response theory. The examples include: Classical test theory defines the association between the true score and linearly observed score making CTT's frameworks simple to comprehend and employ by a lot of investigators and examiners.

Under the CTT, the focus is on overall scores and the number of right responses obtained by examinees. The aggregate score attained by each student is called the observed score and varied from the real mark by a shared error mark (Ojerinde, 2013). The process has produced some benefits and shortcomings. Some of the benefits of CTT are: the analyses involve smaller sample sizes, the mathematical procedures involved are easy to understand, the models in CTT are linear, the model parameter assessment is theoretically direct and entails minimum rules which make the models suitable and generally appropriate because of the entail stringent goodness of fit required for the analyses compared to IRT.

Nevertheless, some of the disadvantages of CTT include; the foundation of various CTT analyses are determined by items difficulty and reliability properties produced. The index is estimated through the test's percentage, (p), of an examinee that responds to the item rightly and the item-total correlation, (r). Though, these indices are not continuous as they depend on the number of students who wrote the test. These are not used to specify the character or value of a test. One more disadvantage is that the student's scores are tested reliant. This means that students may achieve lesser marks on a harder test and achieve higher scores on an easier, and therefore no real mark can be pulled out giving no room to the matching of items and ability level of the student. However, some benefits of IRT over CTT include; Under IRT, test features do not sample dependent and an individual's latent scores are not tested dependent if the designated models fit the data better. Therefore, marks that designate student's ability not depending on test difficulty. The marks may be lesser on harder questions and greater on simpler questions, nonetheless, the estimated score of the examinee remains the same at any time of assessment and the test. IRT also allow the estimation of the probability of a specific examinee choosing a group on a test question. All tests items that are not sample dependent enable the development of computer-adaptive examinations for detail precision when comparing or identifying test-takers.

Furthermore, in scale improvement or development, IRT can be suitably adopted because it is proficient in the estimation of normal errors and consequently offers statistics on the value of all items (Hambleton et al, 1991; Adegoke, 2013; Ojerinde, 2013). This helps in taking decisions on which items to be selected, excluded or included in a test or stated in an instrument. It also provides a guide on the selection of items based on their difficulty and discrimination indices, showing the ability to discriminate low and high latent trait groups. Item response could be used to assess the parameters of an item independent of the characteristics of both examinees.

Aside from the various benefits of IRT models, there are some disadvantages associated with its models. These include: first, the models are hard to understand and the parameter estimation procedures often include difficult numerical procedures. Second, the Latent attributes and item parameters are complex to understand in terms of graphics and numeric. It is only the one-parameter model that is simple to use than the other item response models. Though, the constraints enforced by assumptions of the model are easily managed sometimes through the advent of new studies on the models. Third, the design of the item response models is more difficult than the classical test models. Fourth, the processing and preparation of data for practice with item response model software are arduous and tasteful.

2.2 Conceptual Review

2.2.1 Public Examining Bodies in Nigeria

There are four public examining bodies in Nigeria. These are the West African Examinations Council (WAEC), National Examinations Council (NECO), National Business and Technical Examination Board and Joint Admissions and Matriculation Board. The examining bodies are statutorily charged to conduct public examinations while WAEC, NECO and NABTEB conduct examinations that lead to certification, JAMB conduct unified tertiary matriculation examinations and do not award certificates. The conduct of public examination (s) is a pivotal role of public examining bodies. Test administration represents the implementation phase of the whole procedure of public examining, where a plan drawn up during the pre-examination stage is operationalised. Examination bodies conduct certification examinations like school examinations, vocational examinations, professional examinations or selection of the type or form of test/examinations whether achievement, aptitude, mental ability and other psychological tests (FRN, 2013). These examinations are for school candidates, private candidates and job seekers. Candidates of these examining bodies are awarded certificates after examinations.

2.2.2 The West African Examinations Council (WAEC) and Its Mandate

This examination body was established in 1953 to serve West African Anglo countries of Gambia, Ghana, Nigeria and Sierra Leone curriculum development and innovations. The National Policy on Education supported the statements that state: Nigeria will, for the present, continue to use WAEC as its national examination body. The examination body conducts examinations like the Senior School Certificate Examination (SSCE) and General Certificate in Education (GCE) now called SSCE private candidates' examination (WAEC, 2013). It seeks to stimulate and encourage curriculum development. For feedback from the syllabus, WAEC carefully studies the syllabus and makes radical or mirror changes as deemed necessary. Minutes of international, national and subject panels of WAEC provide the best possible source of information for curriculum development and performance in schools of special note are the annual reports of Chief Examiners for various subjects examined by WAEC. These reports are sent to the principal of schools recognised by WAEC for their Examinations.

Nevertheless, irrespective of all criticisms, WAEC has continued to dictate the tempo and direction of curriculum development and innovations in Nigeria mainly through regulating and standardising secondary schools and related institutions programmes. The various certificates awarded by WAEC provide minimum education requirements and qualifications for entry into a higher level of education in Nigeria. The mandates of the West African Examinations Council include:

(c) To administer final examinations for senior secondary school leavers and private candidates in Nigeria, Ghana, Gambia, Sierra Leone and Liberia.

(d) Construction of aptitude tests and sale of such instruments to organisations like banks, insurance houses and other organisations that needs the instrument for recruitment of staff.

(e) Organising workshops and seminars for teachers in secondary school on a new topic in the WAEC syllabus.

(f) Award of the statement of results and certificates to deserving candidates.

(g) To review the existing syllabus as the need arises. (WAEC, 2013)

The WAEC grading system and Interpretation (by Percentage) are presented in Table 2.1.

Table 2.1: WAEC Grading System and Interpretation

S/N	Grade	Interpretation	Percentage
1	A1	Excellent	75% - 100%
2	B2	Very Good	70% - 74%
3	B3	Good	65% - 69%
4	C4	Credit	60% - 64%
5	C5	Credit	55% - 59%
6	C6	Credit	50% - 54%
7	D7	Pass	45% - 49%
8	E8	Pass	40% - 45%
9	F9	Failure	0% - 44%

Source: WAEC Office, Lagos (2013)

2.2.3 The National Examinations Council (NECO) and Its Mandate

Historically, the Nigerian government constituted the Sogbetun Commission of Inquiry in 1981 due to the public cry about the burden on West African Examinations Councils (WAEC) to investigate and make appropriate recommendations to Government on examination processes. Among other things, the Commission discovered that the West African Examinations Council (WAEC) was overloaded. The Commission, therefore, recommended that the workload for WAEC be drastically shelved to other examination bodies to be set up. This recommendation was not put to effect due to bureaucratic bottlenecks. It is on record that WAEC admitted before the House of Representatives Committee on Education in October 1981, that there was a need to set up other examination boards in Nigeria to reduce the burden on WAEC. This admittance by WAEC led to the setting up of the Angulu Commission in 1982

The Federal Government in 1982, in response to continued public cry out for non-implementation of the recommendations of the Sogbetun Commission, set up the Agulu panel. This panel recommended the setting up of three (3) regional Examination Boards for the conduct of Senior School Certificate Examinations (SSCE) and one Board to conduct GCE type of examination for private candidates. The Federal Government under General Buhari was determined to implement this recommendation when political developments overtook its implementation. One of the last activities of the Abdulsalam Abubakar military administration was the declaration of 1999 decree that created the National Examinations Council (NECO). Even though cries for the formation of a national examination body had been on for over two decades earlier, the birth of NECO was not a saved argument. While some Nigerians received it and saw its arrival as an opening for choice of examination body for candidates, others doubted its capacity to conduct reliable examinations that could command widespread national and international respect and acceptability. Some people believed that, as a federal government parastatal, it would provide subsidised registration to candidates; yet some people queried even its legal position.

Through its obligation, NECO was to take over the tasks of the National Board for Educational Measurement (NBEM) which had been established in 1992, though its allowing decree was promulgated in 1993. Yet, the additional responsibility of NECO over those of its precursor (NBEM) called for some restructuring. Not only was the staff strength to be increased, but there was also the need for offices to be established in every state of the federation and the federal capital territory (FCT) if NECO was to meritoriously cope with the enormity of its mandate. Consequently, within one year of its creation, the staff strength of NECO had quadrupled that of NBEM in 1998.

By January 2000, NECO offices had been created in every state and FCT. While NBEM had functioned with six zonal offices located in Bauchi, Ibadan, Katsina, Makurdi, Owerri and Ilorin, NECO established other zonal offices in Akure, Asaba, Damaturu, Enugu, Lagos, Sokoto, Yola, Port Harcourt and Uyo. Kano replaced Katsina (which became a state office) as a zonal office (Ayanwale 2018). The Federal Government of Nigeria in 1999 finally established the National Examinations Council (NECO), consequently, the first Senior School Certificate Examinations (SSCE) was conducted in June/July 2000 for candidates in Nigeria only and has not to change up till now. Its mandates include:

i. To conduct final examinations for final year junior secondary school students (JSS) in Nigeria.
ii. To conduct SSCE examinations for senior secondary students in Nigeria.
iii. To award certificates/statements of results to deserving candidates.
iv. To review the existing syllabus from time to time as the need arises. (FRN, 2013)

As a developing examining body in Nigeria, the NECO examinations being administered include:

1. National Common Entrance Examination (NCEE)
2. Basic Education Certificate Examination (BECE)
3. National Entrance Examination into Federal Unity Senior Secondary Colleges (NEEFUSSC)
4. Gifted Examination into Federal Academy, Suleja
5. Senior Secondary Certificate Examination (SSCE Internal)
6. Senior Secondary Certificate Examination (SSCE External)

The NECO grading system and Interpretation (by Percentage) are presented in Table 2.2.

Table 2.2: NECO Grading System and Interpretation

S/N	Grade	Interpretation	Percentage
1	A1	Excellent	75% – 100%
2	B2	Very Good	70% – 74%
3	B3	Good	65% -69%
4	C4	Credit	60% -64%
5	C5	Credit	55%-59%
6	C6	Credit	50% – 54%
7	D7	Pass	45% – 49%
8	D8	Pass	40% – 45%
9	F9	Failure	0% – 44%

Source: NECO Office, Minna NECO (2001)

2.3 Importance of Mathematics

Mathematics is fairly rich in perceptions that unswervingly transform into good natural life abilities. The significance of Mathematics to normal living cannot be overemphasised. Mathematics Odilli (2006) refers Mathematics to as the subject that aids students to form the practice of clarity, accuracy, brevity, precision and certainty in expression. To effectively prepare students for the future, the influence of Mathematics in personality and a moral building may be considered. In addition, some efforts are required from the students is t to enable them to serve as good citizens which are most required in the society to align with the agenda of re-branding in Nigeria.

The Federal Government of Nigeria, in acknowledgement of the importance of Mathematics, takes revealed pledge to the teaching and learning of the subject, mostly the advancement and of the study of Mathematics at all education levels in Nigeria (FRN, 2013). Through the various developmental programmes of the National Mathematics Centre (NMC) Abuja. The Centre over the years carried out particular Mathematics Improvement Projects (MIP) in its various attempts to increase Mathematics teaching and learning at the primary and secondary school levels. These projects include training/workshops for teachers, distribution of instructional materials to schools and production of mathematical games to stimulate students' interest in the learning of Mathematics. It also includes national incentives schemes like scholarship awards, certificates of merit, book prizes, among others for pupils, students and teachers of Mathematics in institutions that have attained a measure of quality in the teaching and learning of Mathematics at all levels.

According to Okafor and Anaduaka (2013), learners are confronted in Mathematics with the aim of findings out mathematical associated functions. The findings will move learners to the levels of analysing and interpreting experiences learnt and to make generalisations which could be consequently used in fresh situations. Mathematics similarly exposes students to diverse means of solving similar tasks. Availing the learners the ability to tackle everyday issues with varied processes and numerical skills based on the habits learned with the learning of Mathematics. Further, that individuals need Mathematics every day to survive and its fundamentals are hooked on one's skill to solve Mathematical issues.

Besides the import of Mathematics learned at high-level concerning the realisation of objectives of science and technology for a country, In the lower levels of the Nigerian educational system Mathematics taught are likewise helpful in building sensible, considerate and resourceful citizens. The prominence on sound mathematical information for pupils and students in public primary and secondary schools is meant for the pupils and students to gain the benefits of the attainment of mathematical abilities, living a good being and also be in an improved situation to unswervingly support the growth of the nation and worldwide economy.

2.3.1 Mathematics Education in Nigeria

Mathematics as a subject is an essential part of all human beings life and affects almost all areas of undertakings of people. Everyone needs Mathematics to live no matter how elementary**.** It was illustrated in the study of Usman (2002) that everywhere people go, everything people do or propose to do, the structure and applications of Mathematics play a vital role and that is why most people, races and countries emphasise all the aspects of studying, developing, and applying Mathematics skills in day to day activities. Mathematics is also a body of knowledge that is vital for the attainment of a scientific and technological nation. However, the difference between the developed and the underdeveloped nations is based on the level of mathematical achievement and ingenuity (Ale and Adetula, 2010). According to the authors, Mathematics is an acknowledged agent of national growth and wealth establishment. Also, Nosa and Ohenhen (1998) noted that evidence abounds to show that nations that embrace Mathematics, science and technology enjoy a better standard of living and are less dependent on others. Today, everything in society has become more and more reinforced by mathematical concepts.

However, the major improvement in Mathematics education presently is the increased quantity of Mathematics concepts that all individuals are anticipated to be acquainted with. Technological leaders and political leaders need mathematics education that considers both the new uses of mathematics and technology and new ways in which

Mathematics can be done with information technology. The way Mathematics is presented in the classroom by teachers makes many students see the subject as a very difficult one. Instead of memorising properties and definitions, students should develop personally; meaningful concepts and ways of reasoning that enable them to carefully analyse spatial problems and situations. This was by one of the principles of learning Mathematics as stated by the National Council of Teachers of Mathematics (NCTM, 2000) that students must learn Mathematics with understanding and actively build new knowledge from experience and prior knowledge.

Supporting the concept, Lotfi, Dehkordi and Vaez-Ghasemi (2012) submitted that one of the best teaching methods is that which produces the knowledge and skill needed by students and helps the students to use the acquired knowledge and skills in real-life situations. Buttressing the same point, Bethany and Michael (2012) asserted that Mathematical competency rests on the development of both conceptual and procedural knowledge. Meaning that, in learning mathematics, the balance must be maintained between conceptual and procedural knowledge.

Mathematics is a vital course for the scientific and technological advancement of any country. Studies have shown low enrolment of students in Mathematics Education when compared with other courses in Social Sciences and Arts, particularly at the higher level of Education. This issue is viewed with serious worry for teaching and learning Mathematics in schools. Mathematics Education is a programme or course of study operating by the Nigerian Colleges of Education and Faculties of Education in Nigerian Universities. The programme comprises a combination of Mathematics with Physics, Chemistry and Statistics among others. Students studying this programme at the College of Education and University levels are trainee Mathematics teachers who are anticipated to teach Mathematics at primary and secondary school levels. These groups of undergraduates, after the programme, are considered as professionally qualified teachers to handle the teaching and learning of Mathematics in Nigerian primary and secondary schools. The consequence of this is that efforts must be directed towards adopting instructional strategies that will not only deposit procedural knowledge in students but also promote conceptual skills needed to build conceptual understanding in students. This was in line with the suggestion made by the WAEC Chief Examiner's Report (2013) that teachers should help students improve their achievement in mathematics by reducing its abstraction, and remove their apathy and fear of the subject.

It has been reported also, that many students do not show interest in Mathematics and not be bothered about the offer of the subject. These students are filled with hatred, fear and phobia for this subject. The lack of interest by these students in Mathematics according to Usman and Nwabeze (2011) was caused by the teachers' use of insufficient, uninteresting ways of exposing or imparting the knowledge of Mathematics to students using conventional approach and this is one of the main reasons for the massive failure in Mathematics examinations (Maduabum and Odili, 2006; WAEC, 2013). This ugly situation spurs the implementation of various innovative methods such as the Cooperative learning method (Chianson, 2011), the Montessori Method (Kurumeh, Agogo and Usman, 2010), the Games and Simulation Technique method (Achor, Imoko and Ajai 2010), hitherto, the issue still ensues. Even with the effort of the government on the advancement of mathematics teaching and the establishment of prospects for the development of teaching, the problem persists. Maduabum and Odili (2006) enumerated the following problems:

i. Shortage of mathematics teachers.
ii. Poor government policy.
iii. Lack of curriculum integration.
iv. Poor classroom arrangement by teachers
v. Lack of good instructional strategies and materials.
vi. Teacher's impatience and un-preparedness.
vii. Lack of equipped mathematics laboratory for practical.
viii. Overpopulation of students may impede effective demonstration during practical.
ix. Poor remuneration of teachers.

Ojumba (2012) submitted that causes of student's low achievement in Mathematics in Nigeria are:

i. Acute shortage of qualified professional Mathematics teachers
ii. Exhibition of poor knowledge of Mathematics content by many Mathematics teachers
iii. Overcrowded Mathematics classroom

iv. Students bad attitude towards Mathematics syllabus at the expense of meaning leaning of Mathematics concepts

v. Inadequate facilities and Mathematics laboratory

However, additional identified reasons are a misunderstanding of the subject to be a difficult one, Mathematics anxiety, parent attitude, fear and assessment procedures. It is therefore known that the reasons for students low achievement in senior secondary school Mathematics students are numerous and diverse but fall under student, parent and school-based reasons.

2.4 Concept of Constructed-response Item

Constructed-response items are implicit unrestricted items or other forms of inducements that entail students to supply answers and be examined by content-expert panels or raters. In education, the essay items are the most used items of the constructed-response tests. Constructed-response items involve the students producing in black and white responses to a question. Constructed-response tests have been used to examine students for over 100 years. According to Livingston and Ruppy, 2004, constructed-response (CR) test formats have many fortes and is the only assessment design suitable for testing text abilities like the sufficiency in the skill at writing an equation, creation of sentence and paragraph, ability to organise rational opinions. All constructed-response questions entail non-cued written responses from students. All essay or timely questions comprises a straight inquiry on an exact attentive subject and make available adequate data to students to respond to questions. The applicable directions about responding to the questions, like the probable size of the response, time restrictions, details of the response are specified.

The constructed-response test design may allow the essay reader to grade detailed phases in working through a problem or the reason of each phase used in cognitive or problem solving, which may enable partial credit scoring (as opposed to “all or nothing” scoring). constructed-response test design may be most time-efficient (for the instructor) in assessing small sets of learners since less time will be spent writing essay items. Conversely, some *tasteful* and potential problems are associated with constructed-response items such. According to Downing (2006), constructed-response tests are hard to score objectively, accurately and reliably. The marking of the constructed-response test is consuming much time and is expensive. The content-related validity is sometimes negotiated, especially for huge content domains due to some sampling issues related to assessment time restrictions. Also, there are a lot of potential threats to the validity of CR items, all related to the more subjective nature of essay scores and various biases associated with human essay readers. There are a smaller amount of psychometric quality-control procedures, such as item analysis and cleaning.

During the preparation of the constructed-response test, a perfect response to the item should also be arranged by the author of the test The constructed-response test details, together with its precise instructions for students, the model response, and the real grading rulebook should be arranged well in advance before the administration of the test, to ensure adequate time for revision, editing and review. In the Mathematics selected-response test, students do solve many kinds of the problem but it does not mean that they are capable to construct a mathematical proof. Also, the writing ability in a selected-response test can define if the examinees can differentiate among well written and badly written equations

One other reason for employing constructed-response items is that an examinee that can choose the right answer from the list of possible answers, will possibly supply the answer even when it is seen. Results from some research have proven that SR and CR tests offer similar statistics. Thus, it was concluded that selected-response tests can serve as an alternative to constructed-response tests (Lukhele, Thissen and Wainer, 1994). Welch (2006) stated that it is centred around studies displaying an advanced acceptance of scores on SR and CR tests. Generally, many studies related the selected-response and constructed-response marks of a set of examinees who were once examined with the two kinds of tests and there was confirmation that the high-level of total agreement could cover significant variances among groups of examinees.

However, it has been revealed by some studies that man and woman differences in CR items do not equate the man and woman differences on selected-response items in the same focus (Mazzeo, Schmitt and Bleistein, 1992; Livingston and Rupp, 2004). Naturally, when boys and girls do similarly well on the selected-response tests, the girls outclass the boys on the CR test. However, when boys and girls perform similarly well on the constructed-response items, the boys outperform the girls on the selected-response items. All the observed variances happen notwithstanding the selected-response marks and the constructed-response marks likely to reach agreement strongly within each set. There are two general forms of essays. Namely; the ones that required long responses and those that required short responses.

A long-answer essay may involve the students writing one to two or more sheets in responding to the item, whereas short-answer essay items may entail one or two paragraphs was written responses. Most tests written in pen and paper are expedient in the assessment of cognitive abilities, knowledge, achievement, and skills. The main controlling element in defining the correctness of any testing design correlates to its determination, the anticipated clarifications of marks, the model hypothesised to be measured, and the critical magnitudes of the test. But, the relationship between the selected-response and constructed-response scores can vary in a pool of examinees, For instance, in boys and girls. In addition, the development of abilities assessed by the CR items will not be revealed in the marks on the selected-response items. The answers to CR items can be estimated either in analytical or holistic ways. Under the analytic estimation, reliable scores are obtained among examinees but a large number of types of constructed-response items entail holistic scoring.

2.4.1 Ethics of Writing Constructed-response Items

The authors of performance assessment items must follow similar rubrics of item writing employed in the design of selected-response items. Downing (2006) expressed that constructed-response item text aids the attention to these values, revisions and editing based on an autonomous appraisal by other content experts. However, the simplicity in the connotation is a crucial representative for all items as such tests are extremely scrutinised by students for understated connotation. In all tests processes, it is the content to be examined that is the most central and considered; the design designated for the test is secondary to the content to be examined. In the course of the preparation of the essay-type question, a model or ideal answer to the question should also be prepared by the author of the question, just as a corrector best answer key should be designated by a selected-response question writer. The specification of the model answer must equal the directions to students (Welch, 2006). This ultimate answer will form the source of a scoring rubric (the scoring key for constructed-response questions) employed in the real scoring of the answer to the essay item.

The CR item, has specific instructions for students, the ultimate answer, and the actual scoring rubric should be prepared well in advance of the test administration, so that time for review, revision and editing are available. The set of items in a test is constant at a particular time of measurement. Therefore, the relationship between two items should not differ knowingly from zero, else, it may be said that the answers to the items are influenced by extraneous factors other than what the instrument is intended to measure. The axiom of local independence states that the observed items are independent of each other given a person's score on the latent trait (Vermunt and Magidson, 1996). Local independence means statistical independence. When items are statistically independent, each exhibits its quality and takes students' good display of ability in unfolding the characteristic function about them (Yen, 1993). Based on the assertion by Yen (1993) one can posit that to test items that would not violate the theory of local independence, the interaction between each item must not be high but should be as low as possible concerning the logic of operations and manipulations.

2.5 Concept of Dichotomous and Polytomous Model

Several Item response models have been developed for dichotomous and polytomous data. The frequently used models for dichotomous tests are the logistics models (one, two, three-parameter). The Single/one Parameter logistic model was first printed by the Danish mathematician Georg Rasch in the 1960s. Rasch approached the analysis of test data from a probability theory point of view. The model is one of the most generally used IRT models in many IRT applications. Assume a J binary items, X_1... Xj, where 1 indicates a correct answer and 0 is an incorrect answer. The equation for the Rasch model is defined by the following:

$$\boldsymbol{P_i(\theta)} = \frac{\mathbf{1}}{\mathbf{1}+ e^{-(\theta - b_i)}} \qquad \text{….Eqn.2.1}$$

Where: $P_i(\theta)$- The probability of a correct answer for the ith item, b_i - is the difficulty parameter for the ith item and e*is* the ability level. This likelihood can be explained by the curve in Figure 2.1, which is called the item characteristics curve (ICC) in the field of IRT. From this curve, it can be observed that the probability is a monotonically growing function of ability. That is as the subject's ability increases, the probability of a correct answer increases.

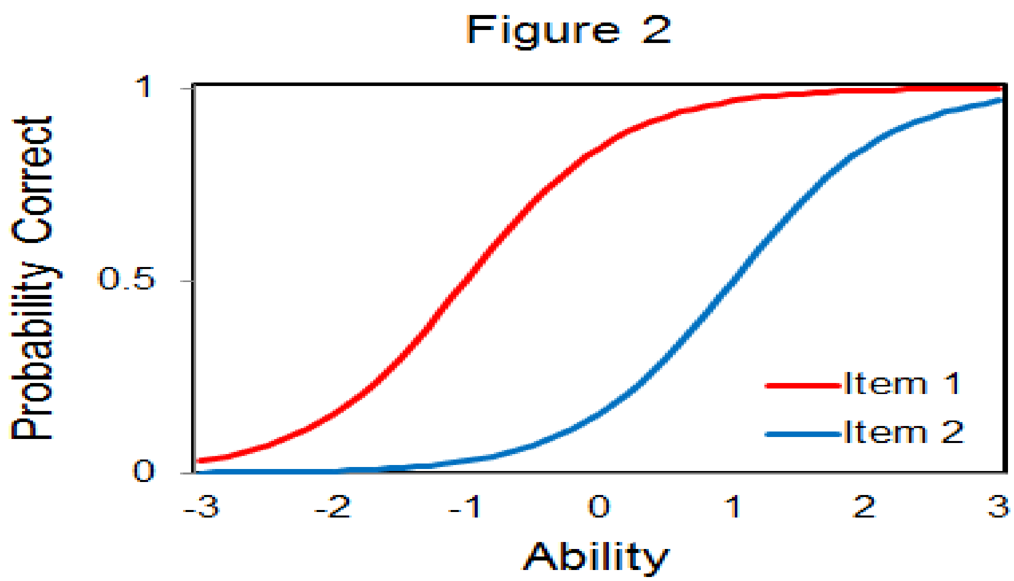


Figure 2.1: Hypothetical example of One- Parameter Logistics Model.
Source: The University of North Carolina at Greensboro

From Figure 2.1, the item response functions for Item 1 and Item 2 with dissimilar values of b_i; the value for Item 1 is -1 (that is., b_1 = -1) and the value for Item 2 is 1 (that is., b_2 = 1). Note in what manner the value of b_i stipulates the horizontal position of the item's item response functions. b_i rises, the item response functions shift to the right and the item turns out to be harder. so, Item 1 is less difficult compared to Item 2, as for a given level of capability there is a greater likelihood of rightly answering to item 1 compared to item 2. Note also, that the possibility of the right answer to item 1 equals 0.5 at a capability value of -1, as would be predicted given that b_1 = -1. Likewise, the probability of correct answer to Item 2 matches .5 at an ability value of 1, as anticipated given that b_2 = -1.

The Two- Parameter logistic model (2-PL) or Birnbaum Model was Birnbaum (1968) modification of the 2-PL model to contain a parameter that denotes the influence of guessing to the probability of correct answer. But, in so doing, some of the nice mathematical properties of the logistic function were lost. IRT modelling of students responses on the test items can be examined by fitting the two-parameter logistic model to the student's answer data. The 2-PL is finest defined by its item characteristic curve (ICC), a mathematical function that correlates the probability of getting the right answer on an item to the ability measured by the test and the characteristics of the item.

In the 2-PL model, at the point of contra flexure (inflexion) consistent to 0.5 probability value, a common tangent is drawn and the slope of the tangent is designated as "item discrimination" and this value is indicated by the letter "a". The "b" and "a" value is estimated for the given items. The equation to the 1 is:

$$\ldots \boldsymbol{P_i(\theta)} = \frac{\mathbf{1}}{\mathbf{1+}\, e^{-D_{ai}(\theta-b_i)}} \qquad \text{.Eqn.2..2}$$

Where, Pi (θ) = Probability of getting the correct answer to the item I of an individual with ability θ

θ = Individual ability

b = Item difficulty

a = Item Discrimination

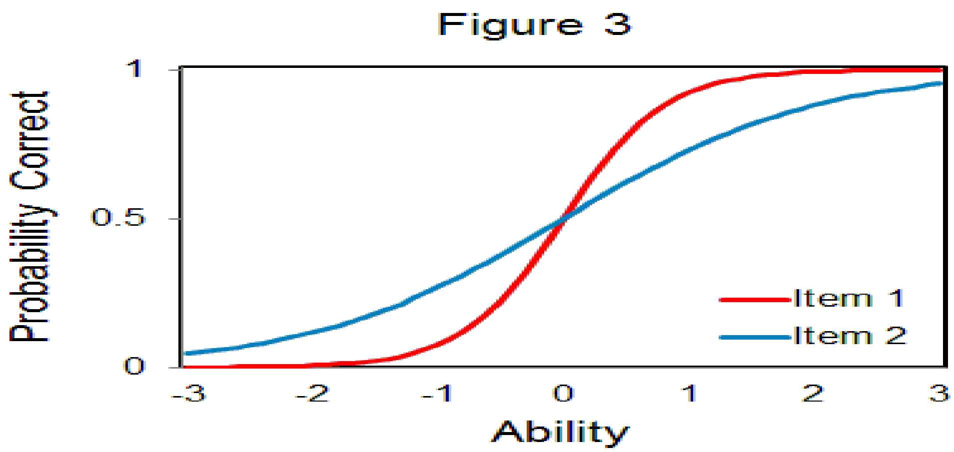


Figure 2.2 Hypothetical example of Two- Parameter Logistic Model.
Source: The University of North Carolina at Greensboro.

In Figure 2.2, the item response functions for Item 1 and Item 2 having diverse values of a_i; the value for Item 1 is 2.5 (that is., a_1 = 2.5) and the value for item 2 is 1 (that is., a_2 = 1). These two items have a value of b_i = 0. Note that the value of a_i stipulates the steepness of the item's item response function; as a_i rises, item response functions turn out to be sharper and the item tends to discriminate. So, Item 1 has an upper degree of discrimination compared to item 2 and therefore would make available detailed information about the person's level of capability. For this reason, items with a higher value of a_i tend to be observed as having more required psychometric properties compared to items with a lower value of a_i.

The Three- Parameter logistic model (3-PL) (Fred Lord's model) Fred Lord's Model (3-PL) is the most general form of parameter model for scoring dichotomous responses. In testing, there is the probability that students will get an item correct by guessing. Under the 3-PL model, there is, in addition, a third parameter called guessing for the item and is designated by the letter "c". This is given by the intercept of the probability axis that indicates the probability of guessing the right answer. The guessing parameter is unique to the item and is independent of the examinee ability. Thus, the guessing parameter remains a constant for all examinees of various abilities. The Lord's or three-parameter logistic model includes a parameter that represents the contribution of guessing to the probability of correct response. The equation for the curve is:

$$\boldsymbol{P_i(\theta) = C_i + (1 - C_i)\frac{1}{1+ e^{-D_{ai}(\theta - b_i)}}} \qquad \text{… .Eqn.2.3}$$

Where,

$P_i(\theta)$- The probability of a correct response for the ith item.

b_i - is the difficulty parameter for the ith item

ai - is the discrimination parameter for the ith item

C_i - is the guessing parameter for the ith item

θ - is the ability level

D - Represents a scaling factor, which is set to 1.7

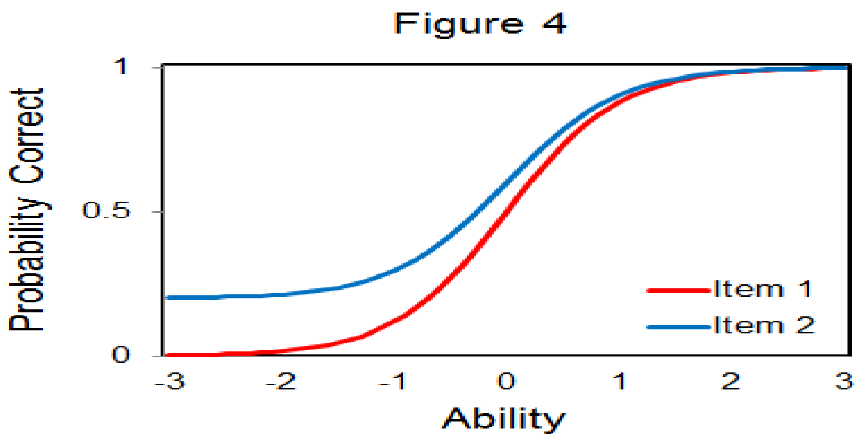


Figure 2.3: Hypothetical example of the Three- Parameter Logistic Model. Source: The University of North Carolina at Greensboro.

In figure 2, item response functions for Item 1 and Item 2 with diverse values of c_i; the value for Item 1 is 0 (that is., $c_1 = 0$) and the value for item 2 is 1 (that is., $c_2 = 0.2$). These two items have the value of $b_i = 0$ and $a_i = 2$. On the other hand, the value of $c_2 = 0.2$ Item 2 causes the lower bound of the item response function for item 2 to be higher compared to item 1, in place of the presence of guessing in Item 2. Whereas the value of b_i denotes the level of capability at which the likelihood of the right answer equals 0.5 under the one parameter logistics and two-parameter logistics models. This does not hold in the three-parameter logistics model when $c_i > 0$. This property is revealed in Figure 2.3, whereby the likelihood of right answer to item 1 equals .5 at $\theta = b_1$ (recall that $c_1 = 0$), but the probability of correct answer to item 2 equals .5 at $\theta < b_2$ (recall that $c_2 = 0.2$).

Despite the robustness of IRT models, the assessment of ability and item parameters can only be attained when the items in the test meet the rules fundamental to its outline. Naturally, three rules are made in stipulating item response models. One relates to the dimensional structure (that is, uni-dimensionality) local independence of the test data and the other relates to the mathematical form of the item characteristics function or curve (ICC) (Wiberg, 2004). Additionally, the selection of the particular IRT model for item standardization is ruled by whichever selecting the framework that fits the data or selecting the data that fits the framework (Yen, 1993). Nevertheless, the choice of the data that fits the model has been criticised by researchers because of its needless negative impact on the content and constructs validity of a test being studied. The item response frameworks are established around the logistic accumulative circulation function. This logistic equation once plotted, produce a graph that refers to as item characteristics curve (ICC) once the ICC is plotted the trait of the students is represented as theta (θ) on the x-axis, but the likelihood of a student properly answers the test is denoted using P (θ) around y-axis as exemplified in Fig. 2.4

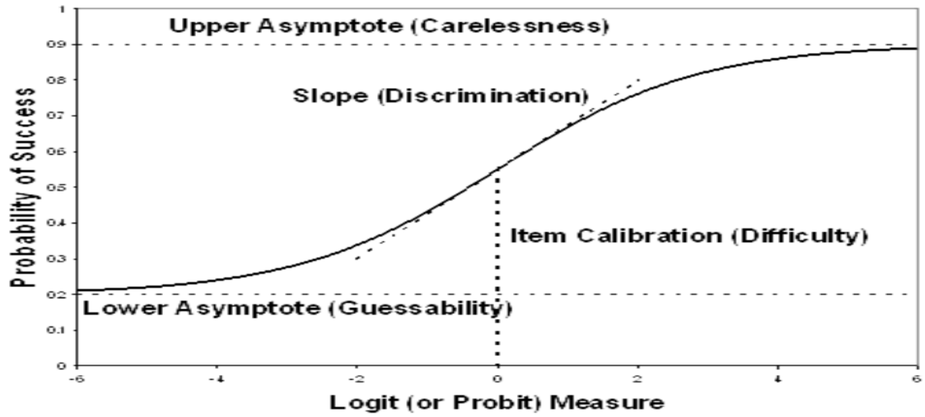


Figure 2.4: Item Characteristics C
Source: An evaluation of the comparability of item analysis results (Ojerinde, 2013).

However, ICC naturally takes the S-shaped curve named ogive (∫). As displayed in Fig 2.4. Skill derived from tests that are established by this theory can be attained independently from the group. More importantly, the fundamental concept of item response theory is centered on the item characteristic curve (ICC). In the dichotomous question, schemed function match ability levels to the likelihood of answering that item properly. All test items have their item characteristic curve. The two commonly known types of item characteristic curves include the normal ogive and logistic ogive models.

In determining personality and social variables dichotomous differences are frequently less in the ability context than inability estimation condition. (Ostini and Nering, 2006). One of the psychometric problems that make polytomous items more attractive than dichotomous items is the measure of that polytomous item through a broader series of the attribute range than dichotomous items. This happens because polytomous items encompass more response categories than dichotomous items. However, the first polytomous model was established by Samejima (1969), which was called the graded response model while Bock and Samejima (1972) established another type of polytomous model (Nominal Response ModeL). The IRT polytomous models came into existence in the early 1980s. The first IRT applications involved mainly unidimensional IRT polytomous models have been established. These models generally are direct extensions of unidimensional models (Liu, 2007).

Each item response model predicts the likelihood that a particular individual will give a definite response to a definite item. In dichotomous models, there are two response functions, the correct (positive response) and the incorrect (negative response). Normally, the response function for one response category is modelled clearly because the functions for each category are the supplement of one another. It is only the correct response that is modelled, with a monotonically cumulative function. The corresponding feature of the category functions shows that identifying the characteristics of one function expresses all that is required to distinguish around the other function.

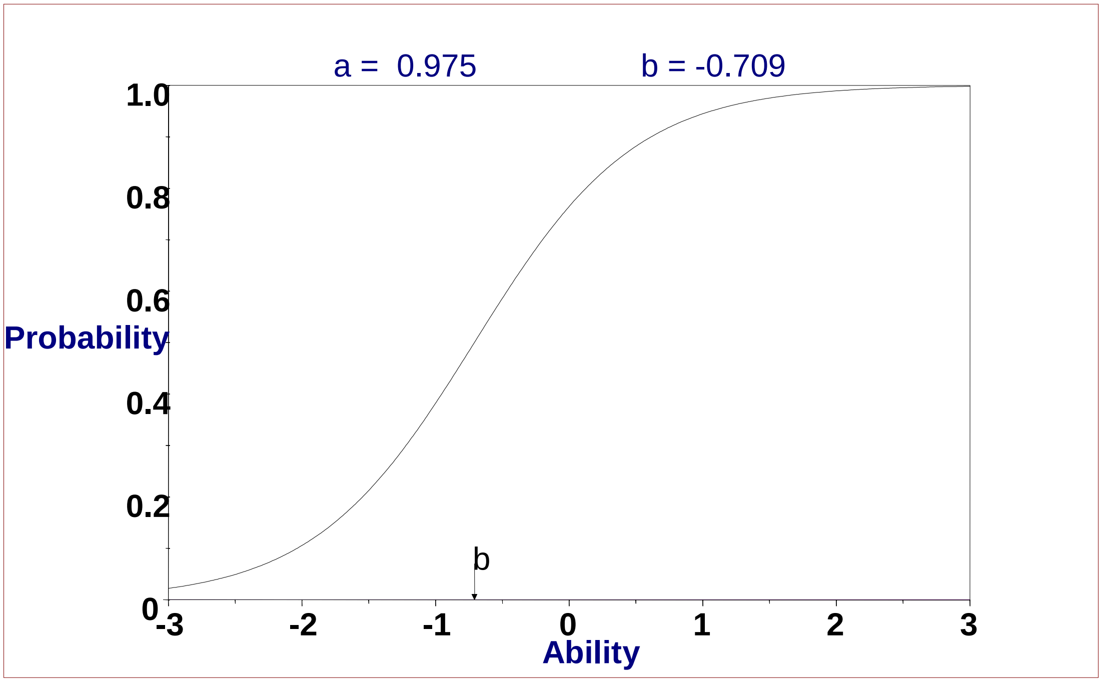


Figure 2.5: Item response function for dichotomous item
Source: Estimation of students ability in physics essay test (Adegoke, 2016)

In nature, some questions in the measurement setting with multiple response options are known as polytomous items. Such measurement tools include rating scales, questionnaires (with Likert-type response format), and ability tests (with partial credit for partially correct responses). The IRT Polytomous models for such items operate quite in a different way from dichotomous models. Here, the information on the features of a response categories functions does not define the features of the other category functions, and each category function, thus, it must be modeled clearly. The effect of this is that the non-define nature of the category response functions is no longer monotonic functions. For questions with the ordered categories, only the functions for the extreme negative and positive categories are monotonically diminishing and accumulative respectively.

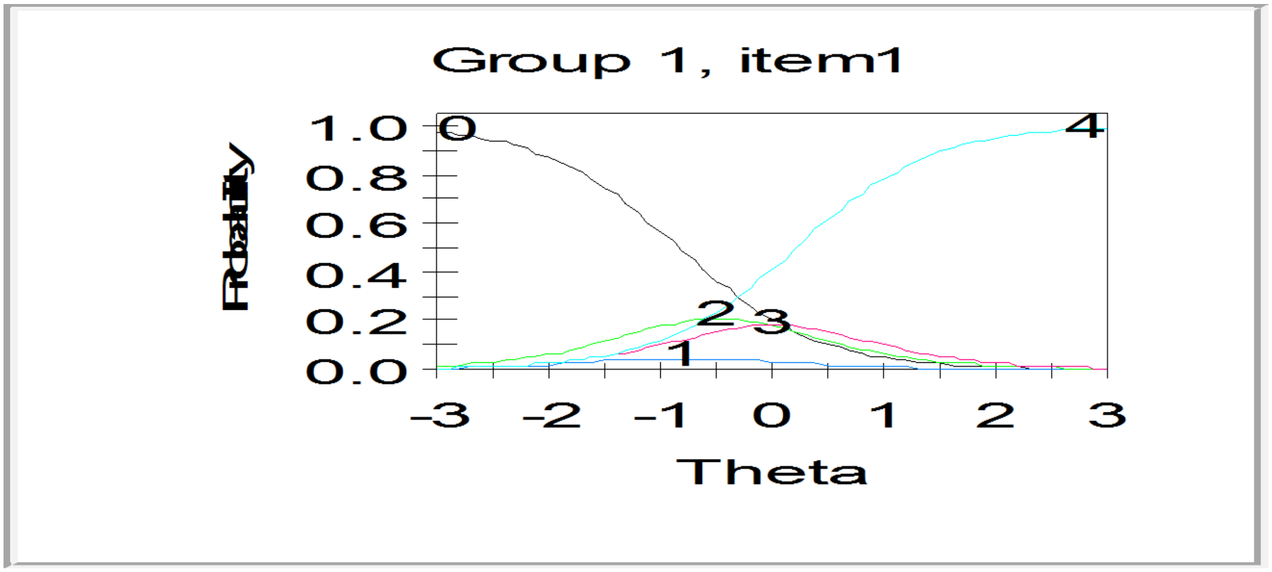


Figure 2.6: Item response function for polytomous item
Source: Estimation of students ability in physics essay test (Adegoke, 2016)

Supporting Figure 2.6, Dawadi (1999) reported that in polytomous items analysis the attribute evaluation was presumed to measure the typical attribute of a person of two likewise significant traits and the main trait of the two inequitably significant traits (75 % of the aggregate sum of questions was examined), the process was usually vigorous to the abuse. On the other hand, once the ability evaluation was presumed to measure one of the two similarly significant abilities (the dimensional strong point was fifty-fifty) so is the minor traits of two unevenly significant traits (twenty-five per cent of the overall amount of items), the assessment process was not vigorous.

However, the polytomous items are desired because of the belief that they tend to increase the test validity. As a result of the long time spent to test and score constructed-response items, many researchers have viewed the worth of using polytomous questions in testing. In a simulation study report, the classification of correctness in computerised testing circumstances gave better precision for polytomous items compared to dichotomous questions. Lesser wrong negative and wrong positive ordering errors and the total error rates were stated for polytomous questions compared to the dichotomous questions. The influence of test length limitation was lesser for polytomous questions compared to dichotomous questions (Lau and Wang, 1998; Adegoke, 2013).

Under the IRT polytomous models scoring process, items are scored beyond just right or wrong through the method of assessing each category of responses and they are scored giving the grade of accuracy or the measure of facts presented as the complete response. One of the advantages of polytomous questions is that, the larger the number of response categories, the additional statistics provided over a wider range of the trait continuum than dichotomous items. Masters (1988c) and Bejar (1977) stated that the entire aim of using more than two categories of responses per item is to obtain detailed facts about the attribute level of persons that were examined and to obtain accurate trait-level estimates. In addition, Masters (1988a) reported additional comprehensive analytic information about persons and question could be acquired from polytomous test questions.

However, the dissimilar forms of IRT polytomous models that have been established include:

i. Partial Credit Model.
ii. Generalised Partial Credit Model or Rating Scale Model.
iii. Nominal Response Model.
iv. Graded Response or Modified Graded Response Model.

2.5.1 The Partial Credit Model and Its Parameter Interpretation

This model is an extension of the Rasch model. The model was developed by Masters (1982) and it was expected that answer category were ordered by the levels of ability its symbolize. This is to conceptualise items with multiple stages that each stage will denote the change in ability levels between two adjacent categories. The PCM was specified for each phase accomplished. Consequently, PCM is an adjacent category model. Mellenbergh (1995) stated that adjacent category models are ordered polytomous item responses divided into pairs of adjacent categories, which mean that (k_{th} and $(k+1)_{the}$ categories, for k=1,2,…,m-1) and applied Bock’s model to the log ratio of the pairs, as follows: That is if m_i is the total of stages in an item and the response categories of the item can be denoted by the partial credit given to them, that is, 0 to m_i. The model is described by m_i log-odds:

$$ln\left[\frac{P_{ik}(\theta)}{P_{i(k-1)}(\theta)}\right] = \theta - b_{ik} \qquad \dots$$

Eqn.2.5

Where k = 1…,m_i

However, there is a basic difference between the dichotomous and polytomous item response models which is the axes mainly on the fundamental measurement belief. According to Ostini and Nering (2006), the difference is in Rasch-type models that endeavour to follow the crowd to essential measurement model and more practically established models that do not conform to the essential measurement concept, at last, entails that item response models follow the rule of explicit objectivity. At the model parameters level, explicit objectivity entails that contrasts between item and individual parameter values and any other item parameters not being compared and dependent. However, the two forms of parameters (person and question) in the measurement framework must be preservative and distinguishable. (Rasch,1977). This means ($\theta - b$) and as in a ($\theta - b$) formulations.

Hence, in estimating dichotomous item response function, it is required that the appraisal of two or three parameters (a, b and or c. This is untrue if devotion to the rule of explicit independence is required. Then, the discrimination (a) and pseudo-guessing c parameters do not need to be assessed. The Rasch dichotomous framework, thus, requires an assessment of an item location parameter. Similarly, item response models studied and identified by Adedoyin (2010) are one, two or three logistics parameters. The modest model is the 1-PL and there was a recommendation that the use of these models should start with the 3PL model wish is the most composite of item response framework. Symbolised thus:

$$P_1(\theta) = C_1(1 - C_1)\frac{1}{1+ e^{-Da_1(\theta-b_1)}} \quad \ldots eqn\ 2.6$$

Where Ci is signified the guessing factor

ai is the item discrimination parameter which is referred to as item slope.

bi is the item difficulty parameter defined as the item location parameter.

D is the arbitrary constant (normally D = 1.7) and

θ denotes the ability level of particular students.

The location parameter of an item is on a similar scale of ability θ and taking the value of θ at the point at which an examinee with the level of ability has a fifty-fifty likelihood of responding to the question rightly. The slope of the target line of the ICC at the point of the location parameter is identified as item discrimination.

The guessing element is signified as zero (0) (C1 = O). The 2PL that must be assessed is specified as

$$\ldots P_1(\theta) = \frac{1}{1+ e^{-Da_1(\theta-b_1)}} \quad \ldots eqn2.7$$

If there are circumstances that all the items have equivalent and permanent discrimination, then a1 will turn into a continuous relatively than a variable, thus the parameter does not require assessment and the item response model is more concentrated to:

$$P_1(\theta) = \frac{1}{1+ e^{-D(\theta-b_1)}} \ldots eqn\ 2.7.1$$

Thus, the 1PL of the item response model limits put on two out of the three likely item parameters that require to be assessed, the 3PL is the most common model than others (1PL and 2PL models) can be measured as models considered in the 3PL model (Lord, 1980; Hambleton and Swaminathan 1985).

2.5.2 The Generalised Partial Credit Model (GPCM) and Its Parameter Interpretation

One of the most powerful polytomous models is the generalized partial credit model (GPCM). This model was developed by Muraki (1992) with an additional slope parameter to the partial credit model already in existence. The slope parameter is similar to item discrimination in the 2-PL dichotomous model. Under the generalized partial credit model, the slope parameter is continuous through the categories within each item but could be varied between items. In the dichotomous 2-PL or 3-PL IRT polytomous models, it is only the slope parameter that is responsible for producing the item discrimination and combine with the formation of item thresholds determined by the discrimination of an item. However, the partial credit model (PCM) and generalized partial credit model (GPCM) log-ratio are the same except that discrimination term is added. to GPCM.

$$ln\left[\frac{P_{ik}(\theta)}{P_{i(k-1)}(\theta)}\right] = a_i(\theta - b_{ik}) \quad \text{… .Eqn.2.4}$$

Where k = 1,m_i

-

Where k= 1…m

$$ln\left[\frac{P(x \geq k)}{P(x < k)}\right] = \alpha_j \left(\theta - \beta_{jk}\right)$$

… .Eqn.2.4.1

$\alpha_j$is the estimated discrimination parameter for item j

$\beta_{jk}$is the estimated category boundary threshold between categories k and k-1 for item j.The estimated variance of the latent trait (θ) is fixed at 1. (Muraki, 1992)

2.5.3 Graded Response Model and Its Parameter Interpretation

This model is an extension of the 2PL model which is better for to attainment of categorical outcomes. The graded response model stipulates m - 1 “boundary” response functions that specify the cumulative likelihood for a response category greater than the choice of concern. The equation for such a boundary response function is closely correlated to the 2-parameter logistic model for dichotomous response data:

$$Pj_1(\theta) = \frac{expa(\theta - b_1)}{1 + e^{xp(a_1(\theta - b_1))}}$$

---equ 2.8

$\alpha_j$is the estimated discrimination parameter for item *j*

$\beta_{jk}$is the estimated category boundary threshold between categories k and k-1 for item j.The estimated variance of the latent trait (θ)

Nevertheless, the graded response model refers to an indirect model in that the likelihood of answering each category is apprehended by attaining the item response functions from the difference between adjacent step functions (Samejima, 1969). The GRM estimates probabilities based on the 2PL specification such that separate b_{ik} parameters are assessed for each step of the item, and one a_i parameter is used for all steps for each item. The GRM stipulates m - 1 “boundary” response functions that indicate the cumulative probability for a response category greater than the option of interest. The equation for such a BRF is closely correlated to the 2PL logistic model for dichotomous response data:

$$Pj_1(\theta) = \frac{expa(\theta - b_1)}{1 + e^{xp(a_1(\theta - b_1))}} - \qquad -\text{eqn. 2.9}$$

The *bik* is interpreted as the target trait value at which *Pi0* (ɵ) = .5, *bim* as the target trait value at which *Pim* (ɵ) = .5, and for values in between steps (*bik* + *bik* + 1)/2 relates to the modal point of the IRF for *Yi* = *k* (Penfield, 2001). The explanation for using a graded response model, or any model built on ordered response categories, with test-based scores, is that test-based scores can hypothetically have an ordered quality if they “correspond to the extent of completeness of the students’ reasoning process within test items” (Lee and von Davier, 2011). This means that the more dichotomously scored measurement chances within one test item are responded to rightly by a student, the more broad the student's ability.

2.5.4 Nominal Response Model and Its Parameter Interpretation

The nominal response model (NRM) handles all item response options that are not ordered or categorised in a pre-specified mode such as the nominal or multiple-choice models which are used to characterise item responses. In NRM, the processes have been traditional to define the likelihood of an examinee answering one of the existing categories of an item (such as the multiple-choice question). The NRM has been employed frequently in test applications with MCQ items (Wainer, 1995). In the nominal response model, the item response function for $Y_i = 0$ is described as

$$Pj_1(\theta) = \frac{1}{1+\sum_{K=1}^{M} \quad e^{xp(c+aik\,\theta)}} \qquad \text{---eqn.2.10}$$

And the IRF for Yi= j> 0 is

$$Pj_1(\theta) = \frac{1}{1+\, e^{xp(a_1(cik+aik_1))}} \qquad \text{----eqn 2.11}$$

Where c_{ik} is a location parameter such that the joining of item response functions for $Y_i = 0$ and $Y_i = k$ is at theta = $-c_{ik}/a_{ik}$. Therefore, in each item, there are 2*m* item parameters. Some other forms of the nominal response model have been proposed as an explanation for guessing behaviour in candidates with low target ability (Thissen, Steinberg, and Fitzpatrick, 1989). More than a few rules have been drawn in the past. These data are not constant but ordered and have different numbers of categories, which hypothetically excludes the adoption of a rating scale model Ostini and Nering, 2006). The residual selections include the adjacent category (that is generalized partial credit model) and cumulative boundaries (that is, graded response model). Samejima (1996) presented explicit mathematical standards moderating the reliability between the cognitive process of response production and the measurement framework.

2.6 Scoring of Constructed-Response Items

The issue of validity challenge for constructed-response test items is associated with scoring. Constructed-response scoring is fundamentally biased and thus entails consideration to several issues in reducing the bad influence of bias on scoring validity. Analytic and holistic ratings are commonly used to score constructed-response (constructed-response and are scored through analytic approaches, CR is graded in some dissimilar categories or for some diverse features. however, the scoring is centered around a set of rules that refers to the *rubric*. The rubric tells the scorer what features of the response to focus on and how to adapt how many points to award to an answer. An analytic scoring rubric lists explicit structures of the response and specifies the number of points to award for each feature. For an item in applied Mathematics, the scorer may award one point for identifying the significant variables, one point for writing an equation that will solve the problem, and one point for solving the equation rightly (Livingston 2004). For instance, analytic scoring might entail ratings of the correctness of the answer to the question and the specificity of the answer, the organisation of the written answer, the writing quality, and so on.

Analytic methods entail the scorer to concentrate on several different aspects of the essay, all of which are presumably related to the quality of the essay answer and the construct intended to be measured by the essay question. Score points are assigned to each analytic segment or aspect of the essay, based on some justification. Holistic or global ratings require the essay scorer to make only one single rating of the overall quality of the written answer. The process of holistic scoring is very different. The scorer reads the response and makes a single judgment of the quality of the response by assigning numerical scores (Livingston and Ruppy, 2004).

A holistic scoring rubric usually comprises statements describing the characteristics of a typical response at each score level. However, to define the score levels in practical terms that the scorers can apply requires exemplars, actual responses written by examinees, selected as examples of a 5-point response, a 4-point response and so on. The exemplars also include borderline cases, for instance, a response that just barely qualifies for a score of 5 or a response that narrowly misses earning a score of 5. Analytic scoring tends to be more consistent from one scorer to another than holistic scoring; the same response, scored by two different scorers, is more likely to receive the same score from both scorers. Analytic scoring works well when:

i. The item developer stipulate the structures of the response for which examinees should receive points or lose points.

ii. The important structures of the response can be assessed distinctly; the feature of the response is independent of interactions among those features.

iii. under the analytic scoring system, the scoring benchmarks can be stated as a set of yes-or-no items. (Did the learner rightly pinpoint the methodical rule? Did the learner present another valid example?) Various analytic scoring systems holistically twist a little, permitting the scorer to award partial credit for some features of the response. For instance, 2 points are for a complete right description and 1 point for a partly right description. Specific kinds of constructed-response questions (some items projected to test writing ability), it is not likely to define the value of an answer in terms of explicit features that are either present or absent.

All responses to these items are scored holistically. One of the greatest difficulties in constructed-response testing is the time and expense complicated in the scoring. The scoring procedure entails considerable quantities of time from extremely trained scorers and often comprises elaborate systems for monitoring the reliability and correctness of the scores. Not long ago, some investigators have made a great deal of progress in using computers to score the responses. Automated scoring offers the possibility of greatly lessening the time and cost of the scoring procedure, making it practical to use constructed-response items in testing situations where human scoring would be impractical or prohibitively expensive.

2.6.1 Constructed-Response Items and IRT Models

Item Response Theory uses statistical techniques to model the association between a student's responses to test items and the underlying latent trait that is measured by the items. The accuracy of a test score (the estimation of the underlying ability) depends on how well the IRT model describes this association and fits the test data. Among the commonly used IRT models for constructed-response items are partial credit models with one- or two- item parameters. These models are additions of the IRT logistic models for multiple-choice items. The two-parameter partial credit (2PPC) model characterises an item with item discrimination and item score level difficulty parameters that differ by item score level. The 1PL model, also known as Masters' partial credit model (Masters, 1982), assumes the same discrimination of all items on a test. Figure 2 depicts two constructed-response items of three score levels (s = 0, s = 1, and s = 2) modeled by the 2PL model. Each curve models the relationship between the probability of getting the designated score level on the item and the student's ability level (0). The s = 0 curve, for instance, is the graphical demonstration of probability P (s = 0 | 0). These curves are regularly called item category characteristic curves (ICCC). The two items differ in item discrimination only. The item in the upper panel has a discrimination parameter value of 1.7 and the lower panel item has a smaller discrimination value of 0.75.

2.6.2. Item Information and Test Information Functions

Under item response, the interest is in the sampling variance of the person's underlying latent trait. To obtain an idea of how well an item and the entire instrument can estimate person trait locations, the item information and test information are examined. The information of an instrument depends on the items used as well as the ability of the examinees. Easy items can tell us very little about the examinee in the upper end, but provide adequate information on the examinee at the lower end. Equally, difficult items tell us very little about the examinee at the lower end, but provide information on the upper end about the examinee ability. As shown on the left plot in Figure 2.1:

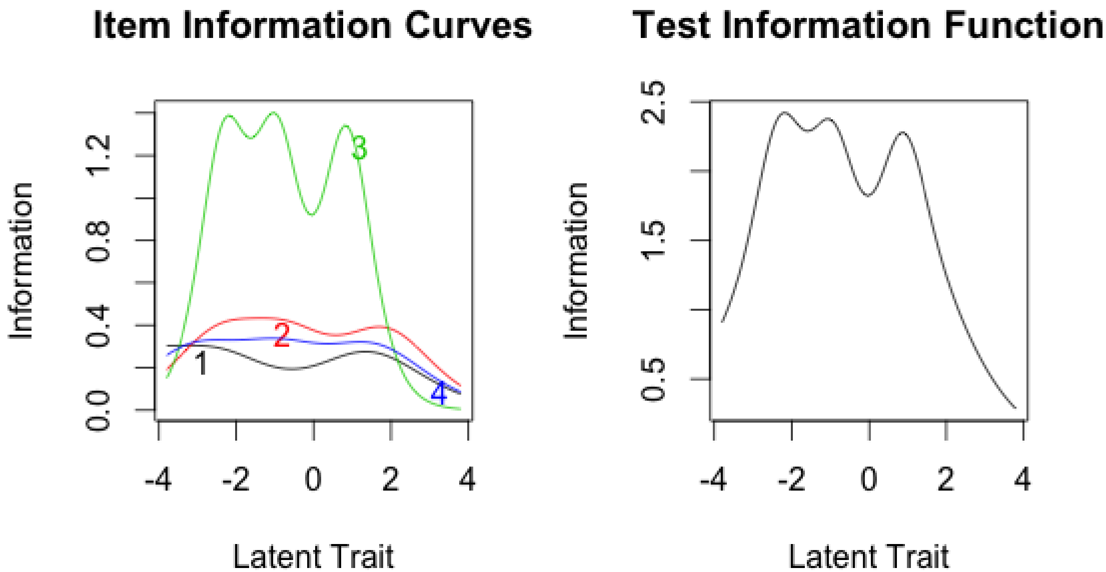


Figure 2.1: Items information cures and test information function of hypothetical items. Source: The University of North Carolina at Greensboro.

Left: Item information curves (IIC) for four items. Item 3 gives the most information (IIC3) about the latent trait of examinees compared to items 1, 2, and 3. Right: Test information curve. The curve resembles the shape of the IIC (3) since item 3 had the largest amount of information. Of which category to the item information are the product of the category information and the probability of the responses, thus, the total information yielded by the items in this example is the weighted sum of the information from each response category.

2.8 Differential Item Functioning

Most of the test is designed conducted and prepared for different commitments such as engagement of examinees in varied achievement marks, employment endowments and so on. The differential item functioning is regarded as the statistical process that is used in test item development to ensure that items are free from bias across the demographic characteristics of the students. It is conducted to examine how items function in various sub-groups in a population (Schipke, Knoll, Friederici and Oberecker ,2012). It refers to the situations in which the probability of responding correctly to an item for different groups of students after controlling their abilities is not equal. Differential item functioning occurs when an item's properties in one group are different from the item's properties in another group. An item is said to exhibit DIF if, after conditioning on any latent differences between groups, the item behaves inversely in the groups. This different behaviour is manifested in the item parameters (Oladele, Adegoke and Longjohn, 2020). In principle, an item exhibiting DIF has dissimilar item parameters depending on which group a respondent is in. For instance, DIF exists when a particular item has one difficulty level for males and a different difficulty level for females. In other words, the presence of differential item functioning means that a male and a female who have the same trait level have different probabilities of answering the item correctly.

The Differential Item Functioning is broadly used for detecting bias in tests and to state if a test is fair or not. Ogbebor and Onuka (2013) describe test fairness to be a test administered to a group of students with the same opportunity to exhibit knowledge. When a test has the element of DIF, it shows that the item or test is not fair to some of the students or subgroups. In this study, emphasis was on gender as a group. This is because of the generally held view that boys are better than girls in Mathematics. Studies such as this will help to establish if such differences are due to bias against girls in the construction of test items.

There are two different types of DIF, namely, uniform and non-uniform DIF. The uniform DIF occurs when there is no interaction between ability level (θ) and group membership. That is, the probability of answering the item rightly is uniform greater for one group than another group overall levels of ability. That is to say, for uniform DIF items only the difficulty parameter (b_i) is different between groups but the discrimination parameter (a_i) is the same. The non-uniform DIF happens when there is an interface between ability level (θ) and group membership. That is, the difference in the likelihoods of a right answer for two sets is not similar at all levels of ability. The non-uniform DIF is reflected by item characteristics curves that are not parallel (Mellenbergh, 1995).

In addition, item bias could be discovered through various analytic methods but the three most popular and widely used will be discussed in this study namely: Item Response Theory-Based Methods (IRTBM), Mantel-Haencel (MH) and Logistics Regression (LR). The item response theory offers a mathematical model that links performance on an item to exact feathers of the item (Difficulty, discrimination, pseudo-guessing) with characteristics of the examinees (typically ability on the unidimensional trait being measured). This mathematical function may take on a variety of forms, depending on the specific item response theory model. The Mantel-Haenszel method is particularly attractive in terms of implementation and having an associated test of significance with a small sample size (Swaminathan and Rogers, 1991, cited in Bichi, 2016). The Logistic Regression (LR) method is based on statistical modelling of the probability of responding appropriately to an item by group membership and a conditioning variable which is usually the scale or sub-scale total score (Zumbo, 1999).

2.8 Test Dimensionality

The process towards the acquisition of detailed correct knowledge of a test's internal arrangement is dimensionality. The item creators and consumers are provided with an enhanced understanding of how test marks make human capabilities solid through dimensionality assessment. Some outcomes from dimensionality assessment allow test authors and consumers to carefully validate detailed clarifications and usages of test scores (Zhang, 20016). A dimension is measured to be nontrivial if it is substantially related to more than five items (Stone and Yeh, 2006). Under the ML estimation, two additional model fit indices, the AkaikeInformation Criterion (Akaike, 1974) and the Bayesian Information Criterion (Schwarz, 1978), are described: a smaller value of either criterion indicates better model fit. Dimensionality also shares a close relationship with equating, a statistical procedure that enables comparability of scores on multiple forms of a test (Kolen and Brennan, 2014). Score reliability is interrelated with dimensionality. Dimensionality is a type of test score and is thus affected by the specific scoring method used to produce those scores. For instance, a special issue could be considered to score item responses.

The choice of the scoring process often affects how diverse dimensionality assessment methods act and what results they offer. The widespread use of a CR item, in contrast, entails students construct responses. Various forms of constructed-response tests have been seen in active tests, like short answers, essays, and speaking prompts. The constructed-response items are normally scored on an integer scale of 0 to 5, based on various prearranged scoring rubrics. Both manual and engine raters are involved in the scoring of constructed-response tests; for tests applying multiple raters, the number of raters and approaches for ratings varies across dissimilar tests. Test dimensionality is estimated in both exploratory and confirmatory ways (Reckase, 2009; Svetina and Levy, 2014). The dimensionality assessment in the exploratory method serves as an integrated part of a preliminary analysis before studying other psychometric procedures (for instance. MIRT equating, by Brossman and Lee, 2013).

2.9 Test Equating

Equating denotes the strictest form of establishing a translation between scores on two or more assessments, assuming that the tests are developed from the same test specifications (Muraki, Hombo and Lee, 2000). Consequently, equating can be considered as a process used to make test scores across different forms of the same test interchangeable. When test forms are created to be similar in content and difficulty, equating adjusts for differences in difficulty. Test score equating is usually adopted in comparing different test scores from different test forms (Kolen and Brennan, 2014; González and Wiberg, 2016). Test forms are considered to be the same, so scores on the two forms can be used interchangeably after equating has adjusted for changes in difficulty. The following situation is projected to develop further the idea of equating. Suppose that a student takes a college admissions test for the second time and earns a higher reported score than on the first testing. One explanation of this change is that the reported score on the second testing reflects a higher level of achievement than the reported score on the first testing.

Test equating is the statistical procedure that is used to establish comparability between tests built to the same content and specifications (Michaelides and Haertel, 2004). Under Item Response Theory, test equating is conducted in three different ways. They are Single group or common subject design; Random or equivalent group design and the common Item non-equivalent group designs (Kolen, 1988; Dorans, 2004). In this study because of its robustness and elegance of statistical analysis involved, emphasis was placed on a single group or common subject design. To compare scores from two different tests, a process of test score equating must be conducted. Through this process, scores from one test are converted into metrics of the other test. Then, there will be a comparison of both scores since the converted scores of the first test are now on the same scale as scores in the second one. Even though test score equating is done in CTT but IRT makes things simpler.

Based on item response procedures, the ability θ of students is invariant across different sets of items. Therefore, two students tested in dissimilar sets of items for which the item parameters are known will have student’s ability measured in the same scale and abilities compared without equating. However, suppose that the student had been administered the same test questions on both testing. Rather than indicating a higher level of achievement, the student's reported score on the second testing might be exaggerated because the student had already been exposed to the test items. Fortunately, a new test form is used each time a test is administered for most college admissions tests. Therefore, a student would not likely be administered the same test questions on any two test dates. The use of different test forms on different test dates might cause another problem, as is explained by the following conditions. Two students apply for the same school admission that is based partly on test scores. The two students take the test on different test dates, and Student one score a higher mark than Student two. One possible description of this change is that Student one is higher achieving than Student two.

But, if learner one took an easier test form than Learner two, then learner one would have a partial advantage over learner two. Then, the variance in scores might be due to changes in the difficulty of the test forms rather than in the levels of achievement of the learners. For this to resolve, equating is used with all tests in educational measurement. If the test forms are effectively equated, then the variance in equated scores for learner one and learner two is not attributable to learner one's taking an easier form. The process of equating is used in circumstances where such alternate forms of a test exist and scores earned on dissimilar forms are compared to each other. Even though test authors attempt to construct test forms that are as similar as possible to one another in content and statistical specifications, the forms usually differ somewhat in difficulty. Equating is intended to adjust for these difficulty variances, allowing the forms to be used interchangeably. Equating adjusts for variances in difficulty, not for differences in content. After successful equating, for instance, students who earn an equated score of, say, 56 on one test form could be considered, on average, to be at the same achievement level as students who got an equated score of 56 on a different test form.

2.9.1 Methods of Test Equating

The traditional equating approaches include mean equating, linear equating and equipercentile equating. Equipercentile equating is the most general among these methods and includes the first two methods (Angoff, 1971). Kernel equating (Lee and von Davier 2011) is a unified approach to test equating which comprises five steps. One, fitting appropriate statistical models to the raw data achieved by the data collection design (pre-smoothing). Two, estimation of the score's probabilities, Three, a continuation of the discrete distributions obtained at the previous step. Four, equating using the equipercentile method and Five, calculating the standard error of equating. Item response model equating (Lord, 1980) is a three-step process. In the first step, item parameters are estimated; in the second step, parameter estimates are scaled to abase IRT scale using a linear transformation in the third step equating is conducted by using different methods, for instance, the equipercentile equating, Linear equating and mean equating. The linear equating is less strict than equipercentile equating: in fact, while equipercentile equating requires that the scores on the two forms have the same distribution, in linear equating it is assumed that only the means and the standard deviations of the scores on the two forms are equal. From this, it is evident that linear equating is a subcase of equipercentile equating mean equating this method is the least strict of all the traditional methods: it only requires that the means of scores on the two forms A and B are equal. It is a particular case of equipercentile and linear equating.

2.9.2 Need for Test Equating

The needs for test equating include:

i. Test equating is needed because large-scale testing programmes often require multiple forms to maintain test security over time or to enable the measurement of change without repeating the identical questions.
ii. Agencies and decision-makers (for instance, the DSM) set scale-score cut points for diagnosis or certification these standards must be held constant over time and test forms Same test in different languages.
iii. The interchangeable use of alternate forms of tests must often be placed on a common scale – a process called test equating.
iv. To determine the extent that behavioural measures are to be used interchangeably, the outcome scores need to be equated or made comparable.
v. To compare test scores of different forms of tests (Strictly speaking, parallel tests) which measure the same latent trait?
vi. To construct the item bank/pool.
vii. For Computerised Adaptive Testing (CAT)

2.9.3 Requirements of Equating

Equating is distinguished from other forms of linking because of its goal which is to allow the scores from two tests form to be used interchangeably. Experience has shown that the scores and tests that produce the scores must satisfy very strong requirements to achieve this severe goal of interchangeability. Naturally, a set of tests are gathered to be rigorously similar with equal psychometric properties. Equating would then be needless. Practically, it is impossible to create multiple test form that is exactingly equivalent, and equating is necessary to adjust the test construction procedure. According to Holland, Pommerich and Dorans (2004), there are five important requirements for linking to be equating. These include:

i. The equal construct requirement: The two tests should both be measures of the same construct (latent trait, skill, ability).

ii. The equal reliability requirement: The two tests should have a similar level of reliability.

iii. The symmetry requirement: The equating conversion for representing the scores of Y to those of X should be the inverse of the equating conversion for representing the scores of X to those of Y.

iv. The equity requirement: It should be a matter of insignificance to students as to which of two tests form the students take.

v. The population invariance requirement: The equating function used to link the scores of X and Y should be the same irrespective of the choice of (sub) population from which it is derived.

2.10 Empirical Review

The recent improvement of the item response model takes its use from being an instrument set adopted completely by behavioural scientists and psychometricians to change to data analytical tools for an extensive adoption by statisticians. By 2000 the Item response study was supported by a new movement of investigators who not only extended the technical parts of the outline (estimation, model identification, and goodness of fit) but also advanced its computational features. Item response study throughout over fifty years was revealed and the growth in the total of software packages considered for analysing item response data from surveys or tests. Many item response profitable software were also produced such as MPLUS, MULTILOG, BILOG, Listrel, WINSTEPS, and HLM and currently IRT PRO and R

More significantly, several item response packages were developed in the open-source R software to estimate various item response frameworks and improved recognition. These included the packages LTM for unidimensional IRT (Rizopoulos, 2006), eRm for extended Rasch models (Mair and Hatzinger, 2007), MIRT for multilevel and Bayesian estimation of some IRT models (Fox, 2010) GPCM (Johnson and Kuhn, 2015) for Bayesian estimation of the generalised partial credit model, MCMC pack for Bayesian IRT (Martin, Quinn, and Park, 2010), and MIRT for multidimensional IRT (Chalmers, 2012). De Boeck and Wilson (2004) made use of the general statistics package lme4 and incorporated Rasch models under the generalised linear mixed model framework. This makes it possible to use SAS PROC NLMIXED for IRT.

2.10.1 Studies on the Calibration of Multiple-choice Items and Constructed-response Items

The appropriateness of calibrating multiple-choice items and constructed-response tests by scrutinizing the structure of the item response information from the model was conducted through Ercikan, Schwarz, Julian, Burket, Weber and Link (1998). Ercikan, et, al (1998) addressed the issue of not examining CR and SR tests. Thus, the residuals would reflect the violation of expectations because the deviations from the unidimensionality, local item dependence and goodness of fit introduce a methodical change in residuals. The loss of data due to the concurrent calibration conversed. It is noted that the constructed-response items offer unique information about the students' abilities. Thus, the concurrent calibration may cause the loss of information. The report shows that to assess the size of the forfeiture of information and the assessment of the results of the item, ability parameters and scores from separate and concurrent calibrations should be done.

From the literature, it was reported by Donoghue (1993) that polytomous items have information when the analysis is experiential than dichotomous. The findings of Grima and Weichun (202) in the scoring of a Mathematics test with mixed item forms and appraised six different scaling methods. They calibrated dichotomous items with the three-parameter model and polytomous items with the generalised partial credit model. The methods they employed included calibrating all items simultaneously, or calibrating components, which were defined on some basis, such as the item form, judgment of experts, or factor analyses result. The study shows that reliabilities coefficient results and correlation analyses comparing students results from different means were positive. Grima and Weichun, 2002 concluded that adjusting all items together resulted in the best fit to the model used. Alagoz (2005) adopted item response theory approaches to a Mathematics achievement test by merging constructed-response (CR) and multiple-choice (MC) forms of the item and established that the partial credit (PC) did not fit the data compared to the generalised partial credit (GPC) models. Also, test reliability is better when CR items are employed. The present study made use of a constructed-response Mathematics test item of two examination bodies. Other empirical studies compared the precision of the ability estimate (De Ayala, Dodd, Koch, 1997), calibrated dichotomous item exploring the two-parameter partial credit model and the three-parameter item response model using the generalised partial credit model (Erickan, Schwarz, Julian, Burket, Weber and Link, 1998). Nevertheless, this study used more than one polytomous IRT model to calibrate constructed-response items.

2.10.2 Studies on Software and Items Parameters

In a study, a comprehensive analysis of the methods and procedures for estimating the parameters of test items and students' ability levels was established by Baker and Kim (2004). Practically, the LOGISTTM programme as a method was implemented in that programme which is referred to as Joint Maximum Likelihood Estimation (JMLE) and was formulated by Birnbaum (1968). The estimation of θ (ability level) and item parameters was simultaneously carried out. Other popular programmes such as BILOG-MG 3TM (Zimowski, Muraki, Mislevy and Bock, 2003), MULTILOG V7TM (Thissen, Chen and Bock, 2003), XCALIBRE 4.2.2 (Assessment Systems Cooperation [ASC], 2014; Pido, 2012; Ojerinde, 2013) used the marginal maximum likelihood estimation (MMLE), and an expectation-maximisation (EM) algorithm. There is also a recent software developed called IRT PRO module software that can be used for estimation of item parameter and ability parameters of test items and examinee abilities.

This method estimated the items and parameters in sequential stages. The advantage is, that convergence can be extended with a fixed number of items devoid of calling upon an arbitrary prior ability distribution. The XCALIBRE programme is used to meet the needs of users requiring comprehensive 2 - parameters and 3-parameters item response theory (IRT), also for calibrating items, estimating item parameters. It is user friendly and produces a graphical user interface that makes it easy to run. It performs IRT item parameters calibration for polytomous IRT models.

However, dichotomous items were calibrated by employing the three-parameter item response model, and the two-parameter partial credit model (using the generalised partial credit model) by Ercikan, Schwarz, Julian, Burket, Weber and Link (1998). Also, De Ayala, Dodd, and Koch, (1992) use partial credit and graded response models and equated items in terms of the accuracy of the ability estimate under a computer adaptive testing setting. Under this study, the strength of the partial credit and graded response model-based ability estimation to the use of items, which did not fit these models, was also examined. The research employed the likelihood ratio statistic for the model-fit analysis. Results showed that the partial credit computer adaptive test provided reasonably accurate ability estimation despite adaptive tests, which on average contained up to 45% misfitting items.

Also, Schumacker, Si and Mount (2003) carried out research work on the ability estimation under different item parameterisation and scoring models and the result revealed that the different IRT models and scoring models yielded different ability estimates in combination with the factors investigated in the study. The study demonstrated that polytomous models have better accuracy in ability estimation, both in terms of higher recovery rates and lower RMSEs, in all combinations of the prior distribution and threshold configuration. It was reported dichotomous models had less information about students' ability by ignoring their differences in choosing different categories other than the most completed answer. The present study will make use of two ordinal polytomous models.

2.10.3 Studies on Differential Item Functioning

Enu (2014) calibrated Mathematics and Geography item banks for Joint Command Secondary Schools Promotion Examinations (JCSPE) of West African Examinations Council’s items with a view of ensuring that items in the bank are calibrated with high quality. A survey design was used with 600 junior secondary school students (JSSS). Past JCSPE items were pooled for four years. The IRT parameter approach was used under

Bilog-MG software. In addition, DIF analysis was conducted for locations, gender and mode of schooling. It was concluded that there are needs to involve item developer in their assessment. This present study sampled more than 600 students which make the study more robust and possessed more valid precision outcomes on student's ability estimation. Ogbebor (2017) conducted research work on the construction of the Mock Economics test for senior secondary school students in Delta State, Nigeria using classical test and item response theories to validate test and estimate test parameters. The study concluded that item response theory is more effective in the calibration of test items as it enhances selecting of items that best measure student's ability and also giving adequate information concerning the behaviour of an item as well as the students. Also, IRT was considered to be the best method for measuring DIF in a test thus, assisting in the aspect of identifying items that favour group membership about gender, school type and school location. This present study focused on the efficiency of the polytomous models in scoring student responses in constructed-response tests.

Metibemu (2015) compared CTT and IRT in the development, scoring and equating of senior secondary school Physics tests with a view of identifying which among the two methods of test theory best for the estimation of students' scores. A descriptive survey design was used with a 1423 sample size from 48 sampled secondary schools. After item analysis was conducted on the developed item, 50 multiple-choice items were compared with 50 multiple-choice items of 2014 WAEC. For the data analysis, Listrel and Bilog were used and Mante-Hanzel DIF shows the difference in gender and school location. Also, IRT and CTT frameworks were similar in producing DIF conducted on the test items and scores. IRT scoring method was seen to be better than CTT in the estimation of individual-based test performance. The present study was conducted on two examination bodies which make it an advanced study compare to Metimemu's own. Adegoke (2016) estimated the ability of students in Physics essay test under two IRT polytomous models and the results showed that GPCM was better in the estimation of students ability estimate. The GPCM gave a higher mean ability estimate than RPCM. The study used a small sample size of 123 students but this present study used a large sample of over 1,000 students.

2.10.4 Studies on Test Equating

Olonode (2017) equated and scale the multiple-choice Mathematics items of WAEC and NECO to determine their comparability. The study revealed that the Mathematics items were similar in function. This study is designed to do the same to constructed-response items of WAEC and NECO Mathematics respectively. Furthermore, Fakayode (2017) compared the relative effectiveness of the classical test and item response theories in equating West African Examinations Council Mathematics test scores for June and November 2015 to identify the best approach to equating that can be adopted between the classical test theory equating approaches and item response theory equating approaches. Multiple choice Mathematics items of two sets of WAEC examinations (June and November 2015) were used as instruments of the study. The study sampled senior secondary students three in Osun State. It was concluded that the items for the two examinations were unidimensional, locally independent and could be modelled using Item Characteristic Curve. Also, the linear and equipper-centile equating methods were found better in the CTT framework while mean/sigma was found to be the best method both under the CTT and the IRT framework. This present study is different because it focused on equating the Mathematics constructed-response tests of two examining bodies. That is, WAEC and NECO. This study also used IRT polytomous model to score students responses.

Daniel (2014) discovered that more items on NECO SSCE were at a higher level of difficulty and were generally less discriminatory than WASSCE Mathematics items while more WASSCE items were at the appropriate level of difficulty and had better discriminating indices. The above study was carried out on multiple-choice items while this study will be on constructed-response items. Also, Ajayi and Awogbemi (2012) conducted a correlation analysis of students' achievement in WAEC and NECO Mathematics from four selected secondary schools in Ifedayo, Osun State revealed that there is a significant positive relationship between the Mathematics performance of students WAEC and NECO. Bandele and Adewale (2013) reported that WAEC, NECO and NABTEB Mathematics achievement examinations are highly reliable and valid and as well equal and comparable. In the aforementioned study, only multiple-choice items were used but this present study further confirms if WAEC and NECO items were similar or comparable using only Mathematics constructed-response test items of the two examination bodies.

2.11 Appraisal of Literature and Gaps Filled

From the literature, few studies have reported the use of Item Response Theory (IRT) for designing, analysing, scoring, and comparing tests and similar instruments whose purpose is to measure unobservable characteristics of the respondents. Over time, it seems that Classical Test Theory (CTT) was being used for the scoring of both dichotomous and polytomous items.

Most of the past studies on IRT in sub-Saharan Africa have concentrated on dichotomous items and IRT logistic models. But this study concentrated on the polytomous items scoring with IRT models. Indeed, many studied had been conducted on the performance of students in the internal and external examination but very few among these research works were conducted on Mathematics constructed-response test items. Hence, this research evaluated the performance of students in Mathematics constructed-response tests in pursuance of vision 2020 as stipulated by the Nigerian Government for effective possession of mathematical ability.

Moreover, in this study, senior secondary school student's ability in Mathematics of the two prominent public examining bodies in Nigeria were compared to determine the level at which the examining bodies test items concerning Mathematics are equivalent and parallel. In addition, the item parameters of test items were established using IRT polytomous framework and the extent to which variants of the models affect the scoring of students' responses to constructed-response test items were examined. The fairness of the public examination test to gender norm was examined to determine the differential item functioning of the test items.

Results of studies such as this would not only help in finding out the psychometric properties of constructed-response items but help in familiarising education stakeholders, evaluators and public examining bodies such as WAEC, NECO and NABTEB with the techniques of establishing items parameters and validly and reliably scoring of the test items for quality attainment. In addition, knowledge of and clues on how to score polytomous items using the IRT polytomous models was provided. The dimensionality of constructed-response test items of the public examination was confirmed to ascertain the number of traits being measured. The findings of the study, therefore, enlighten prospective students who prepare to write WAEC and NECO on the traits to possess to score high in Mathematics.

CHAPTER THREE

METHODOLOGY

This chapter presents the research design, target population, sampling techniques and sample, instrumentation, data collection procedures, and data analysis frameworks adopted for this study.

3.1 Research Design

The study adopted a non-experimental design of descriptive survey research type. This design was adopted because the researcher had no control over the variables of the study but only investigated the relationship between variables of the study without any form of variable manipulation. Single group design was employed in the administration of the tests. In the single group design, all students in a single sample of students from the population take both tests. von Davier, Holland and Thayer (2004) described single group design as a design that caters for any likelihood of differential students' ability by having similar students write the two test forms and provide accurate equating results.

3.2 Population

The target population for this study consists of all senior secondary three (SS3) students in Ibadan Metropolis, Oyo State, Nigeria. The choice of senior secondary three (SS3) students was because they are the category of students that are officially permitted to take SSSCE May/June of West African Examinations Council (WAEC) and SSSCE June/July of National Examinations Council (NECO). Also, at the SS3 level, the students have covered significant sections of the syllabus to prepare them for these examinations. In Ibadan Metropolis, official data from the Ministry of Education indicates that there are 96 public senior secondary schools. In all, the total population of SS3 students in Ibadan Metropolis in 2018/2019 session was about 10,000(boys - 5,500) and (girls - 4,500).

3.3 Sampling Techniques and Sample

The multi-stage sampling procedure was used to select the target participants for the study. First, a simple random sample procedure was used to select one educational zone out of the two educational zones in Ibadan, Oyo State. This educational zone is called Ibadan Metropolis which has five Local Government Areas. These are Ibadan North, Ibadan North East, Ibadan North West, Ibadan South-West and Ibadan South East. All the five local government areas were sampled. In all the five LGAs, there are 96 co-educational public senior secondary schools. Second, 25% of the 96 public senior secondary schools were sampled using a proportionate to size sampling method to give a total of 24 schools. Third, in each school, two intact SS3 classes were randomly selected. This gave a total of 48 SS3 classes. The total number of students who partook in the study was 1,151. The distribution of the number of students sampled in the study across the schools is presented in Table 3.1.

Table 3.1: Distribution of the Number of Students Sampled

S/N	Selected LGAs	No of secondary schools in the LGAs	No of selected public secondary schools	No of students sampled
1	Ibadan North	22	6	343
2	Ibadan North East	21	5	242
3	Ibadan North West	19	5	233
4	Ibadan South East	16	4	177
5	Ibadan South West	18	4	156
	Total	**96**	**24**	**1151**

Source: Field survey 2019

3.4 Instrumentation

The instruments for this study were:

i. WAEC Mathematics Constructed-Response Achievement Test
ii. NECO Mathematics Constructed-Response Achievement Test

3.4.1 WAEC Mathematics Constructed-Response Achievement Test (WAEC_MCRAT)

The WAEC_MCRAT consists of 15 adopted CR test items. The WAEC May/June mode of Mathematics question papers for three years (2013, 2014 and 2015) was collected. The five compulsory questions under section A of Mathematics Paper-II for the three years formed the 15-items that each student answered. The test items reflected the themes and topics in the syllabus of WAEC and the teaching curriculum of the Nigerian Educational Research and Development Council (NERDC). The themes and topics were:

i. Number and Numeration; number base system, arithmetic, indices, logarithms, sets, approximations, sequence and series, surds, matrices and determinant.
ii. Algebraic Process; simple equation and variation, quadratic equation, logical reasoning, gradient of a curve, linear inequalities and algebraic fractions.
iii. Geometry; constructions, proof of some basic theories, trigonometric ratios, mensuration, chord property, circle theorem, bearings, trigonometry graphs of trigonometric ratio, surface area and volume of a sphere, longitude and latitude, coordinates geometry of straight lines
iv. Statistics; data presentation, measures of central tendency, measures of dispersion, histograms of grouped data, cumulative frequency curve and probability.

Table 3.2: Test Blue Print of WAEC Mathematics Constructed-Response Test Items

Themes	Knowledge	Comprehension	Thinking	Total	Percentage
Number and Numeration			1, 2, 3 (3)	3	20%
Algebraic Process	4, 5 (2)			2	13%
Geometry		6, 11, 12, (3)	7, 8, 9, 10 (4)	7	47%
Statistics	13 (1)	14 (1)	15 (1)	3	20%
Total				**15**	**100%**
Percentage	**20%**	**33%**	**47%**	**100%**	

Table 3.2 presents the test blueprint. This is to ensure the content validity of the test. The tests covered four major themes in the syllabus; number and numeration, algebra, geometry and statistics. The representation of contents is 20%, 13%, 47% and 20% respectively.

3.4.2 NECO Mathematics Constructed-Response Achievement Test (NECO_MCRAT)

The NECO_MCRAT consists of 15 adopted CR test items. The NECO June/July mode of mathematics question papers for three years (2013, 2014 and 2015) was collected. The five compulsory items under section A of Mathematics Paper-II for the three years formed the 15 items that each student answered. The test items reflected the themes and topics in the syllabus of NECO and the teaching curriculum of the Nigerian Educational Research and Development Council (NERDC). The themes and topics were:

ii. Number and numeration; number base system, arithmetic, indices, logarithms, sets, approximations, sequence and series, surds, matrices and determinant.

iii. Algebraic process; simple equation and variation, quadratic equation, logical reasoning, gradient of a curve, linear inequalities and algebraic fractions.

iv. Geometry; constructions, proof of some basic theories, trigonometric ratios, mensuration, chord property, circle theorem, bearings, trigonometry graphs of trigonometric ratio, surface area and volume of a sphere, longitude and latitude, coordinates geometry of straight lines

v. Statistics; data presentation, measures of central tendency, measures of dispersion, histograms of grouped data**,** cumulative frequency curve and probability

Table 3.3: Test Blue Print of NECO Mathematics Constructed-Response Test Items

Themes	Knowledge	Comprehension	Thinking	Total	%
Number and Numeration		4 (1)	1, 2, 3, 5, 6(5)	6	40%
Algebraic Process			7, 8, 9 (3)	3	20%
Geometry		10, 11(2)	12, 13 (2)	4	27%
Statistics		15 (1)	14 (1)	2	13%
Total				**15**	**100%**
Percentage	**3%**	**20%**	**77%**	**100%**	

Table 3.3 presents the test blueprint. This is to ensure the content validity of the test. The tests covered four major themes in the syllabus; number and numeration, algebra, geometry and statistics. The representation of contents is 40%, 20%, 27% and 13% respectively.

3.4.3 Validity and Reliability of Instruments

The instruments were subjected to content and face validity with the support of the researcher's supervisor. For the reliability of the instruments, the Public examination bodies were assumed to have established the reliability for the test items. Nevertheless, in this study, the reliability coefficients of WAEC_MCRAT and NECO_MCRAT were established using the classical test and item response models. The results of reliability coefficient obtained under each test models include: (WAEC_CTT_rel = 0.47), (WAEC_IRT (GPCM)_rel=0.70, IRT (GRM)_rel = 0.72), and (NECO_CTT_rel= 0.59), (NECO_ IRT (GPCM)_rel = 0.70, IRT (GRM) _rel = 0.71).

3.4.4 Scoring of the Instruments Items

Each item was scored over eight. Two types of scoring were adopted. These were: The point score scoring model of CTT and the polytomous scoring module of IRT Software. For the point-scoring mode of CTT, the score on each item was summed up to give a composite score. For the IRT, the difficulty levels of each item were taken into consideration and the scores obtained in each item by the students ranged from 0, 1, 2, 3, 4, 5, 6, 7 and 8. This shows that for each item, there were eight location boundaries. The location boundaries are designated as b*i* where b represents the difficulty level and *I* represent the category. Also, the discrimination value of *ai* was estimated. (See pages 94 to 99). For comparison of CTT and IRT, the students' ability estimates were obtained through the conversion of. Z-score to T-score using the formula:

$T = 10z + 50$

Where z is the standard score

3.5 Data Collection Procedure

The selected schools were visited to seek permission from the school authorities (the principals and Mathematics teachers) and a letter of introduction collected from the

Institute of Education was presented. The students were informed about the tests and informed consent was sought. The tests were administered during the second term of the 2018/2019 session; this was to ensure that the mathematics syllabus for the senior secondary school mathematics was covered to a large extent. Different dates were fixed for the tests in all the participating schools.

Further, the assistance of the 24 research assistants was sought during data collection. One research assistant was assigned to a school. The research assistants were different from the Mathematics teacher in the school. This was to ensure that there was no interference in the administration and answering of the test items. For, the test administration, a single group design with elements of counterbalance procedure was adopted. There were two administrations of tests.

In the first administration, 15 items were administered. This consists of 8 WAEC_MCRAT and 7 NECO_MCRAT. The items were alternated. For this administration, two hours 30 minutes was allotted. In the second administration, 15 items were administered. This consists of 8 NECO_MCRAT and 7 WAEC_MCRAT. The items were alternated for this administration, 2 hours and 30 minutes was allotted. The researcher developed a marking guide through adopting WAEC and NECO frameworks to score the answer script of students through CTT and IRT polytomous frameworks.

3.6 Method of Data Analysis

The data collected for this study were analysed using mean, standard deviation, t-test, linear equating, Pearson product-moment correlation, models of Classical Test Theory ((Factor analysis and Monte Carlo parallel analysis) via IBM-SPSS version 25, Models of Item Response Theory (Graded Response Model and Generalised Partial Credit Model) via Module of IRT PRO 3, DIFAS and R 3.6.2 software. All analyses were determined based on 0.05 ($p<0.05$) significance level and graphs were produced where necessary. The details of the analytical procedures for each question are presented in Table 3.5.

Table 3.5: Showing Method of Data Analytical Procedure

Research Questions	Method of Analysis
1	Item Analysis using CTT (Factor Analysis) and IRT Polytomous models (GRM and GPCM),
2	IRT Polytomous models (GRM and GPCM)
3	Paired t-test,
4	Linear equating and Pearson Correlation
5	DIF using IRT-log ratio methods in IRT PRO3 and Mantel-Haenszel in DIFAS

3.7. Methodological Challenges

One of the methodological challenges encountered in this study was that the researcher planned to use one research assistant for each selected 24 schools. At the beginning of the exercise, four research assistants declined but this was overcome by using four students teacher in the affected schools. Some schools disturbed the process of the test administration; hence the research activities for some time were truncated. They complained of the test length (of 15 constructed-response items) that took 2 hours to complete which could rob part of the student's time of studying. Also, at the point of instruments administration, some students refused to participate in the test because they disliked the subject but they were later persuaded by the researcher and research assistants. The software for analysis was not readily available coupled with the high price of obtaining this from the internet since some were not free for download. The researcher had to make enquiries on how to get the software and as well, contacted people around for manuals. Efforts were also made to study how to use the software package effectively in the analysis of the collected data.

CHAPTER FOUR

RESULTS AND DISCUSSION

This chapter presents the results of data analysis and a discussion of the findings of the study. The results are presented based on the research questions stated in chapter one.

4.1. Socio-demographic Data of Students

First, the socio-demographic data analyses of the respondents are presented as follow

Table 4:1: Socio-Demographic Data of Students

Students Characteristics	Number	Percentage
Gender		
Male	565	49.1
Female	586	50.9
Total	**1151**	**100**
Age		
12-16	537	46.7
17-19	603	52.4
20-22	11	1.0
Total	**1151**	**100**

Table 4.1 presents the gender and age distribution of students wherein 49.1 % of the respondents were males while 50.9 % were female. This indicates that on average there were more female students than the male students in the senior secondary schools sampled. Also, the table shows that 46.7% of the ages of respondents were between 12 and 16 years, 52.4% were between 17 and 19 years while 1.0% was between 20 and 22 years old respectively. This indicates that on average the age distributions of students in the senior secondary schools aligned with the age range stipulated in the 6 3 3 4 education system documents of the sampled city.

The research questions raised in this study were answered as follows:

4.1.1 Research Question One: How many dimensions do the 15-item of WAEC and NECO Mathematics constructed-response have under:

(a) Classical Test Theory (CTT) framework?

(b) Item Response Theory (IRT) framework vis a vis:

i. Graded Response Model (GRM)?

ii Generalised Partial Credit Model (GPCM)?

To answer research question 1a, exploratory factor analysis was conducted under the CTT framework. The KMO and Bartlett's Test of Sphericity was carried out to establish test normality and sample adequacy. The results of the analysis for WAEC Mathematics constructed-response tests are presented in Table 4.2

Table 4.2 KMO and Bartlett's Test Statistics of 15 WAEC Mathematics Constructed Response Achievement Test Items

	Criterion	**Value**
Bartlett's Test of Sphericity	KMO	0.65
	Approx. Chi-Square	2209.52
	df	105
	p-value	0.00

Table 4.2 shows that the statistics obtained for KMO sampling adequacy was 0.65, which was relatively good. The statistics of Chi-Square for Bartlett's Test of Sphericity was significant at p-value < 0.05. This implies that the test data was adequate and followed a normal distribution. Thus, exploratory factor analysis was performed on test data to determine the number of dimensions. To verify the number of dimensions of the WAEC Mathematics constructed-response test items, the Explanatory Factor Analysis (EFA) was employed using SPSS version 25. Table 4.3 presents the statistics of the dimensionalities of the WAEC constructed-response items.

Table 4.3: Total Variance (WAEC Mathematics Constructed-response Achievement Test)

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.55	16.99	16.99	2.55	16.99	16.99
2	1.77	11.82	28.81	1.77	11.82	28.81
3	1.54	10.26	39.08	1.54	10.26	39.08
4	1.24	8.26	47.34	1.24	8.26	47.34
5	1.10	7.32	54.65	1.10	7.32	54.65
6	1.01	6.73	61.38	1.01	6.73	61.38
7	0.86	5.74	67.13			
8	0.80	5.34	72.46			
9	0.76	5.03	77.50			
10	0.74	4.92	82.42			
11	0.63	4.18	86.60			
12	0.54	3.62	90.22			
13	0.54	3.57	93.79			
14	0.49	3.25	97.04			
15	0.44	2.96	100.00			

The Total variance explained output was used, as revealed from Table 4.3, the highest eigenvalue was 2.55 for component one. This indicated that the highest component explained was 16.99% with an eigenvalue of 2.55. The acceptable rule is that extracted factors when put together explaining 50% to 60% of the variance with eigenvalues greater than one should be kept as good extracted values. This implies that the test data has six underlying factors. The six factors showed that the test data has more than one dimension, and then the test data may be considered approximately multidimensional. This indicates that the WAEC Mathematics constructed-response item is multidimensional with six dimensions. In addition, the scree plot was also conducted to further confirm the dimensionality of the test.

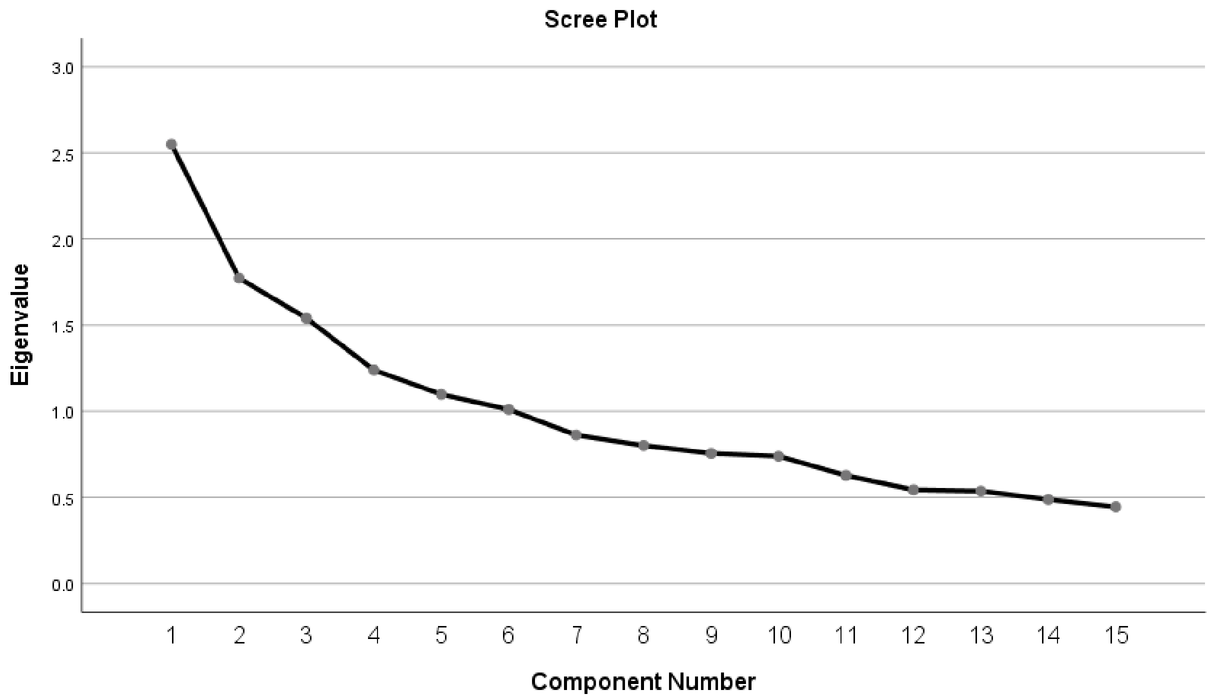


Figure 4.1: Scree plot of WAEC Mathematics constructed-response test items

From Figure 4.1 it could be deduced that the components in the y-axis, go down towards the X-axis. This downward slope is shown with the dots in terms of their contribution to the variance. Each space between two dots means a factor. The acceptable rule is to retain all components or factors within the sharp descent before the Eigenvalues trail off (Adegoke, 2013). Hence, the number of factors underlying the test data is more than one which points to multidimensionality. In addition, further analysis was carried out to substantiate the dimensionality of the test data. Parallel analysis was conducted using Monte Carlo Principal Component Analysis for parallel analysis software. Parallel analysis requires that a set of random correlation matrices be generated based upon the same number of variables and respondents as the experimental data. These random correlation matrices are then subjected to principal components analysis and the average of their eigenvalues is computed and compared to the eigenvalues produced by the experimental data (Watkins, 2005). This is hereby presented in Table 4.4.

Table 4.4: Monte Carlo PCA, Parallel Analysis Statistics for WAEC Mathematics

	Real Data Eigenvalues	Randomly Generated Data Eigenvalues	Standard Deviation
1	2.55*	1.20	0.023
2	1.77*	1.27	0.018
3	1.54*	1.16	0.014
4	1.24*	1.21	0.012
5	1.10	1.12	0.012
6	1.01	1.17	0.010
7	0.86	1.09	0.009
8	0.80	1.12	0.011
9	0.76	1.07	0.011
10	0.74	1.09	0.012
11	0.63	1.04	0.011
12	0.54	1.05	0.011
13	0.54	1.02	0.014
14	0.49	1.02	0.014
15	0.44	1.00	0.819

*Suggested number of dimensions: 4

Table 4.4 shows the comparison of the eigenvalues (experimental and generated data) while there were four components of the real data set with eigenvalues (2.55, 1.77, 1.54, and 1.24) greater than the eigenvalues (1.31, 1.27, 1.24, and 1.21) of the generated data set respectively. The result implies that there are likely four factors that underlie the WAEC constructed-response items. This also suggests that WAEC constructed-response items are multidimensional with a minimum of four dimensions and consequently measured four traits.

Furthermore, exploratory factor analysis was conducted again, based on the minimum number of dimensions suggested by the Parallel Analysis Statistics (PAS) to identify items that measure the objectives of WAEC Mathematics. The results of the analysis are presented in Table 4.5.

Table 4.5: Rotated Factor Matrix of WAEC Constructed-response Items

	Factor			
Item	**1**	**2**	**3**	4
1	**0.547***	**0.408***	0.138	0.078
2	**0.651***	-0.01	-0.041	-0.390
3	**0.655***	-0.149	-0.23	-0.131
4	**0.703***	0.057	0.05	0.132
5	**0.511***	0.348	0.228	-0.245
6	-0.021	0.114	**0.805***	-0.016
7	0.117	0.342	-0.095	**-0.490***
8	-0.007	0.017	**-0.748***	-0.075
9	-0.074	**0.588***	0.007	0.075
10	-0.185	**0.562***	-0.263	-0.226
11	0.115	**0.714***	0.078	-0.038
12	0.179	**0.457***	0.136	0.015
13	**0.442***	-0.019	-0.272	**0.545***
14	-0.051	-0.211	0.189	0.202
15	-0.129	0.162	0.090	**0.732***

*Absolute loadings values > 0.4

Table 4.5 shows the four factors and item loading on them. Varimax rotation method with Kaiser normalisation which is an Orthogonal rotation technique was applied. Thus, Table 4.4 shows that items 1, 2, 3, 4, 5, and 13 highly loaded on factor one, Items 1, 9, 10, 11, and 12 highly loaded on factor two, items 6 and 8 highly loaded on factor three while items 7, 13 and 15 highly loaded on factor four respectively. However, it was only item 14 that did not load highly on any factor. This result shows that four substantial factors underlie WAEC constructed-response tests. Furthermore, the result shows that items 1 and 13 loaded highly on more than one factor. For instance, item 1 loaded highly on factors one and two while item 13 loaded highly on factors one and four respectively. This result suggests that WAEC constructed-response test items are multidimensional, that is, four abilities accounted for the observed variation in the performance of candidates in the test.

Similarly, this suggests that the WAEC constructed-response test items measure more than one trait. Hence, there is a need for the students to possess more than one trait to provide the correct answer to the items. These mathematics abilities include factor one (mathematical competency and computational skills), factor two (understanding of mathematical concepts and their relationship to the acquisition of entrepreneurial skills for everyday living in the global world), factor three (translate problems into mathematical language and solve them using appropriate methods) and factor four (accurate to a degree relevant to the problem at hand and logical, abstract and precise thinking).

Furthermore, Item Response Theory Models (Polytomous Graded Response Model and Generalized Partial Credit Model of IRT-PRO Version 3.0 were used to determine the minimum level of the dimension of the WAEC Mathematics constructed-response items. The results of the analysis are presented in Table 4.6.

Table 4.6: Dimensionality of WAEC Mathematics Construction Response Test Using IRT (Generalised Partial Credit Model)

Dimension	**Loglikelihood**	**Difference**	**Ratio of difference**	**Remark**
1	48450.68			
2	47982.05	468.63		
3	47682.34	299.71	0.64*****	Dimension limit

*****Ratio of difference <1

Table 4.6 shows the minimum dimensionality of the level of WAEC test items using the generalised partial credit model (GPCM) as 3. This was determined based on the differences obtained when loglikelihood values were compared showing the ratio of difference (0.64) to be less than 1.

Table 4.7: Dimensionality of WAEC Mathematics Construction Response Test Using IRT (Graded Response Model)

Dimension	**Loglikelihood**	**Difference**	**Ratio of difference**	**Remark**
1	48411.17			
2	47835.61	755.56		
3	47527.63	245.61	0.33*	Dimension limit

*****Ratio of difference <1

Table 4.7 shows the minimum dimensionality of the level of WAEC test items using the graded response model (GRM) as 3. This was determined based on the differences obtained when loglikelihood values were compared showing the ratio of difference (0.33) to be less than 1.

For the NECO test, the KMO and Bartlett's Test of Sphericity was carried out to establish test normality and sample adequacy. The results of the analysis for NECO Mathematics constructed tests are presented in Table 4.7

Table 4.8 KMO and Bartlett's Test Statistics of NECO Mathematics Constructed Achievement Test

	Criterion	Value
	KMO	0.72
Bartlett's Test of Sphericity	Approx. Chi-Square	2215.40
	Df	105
	p-value	0.00

Table 4.7 shows that the statistics obtained for KMO sampling adequacy was 0.72, which was relatively good. The statistics of Chi-Square for Bartlett's Test of Sphericity was significant at p-value < 0.05. This implies that test data was adequate and followed a normal distribution. Thus, exploratory factor analysis was performed on test data to determine the number of dimensions. To verify the number of dimensions of the NECO Mathematics constructed-response test items, the Explanatory Factor Analysis (EFA) was employed using SPSS version 25.

Table 4.9: Total Variance Explained (NECO Mathematics Constructed-response Achievement Test)

Component	Initial Eigen values			Extraction Sums of Squared Loadings		
S/N	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.80	18.69	18.69	2.80	18.69	18.69
2	1.85	12.35	31.04	1.85	12.35	31.04
3	1.41	9.40	40.44	1.41	9.40	40.44
4	1.07	7.15	47.59	1.07	7.15	47.59
5	1.03	6.84	54.43	1.03	6.84	54.43
6	0.92	6.14	60.57			
7	0.84	5.60	66.17			
8	0.77	5.15	71.32			
9	0.76	5.06	76.38			
10	0.71	4.72	81.10			
11	0.69	4.57	85.67			
12	0.62	4.15	89.82			
13	0.55	3.66	93.48			
14	0.52	3.49	96.97			
15	0.45	3.03	100.00			

Table 4.9 presents the statistics of the dimensionalities of the NECO constructed-response items. The total variance explained output was used, as revealed from Table 4.8, where the total variance explained by the highest eigenvalue was 2.80 for component one. This indicated that the highest component explained was 18.69% with an eigenvalue of 2.80. The acceptable rule is that extracted factors when put together explaining 50% to 60% of the variance with eigenvalues greater than one should be kept as good extracted values. This implies that the test data has five underlying factors. The five factors showed that the test data has more than one dimension, thus, the test data may be considered approximately multidimensional. This indicates that the NECO Mathematics constructed-response item is multidimensional with five dimensions. In addition, the scree plot was also conducted to further confirm the dimensionality of the test.

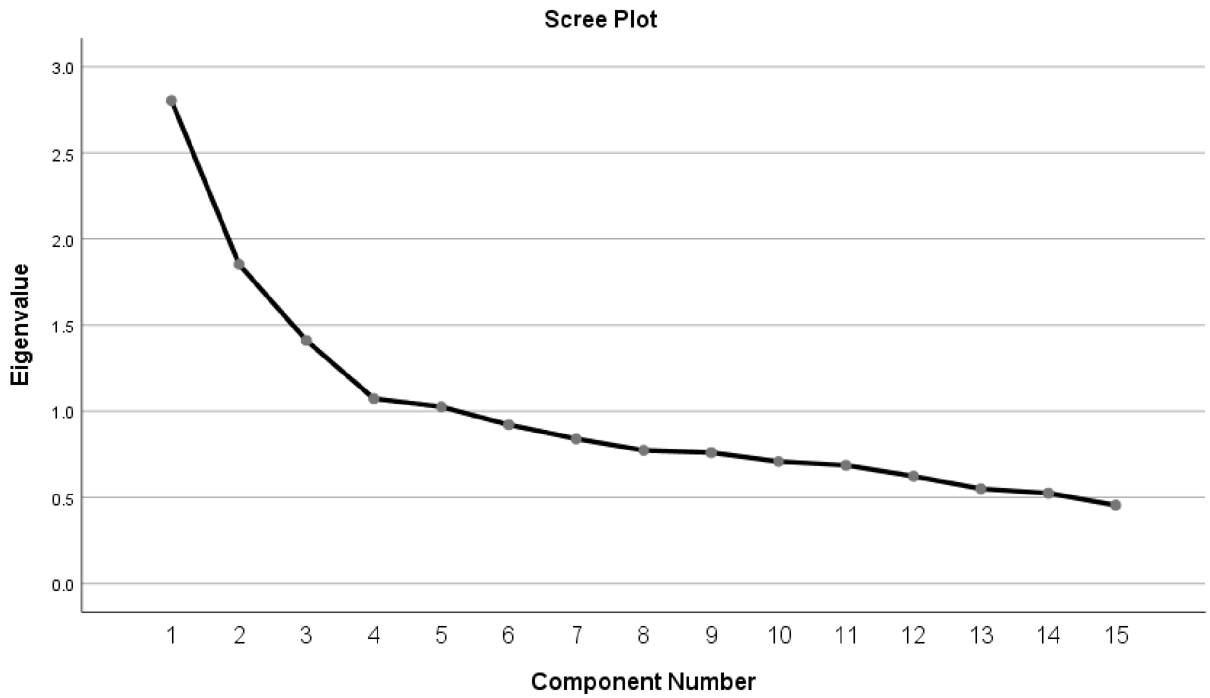


Figure 4.2: Scree plot of NECO mathematics constructed-response test items

From Figure 4.2 it could be deduced that the components in the y-axis, go down towards the X-axis. This downward slope is shown with the dots in terms of their contribution to the variance. Each space between two dots means a factor. The acceptable rule is to retain all components or factors within the sharp descent before the Eigenvalues trail off (Adegoke, 2013). Hence, the number of factors underlying the test data is more than one which points to multidimensionality. In addition, further analysis was carried out to substantiate the dimensionality of the test data. Parallel analysis was conducted using Monte Carlo Principal Component Analysis for parallel analysis software. Parallel analysis requires that a set of random correlation matrices be generated based upon the same number of variables and respondents as the experimental data. These random correlation matrices are then subjected to principal components analysis and the average of their Eigenvalues is computed and compared to the eigenvalues produced by the experimental data (Watkins, 2005). The results of the analysis are presented in Table 4.9.

Table 4.10: Monte Carlo PCA, Parallel Analysis Statistics for NECO Mathematics

Items	Real Data Eigenvalues	Randomly Generated Data Eigenvalues	Standard Deviation
1	2.80*	1.20	0.023
2	1.85*	1.27	0.018
3	1.41*	1.16	0.014
4	1.07	1.21	0.012
5	1.03	1.12	0.012
6	0.92	1.17	0.010
7	0.84	1.09	0.009
8	0.77	1.12	0.011
9	0.76	1.07	0.011
10	0.71	1.09	0.012
11	0.69	1.04	0.011
12	0.62	1.05	0.011
13	0.55	1.02	0.014
14	0.52	1.02	0.014
15	0.45	1.00	0.819

*Suggested number of dimensions: 3

Table 4.10 shows the comparison of the eigenvalues (experimental and generated data) while there were four components of the real data set with eigenvalues (2.80, 1.85, and 1.41) greater than the eigenvalues (1.32, 1.28 and 1.25) of the generated data set respectively. The result implies that there are likely three factors that underlie the NECO constructed-response items. This also suggests that NECO constructed-response items are multidimensional with a minimum of three dimensions and consequently measured three traits.

Furthermore, exploratory factor analysis was conducted again, based on the minimum number of dimensions suggested by the Parallel Analysis Statistics (PAS) to identify items that measure the objectives of WAEC Mathematics. The results of the analysis are presented in Table 4.11.

Table 4.11: Rotated Factor Matrix of NECO Constructed-response Items

	Factor		
Item	**1**	**2**	**3**
1	0.388	0.124	0.052
2	**0.664***	-0.074	-0.254
3	**0.627***	-0.217	-0.278
4	**0.467***	0.175	-0.224
5	**0.519***	0.057	0.222
6	0.250	**0.586***	0.026
7	**0.483***	-0.140	-0.109
8	0.089	**0.515***	-0.046
9	**0.584***	0.353	0.100
10	0.316	**0.543***	0.154
11	**0.562***	0.210	0.264
12	-0.22	**0.527***	-0.278
13	-0.177	**0.534***	0.125
14	0.095	-0.066	**0.785***
15	-0.224	0.095	**0.743***

*****absolute loadings values > 0.4

Table 4.11 shows the three factors and item loading on them. Varimax rotation method with Kaiser normalisation which is an Orthogonal rotation technique was applied. Thus, Table 4.10 shows that items 2, 3, 4, 5, 7, 9 and 11 were highly loaded on factor one, Items 6, 8, 12 and 13 were highly loaded on factor two, while items 14 and 15 were highly loaded on factor three respectively. However, it was the only item1 that did not load highly on any factor. This result shows that three substantial factors underlie WAEC constructed-response tests. This result suggests that WAEC constructed-response tests are multidimensional, that is, three abilities accounted for the observed variation in the performance of candidates in the test.

Similarly, this suggests that the WAEC constructed-response test items measure more than one trait. Hence, there is a need for the students to possess more than one trait to provide the correct answer to the items. These Mathematics abilities include factor one (mathematical competency and computational skills), factor two (understanding of mathematical concepts and their relationship to the acquisition of entrepreneurial skills for everyday living in the global world), factor three (translate problems into mathematical language and solve them using appropriate methods).

Furthermore, Item Response Theory Models (Polytomous Graded Response Model and Generalized Partial Credit Model of IRT-PRO Version 3.0 were used to determine the minimum level of the dimension of both NECO Mathematics constructed-response items. The results of the analysis are presented in Table 4.12

Table 4.12: Dimensionality of NECO Mathematics Construction Response Test Using IRT (Generalised Partial Credit Model)

Dimension	Loglikelihood	Difference	Ratio of Difference	Remark
1	51260.94			
2	50706.25	554.69		
3	50458.34	247.91	0.45*	Dimension limit

*Ratio of difference < 1

Table 4.12 shows the minimum dimensionality of the level of NECO test items using the generalised partial credit model (GPCM) as 3. This was determined based on the differences obtained when loglikelihood values were compared showing the ratio of difference (0.45) to be less than 1.

Table 4.13: Dimensionality of NECO Mathematics Construction Response Test Using IRT (Graded Response Model)

Dimension	Loglikelihood	Difference	Ratio of difference	Remark
1	51307.31			
2	47835.61	347.17		
3	47527.66	307.95	0.89	Dimension limit

*Ratio of difference < 1

Table 4.13 shows the minimum dimensionality of the level of NECO test items using the graded response model (GRM) as 3. This was determined based on the differences obtained when loglikelihood values were compared showing the ratio of difference (0.89) to be less than 1.

The findings of this study showed that WAEC Mathematics constructed-response test items have four dimensions while NECO Mathematics constructed-response test items have 3 dimensions based on the calibration with classical test models. On the other hand, it showed that WAEC and NECO Mathematics constructed-response test items have 3 dimensions respectively based on the calibration with Item response models. This implies that both test items are multidimensional and measures more than one trait. This is in line with the fact that CR tests require the students to demonstrate many abilities to be able to solve such mathematical problems that are couched in essay form.

More importantly, WAEC and NECO syllabi show that students must be able to demonstrate critical thinking skill, the ability to manipulate data and the ability to properly presents their work logically rather than just picking their answer from provided options in multiple-choice tests. The result supported Alu and Adediwura (2019) that conducted a dimensionality test on 2015 and 2016 NECO Mathematics tests and concluded that the test items were multidimensional. This result also corroborated the findings of Metibemu (2020) that confirmed the WAEC Physics paper 1 having more than one dimension and measuring more than one trait.

However, this result contradicts the findings of Ayanwale (2019) that established the 2015 NECO Mathematics constructed-response items to be unidimensional based on the framework used in establishing the number of dimensions of the test. From all these submissions, it could be inferred that most of the constructed-response test items of the public examinations are multidimensional and required the students to possess more than one latent trait for them to score good grades. Consequently, this could be one of the reasons why most of the students perform below average in Mathematics. This gives credence to the findings of Resckase (1985) that some test items demand more than one latent trait or ability to deal with, for instance, arithmetic and algebraic manipulations in Mathematics.

4.1.2. Research Question Two: How comparable are the parameter estimates of the 15-item of WAEC and NECO Mathematics constructed-responses tests under the framework of:

i. Item Response Theory (IRT):

a. Graded Response Model (GPCM)?

b. Generalised Partial Credit Model (GRM)?

Table 4.14: Discrimination and Difficulty Parameters of WAEC_MCRAT (GPCM and GRM)

Item	Model	a	b1	b2	b3	b4	b5	b6	b7	b8
1.	GPCM	0.41	0.24	-5.60	0.53	-4.36	5.06	-3.72	5.08	-3.97
	GRM	1.41	-3.16	-2.79	-1.64	-1.23	0.17	0.32	1.71	1.34
2.	GPCM	0.32	7.51	-9.52	3.85	-7.40	5.33	-3.84	8.69	-8.33
	GRM	1.21	-2.27	-2.20	-1.37	-1.21	0.06	0.29	1.21	1.29
3.	GPCM	0.23	12.00	-13.44	6.94	-12.95	5.79	-7.43	6.28	-3.23
	GRM	0.75	-3.81	-3.74	-2.69	-2.54	-0.79	-0.44	1.54	2.21
4	GPCM	0.24	10.74	-10.46	6.57	-8.53	6.63	-4.39	4.65	-1.90
	GRM	0.97	-1.60	-1.50	-0.62	-0.46	0.67	0.94	1.97	2.56
5	GPCM	0.37	6.70	-6.53	5.66	-5.51	5.81	-2.96	6.20	-2.07
	GRM	1.38	-0.81	-0.72	0.14	0.25	1,27	1.46	2.51	2.76
6	GPCM	0.03	176.67	-133.79	69.73	-85.13	110.24	-59.21	92.33	-98.77
	GRM	0.31	-0.76	1.61	1.36	1.74	5.76	6.13	8.26	8.54
7	GPCM	0.14	36.07	-25.19	20.01	-31.31	28.71	-11.78	7.22	-8.04
	GRM	0.65	-0.70	-0.71	-0.14	-0.11	3.35	3.47	4.29	4.73
8	GPCM	-0.00	-1609.57	905.70	-563.02	1290.77	1205.20	755.68	-838.93	1661.23
	GRM	-0.15	10.95	10.90	10.14	10.02	3.93	3.83	2.67	2.59
9	GPCM	0.08	84.87	-63.68	38.38	-47.53	24.87	39.38	-31.12	NA
	GRM	0.34	1.03	1.04	2.56	2.65	8.40	11.38	11.63	NA
10	GPCM	0.06	102.87	-80.86	46.04	-53.39	32.61	3.92	NA	NA
	GRM	0.26	1.08	1.10	3.56	3.71	11.38	14.69	NA	NA
11	GPCM	0.21	28.62	-19.68	13.26	-15.38	10.54	9.97	-6.59	NA
	GRM	0.80	0.90	0.91	1.61	1.67	4.17	5.19	5.44	NA
12	GPCM	0.16	41.00	-29.04	16.09	-22.61	27.68	-12.01	17.63	-6.66
	GRM	0.64	0.68	0.69	1.33	1.39	5.33	5.49	7.79	8.17
13	GPCM	0.08	59.43	-37.37	26.06	-33.19	39.95	-15.46	6.90	5.21
	GRM	0.25	1.99	2.08	4.03	4.30	10.70	11.25	14.09	17.70
14	GPCM	-0.33	-0.28	-0.34	-1.88	-2.07	-6.89	-7.68	8.40	-8.44
	GRM	-0.36	-0.50	-0.51	-3.24	-7.18	-7.93	-9.46	-10.13	-10.13
15	GPCM	-0.07	-68.56	47.27	-30.60	40.81	-35.27	-4.35	-41.69	60.89
	GRM	-0.33	-0.28	-0.34	-1.88	-2.07	6.89	-7.68	-8.40	-8.44

Positive parameter = high ability, negative parameter = low ability

Table 4.14 presents the comparison of the discrimination and difficulty parameters of WAEC under the two IRT models. From this table, it shows that for a student to score 1, a student must be above average level of ability in Mathematics (above-average level of ability 0.24) for the first item (considering *b1* parameter). Whereas a weak student can easily get 1 using the GRM model (low ability level of -3.16), for item 2, the average level of ability required is 7.57 while low-level ability is -2.27. The items that required low student ability to pass in WAEC test are 14 and 15. The implication of this is that WAEC _MCRI are very difficult.

Table 4.15 Discrimination and Difficulties Parameters of NECO_MCRAT (GRM and GPCM)

Item	Model	a	b1	b2	b3	b4	b5	b6	b7	b8
1	GPCM	0.17	2.10	-8.18	3.40	-9.62	9.73	-3.77	9.08	-8.706
	GRM	0.60	-4.95	-4.05	-2.28	-1.63	0.92	1.46	2.79	3.21
2	GPCM	0.47	-12.01	6.44	-2.98	2.57	2.67	NA	NA	NA
	GRM	1.09	-5.84	0.51	0.64	1.26	1.53	NA	NA	NA
3	GPCM	0.21	16.42	-9.96	6.00	-8.31	11.86	-8.78	7.35	-5.29
	GRM	0.79	-0.33	-0.25	0.34	0.51	1.69	1.81	2.90	3.26
4	GPCM	0.23	7.31	-6.47	5.17	-6.81	13.21	-7.55	8.98	-5.24
	GRM	0.91	-1.35	-1.07	-0.08	0.21	1.91	2.05	3.20	3.50
5	GPCM	0.18	11.65	-5.29	2.81	-5.14	10.37	-4.37	7.63	-4.97
	GRM	0.82	-0.58	-0.32	0.34	0.76	2.10	2.40	3.37	3.79
6	GPCM	0.16	24.93	-17.44	13.33	-14.97	11.00	-1.60	21.95	-17.184
	GRM	0.79	0.02	0.07	0.96	1.07	2.77	3.32	4.67	4.74
7	GPCM	0.16	16.05	-11.21	4.43	-13.12	12.47	-5.48	7.40	-8.506
	GRM	0.55	-2.59	-2.42	-1.57	-1.20	1.38	1.77	2.92	3.40
8	GPCM	0.07	32.77	-12.14	7.73	-23.37	22.90	-21.05	29.19	-41.428
	GRM	0.33	-3.80	-3.36	-2.49	-2.01	0.02	0.45	2.23	2.50
9	*GPCM*	0.33	8.60	-6.39	2.24	-4.76	7.55	-3.24	5.4422.	-3.544
	GRM	1.45	-0.71	-0.65	-0.20	-0.01	1.22	1.37	0.40	2.23
10	GPCM	0.18	19.28	-10.75	13.50	-19.00	9.79	7.84	-6.78	NA
	GRM	0.86	-0.16	-0.08	0.47	0.52	2.60	3.43	3.76	NA
11	GPCM	0.26	15.75	-9.87	9.62	-10.70	13.19	-7.20	10.80	-7.145
		1.19	0.29	0.33	0.90	0.96	2.30	2.38	3.25	3.36
12	GPCM	-0.01	-497.81	340.28	-171.08	224.30	-393.52	197.86	-298.15	87.426
		0.16	2.51	2.58	5.57	6.08	14.91	15.12	17.25	17.34
13	GPCM	-0.01	-379.32	232.13	-161.92	173.32	-257.61	35.99	-75.13	NA
		0.12	8.49	8.71	14.29	15.16	35.31	38.42	49.02	NA
14	GPCM	-0.00	-3082.97	-1981.36	-134.94	685.66	-765.71	-110.16	-813.75	460.181
		0.11	4.49	4,52	6.93	8.89	21.93	27.59	36.54	39.54
15	GPCM	-0.09	-51.34	34.90	-18.05	22.36	-29.98	13.76	10.28	NA
		-025	-0.54	-0.63	-2.69	-3.09	-6.95	-7.31	-8.84	NA

Positive parameter = high ability, negative parameter = low ability

Table 4.15 presents the comparison of the discrimination and difficulty parameters of WAEC under the two IRT models. From this table, it shows that for a student to score 1, a student must be above average level of ability in Mathematics (above-average level of ability 2.10) for the first item (considering *b1* parameter category). Whereas a weak student can easily get 1, using the GRM model (low ability level of -4.95). For item 2, the low-level ability of -12.00 and -.5.84. Also, the table shows that most of the parameters are negative. The implication of this is that NECO _MCRI are not too difficult to compare to WAEC.

The findings of the study showed that an upward discrimination (a) index with values of when polytomous generalized partial credit model was used to calibrate WAEC and NECO CR items (see table 4.16 and 4.18). The implication of this is that both examination items discriminate well but NECO constructed-response items are of quality when compared to WAEC under the graded response model. However, the WAEC constructed-response test items discriminated among the students of low and high ability very well with the values ranged from -0.00 to 1.41 while NECO discriminating index ranged above WAEC values which ranged from 0.00 to 1.45. This indicates that NECO test items produced better parameters and gives good estimates. This corroborated Ayanwale (2019) that concluded that NECO constructed-response test items discriminate well among students. Similarly, it was observed that IRT frameworks did not give the estimates of all the item parameters of both WAEC and NECO Mathematics test items under the item analysis process due to the student's failure to attempt all questions. This is in disagreement with Ayanwale (2019) that indicated that the IRT model used gave all item parameters. These submissions also agreed with Adegoke (2016) that there were no values for some items because the value for r (Pearson correlation between the item and the theta estimates) was very low.

Conversely, the result disagreed with Muraki (2002) and Olonode (2017) that IRT frameworks gave the estimates of all the item parameters of the constructed-response items. The result of the study corroborates Udofia and Udo (2017) that attempted to show that as far as the distribution of questions across the various levels of the cognitive domain, the themes and topics in the syllabus is concerned; WAEC and NECO are comparable concerning Mathematics items. It could be concluded from the findings of this study that the provision of all parameters depends on the examinee's ability to respond to all items so administered to them. Also, it shows that NECO constructed-response items were found to be less difficult compared to WAEC items. The results further showed that 8 of the items out 15 for while 10 out of 15 for NECO were not appropriate for the student's ability level because it required high-level ability.

4.1.4 Research Question Three: Is there any statistically significant mean difference between the ability scale score of students in WAEC and NECO mathematics constructed-response tests along?

i. Classical Test Theory (CTT)

ii Graded Response Model (GRM)?

iii. Generalised Partial Credit Model (GPCM)?

For comparability purposes, the z-scores of students were converted using t-score as earlier explained in chapter 3. The mean differences in student's ability scores in WAEC and NECO were estimated using CTT and IRT (GPCM and GRM). The following analyses were carried out to that effect:

Table 4.16: Paired t-test for CTT and IRT (GPCM) Model of Students Ability in (WAEC_MCRAT)

Model	Mean ($\bar{x}$)	Std. Deviation	Mean Difference	df	t	P-Value
WAEC CTT	35.88	10.02	-5.82	1150	-34.83	0.00*
IRT(GPCM)	41.70	7.0				

*= Significant p<0.05, Not Significant p>0.05

Table 4.16 present the estimated mean score of students' ability in WAEC_MCRAT under CTT as 35.88 while the mean score under IRT (GPCM) was 41.70 respectively. From Table 4.16, it is evident that the mean difference between students' score in CTT and IRT (GPCM) of -.5.82 is statistically significant (t= -34.83, df = 1150, p = 0.00). This implies that the IRT (GPCM) model gives a better score compared to CTT.

Table 4.17: Paired t-test for CTT and IRT (GRM) Model of Students Ability in (WAEC_MCRAT)

Model	Mean ($\bar{x}$)	Std. Deviation	Mean Difference	Df	t	P-Value
WAEC CTT	35.88	10.02	5.84	1150	34.86	0.00*
IRT(GRM)	41.71	7.04				

*= Significant p<0.05, Not Significant p>0.05

Table 4.17 present the estimated mean score of students' ability in WAEC_MCRAT under CTT as 35.88 while the mean score under IRT (GRM) was 41.71 respectively. From Table 4.17, it is evident that the mean difference between students' score in CTT and IRT (GRM) of 5.84 is statistically significant (t= -34.86, df = 1150, p = 0.00). This implies that the IRT (GRM) model gives a better score compared to CTT.

Table 4.18: Paired t-test for CTT and IRT (GPCM) Model of Students Ability in (NECO_MCRAT)

Model	Mean ($\bar{x}$)	Std. Deviation	Mean Difference	df	t	P-Value
NECO CTT	33.49	12.39	-8.19	1150	-33.32	0.00*
IRT(GPCM)	41.67	6.98				

*= Significant p<0.05, Not Significant p>0.05

Table 4.18 present the estimated mean score of students' ability in NECO_MCRAT under CTT as 33.49 while the mean score under IRT (GPCM) was 41.67 respectively. From Table 4.18, it is evident that the mean difference between students' score in CTT and IRT (GPCM) of -.8.19 is statistically significant (t= -33.32, df = 1150, p = 0.00). This implies that the IRT (GPCM) model gives a better score compared to CTT.

Table 4.19: Paired t-test for CTT and IRT (GRM) Model of Students Ability in (NECO_MCRAT)

Model	Mean ($\bar{x}$)	Std. Deviation	Mean Difference	df	t	P-Value
NECO CTT	33.49	12.39	-8.18	1150	-35.04	0.00*
IRT(GRM)	41.67	7.01				

*= Significant p<0.05, Not Significant p>0.05

Table 4.19 present the estimated mean score of students' ability in NECO_MCRAT under CTT as 33.49 while the mean score under IRT (GRM) was 41.67 respectively. From Table 4.19, it is evident that the mean difference between students' score in CTT and IRT (GRM) of -8.18 is statistically significant (t= -35.04, df = 1150, p = 0.00). This implies that the IRT (GRM) model gave a better score compared to CTT.

The results in tables 4.15, 4.16, 4.17, and 4.18 shows that the mean score of students when CTT was used was less compare to IRT (GPCM and GRM) frameworks. When the graded response model (GRM) was used it was slightly greater than the mean score of the student's ability when generalized partial credit model (GPCM) framework and CTT. This implies that the GRM model gives better estimated scores compared to GPCM and CTT. These findings of this study is not in tandem with the findings of studies carried out by Fan (1998), Courville (2004), Wiberg (2004), Adegoke (2014), Metibemu (2016) and Ayanwale (2018) that there was no significant difference in students' test scores when estimated under CTT and IRT frameworks.

However, the IRT framework gives better estimates because it shows that there are significant differences in the scores obtained by students. But the findings of this study in tandem with Ajao and Awogbemi (2012) that revealed that a student who had credit in WAEC mathematics would have at least a credit or pass in NECO mathematics. The majority of the Students who had credit and above in NECO mathematics obtained at least passes in WAEC mathematics and those who failed in NECO mathematics also failed in WAEC mathematics.

Similarly, the result of the analysis further shows that the mean difference between WAEC and NECO test items was statistically significant. The implication of this is that students' ability from the two combined test forms under CTT was not different. Under the two IRT frameworks, the mean differences were not statistically significant for GPCM and GRM respectively. This implies that the student's ability in both examinations of the students was different and this means that CTT gave a low score compared to IRT. It could be concluded that IRT is given better estimates of student ability in mathematics.

4.1.4 Research Question Four: How equivalent are the scores of the students in 15-item of WAEC and 15-item of NECO Mathematics constructed-response tests under the framework of:

i. Classical Test Theory (CTT)?

ii. Item Response Theory (IRT)?

In other to answer this research question, the linear equating model was used. The student's ability estimate gotten from CTT score converted to z score for WAEC were placed on a common scale with the ability score for NECO test items. According to Livingston (2004), the linear equating adjustment is written as a mathematical formula such as If X represents a score on the new form and Y represents a score on the reference form, then X and Y are equivalent in a group of students.

$$Y – Mean (Y)/ SD(Y) = X – Mean (X)/SD(X) \quad \text{----------- Equ. (1)}$$

Where the means and standard deviations are computed in that group of students (for instance. the target population). Solving this equation for the reference form score Y will give us a formula for adjusting any given raw score X on the new form.

$$Y = [SD(Y)/SD(X)] X + \{mean (Y) – [SD(Y)/SD(X)] mean (X)\} \text{-----Equ. (2)}$$

Thus, this formula was used to calculate for the adjusted WAEC students score

For instance, let W signify WAEC(Y) and N signify NECO (X). Then the value of mean (WAEC = 32.30 and NECO = 30.78 and standard deviation (WAEC= 10.02 and NECO =12.39) when substituted in the equation using the score of the first candidate in NECO, we will have:

Wx - Wmean-x/SD = Nx - Nmean-x

Wx-32.30/10.02 = Nx – 30.78/12.39

Therefore:

Wx – 32.02	=	10.02/12.39 (Nx – 30.78)
	=	0.81 (Nx – 30.78)
	=	0.81Nx – 24.93 + 32.02
	=	081Nx + 7.37

Substitution:

$$Wx = 0.81\ (62.3) + 7.37$$
$$= 50.46 + 7.37$$
$$= 57.83$$

Table 4.20: Ability Estimate of NECO, WAEC and Adjusted WAEC for CTT

Students	NECO	WAEC	Adjusted WAEC	Students	NECO	WAEC	Adjusted WAEC
1	62.30	29.75	57.79	1134	35.63	48.92	36.22
2	32.30	19.75	33.53	1135	56.46	47.25	53.07
3	26.46	34.75	28.81	1136	26.46	33.09	28.81
4	18.13	38.92	22.07	1137	13.13	31.42	18.03
5	7.30	29.75	13.31	1138	30.63	30.59	32.18
6	28.13	26.42	30.16	1139	33.96	47.25	34.87
7	18.13	38.09	22.07	1140	43.96	51.42	42.96
8	10.63	26.42	16.00	1141	40.63	50.59	40.27
9	15.63	40.59	20.05	1142	45.63	41.42	44.31
10	13.96	7.25	18.70	1143	24.80	49.75	27.46
11	18.13	18.09	22.07	1144	30.63	21.42	32.18
12	13.13	16.42	18.03	1145	56.46	49.75	53.07
13	11.46	20.59	16.68	1146	34.80	38.92	35.55
14	7.30	33.92	13.31	1147	35.63	21.42	36.22
15	7.30	23.92	13.31	1148	56.46	17.25	53.07
16	8.13	25.59	13.98	1149	43.96	33.09	42.96
17	11.46	27.25	16.68	1150	40.63	31.42	40.27
18	37.30	26.42	37.57	1151	56.46	30.59	53.07
19	19.80	37.25	23.42	**Mean**	30.78	32.30	32.30
20	23.96	12.25	26.79	**SD**	12.39	10.02	10.02
+	+	+	+				
+	+	+	+				
+	+	+	+				
1132	56.46	43.09	53.07				
1133	34.80	47.25	35.55				

Table 4.20 revealed that when students ability scores in WAEC and NECO were placed on a common scale the WAEC adjusted ability gave almost the same estimate (M = 32.30; SD = 10.02) with the NECO ability estimate, (M = 30.78 SD= 12.39).

It can be inferred that the NECO and WAEC tests measured relatively the same construct and could be stated to have equal status. Similarly, the comparability level of the two tests was also confirmed using Person Product Movement Correlation (PPMC) as shown in table 4.21.

Table 4.21: Analysis of the Correlation between Students Ability in WAEC and and NECO Tests (CTT Estimate)

Achievement Tests	**Number**	**Mean**	**Std. Dev.**	***r***	**p-value**	**Remark**
NECO	1151	30.78	12.39	0.10	0.00*	Significant
WAEC	1151	32.30	10.02			

*=Significant at p<0.05

Table 4.21 presents the result of the Pearson Product Movement Correlational coefficient showing the relationship between students' ability in WAEC and NECO Mathematics tests. The result displays that there is a positive, low relationship between student's ability in the two achievement tests (r= 0.10; p = 0.00). This implies that a rise in students' ability in the WAEC test will cause a corresponding positive rise in students' ability in the NECO test. In addition, the result shows that the two tests were positively correlated.

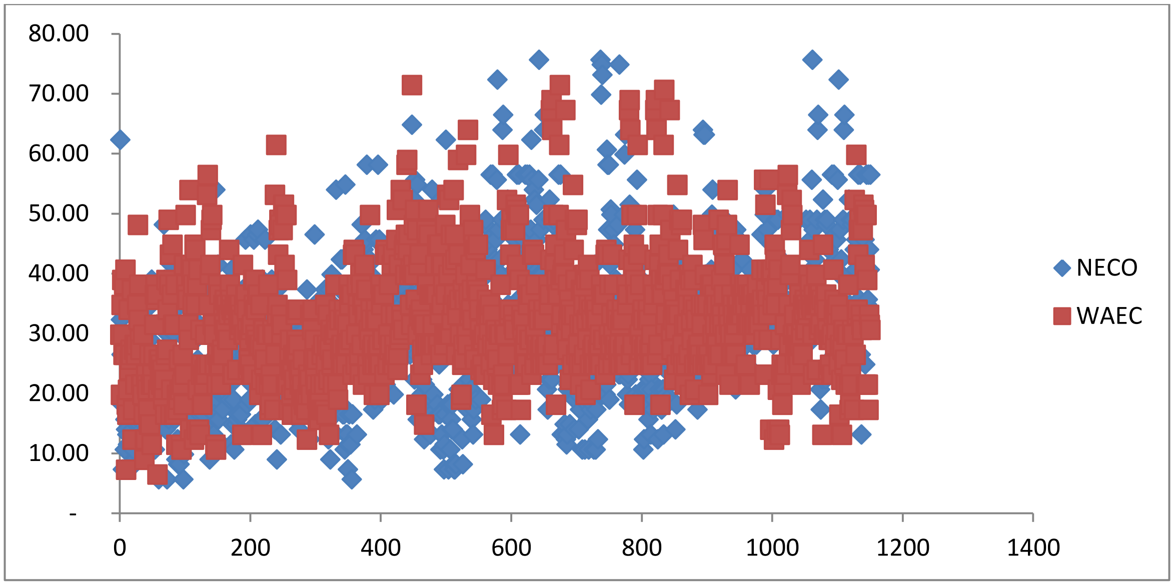


Figure 4.3: Scatter plot showing the pattern of correlation of student's ability in WAEC and NECO Mathematics constructed-response tests (CTT estimate)

From Figure 4.3, it could be deduced that the relationship between the students' abilities in WAEC and NECO Mathematics constructed-response tests under the CTT framework are relatively similar and this shows that student achievement in WAEC Mathematics would likely be the same in NECO Mathematics.

For IRT, The student's ability estimate gotten from IRT Pro 3 output for WAEC were placed on a common scale with the ability score for NECO test items.

$$Y = [SD(Y)/SD(X)]\ X + \{mean\ (Y) - [SD(Y)/SD(X)]\ mean\ (X)\} \text{ -----Equ. (2)}$$

Thus, this formula was used to calculate for the adjusted WAEC students score

For instance, let W signify WAEC(Y) and N signify NECO (X). Then the value of mean (WAEC = 0.001 and NECO = 0.004 and standard deviation (WAEC= 0.838 and NECO =0.840) when substituted in the equation using the score of the first candidate in NECO, we will have:

Wx - Wmean-x/SD = Nx - Nmean-x

Wx-0.001/0.838 = Nx – 0.004/0.840

Therefore:

$$\begin{aligned} Wx - 0.001 &= 0.838/0.840\ (Nx - 0.004) \\ &= 0.998\ (Nx - 0.004) \\ &= 0.998Nx - 0.004 + 0.001 \\ &= 0998Nx - 0.003 \end{aligned}$$

Substitution:

$$\begin{aligned} Wx &= 0.998\ (1.35) - 0.003 \\ &= 1.347-0.003 \\ &= 1.34 \end{aligned}$$

Table 4.22: Ability Estimate of NECO, WAEC and Adjusted WAEC for IRT

Students	NECO	WAEC	Adjusted WAEC	Students	NECO	WAEC	Adjusted WAEC
1	1.35	-0.55	1.36	1134	-0.06	2.11	-0.06
2	-0.22	0.92	-0.22	1135	2.21	1.72	2.22
3	-0.42	0.03	0.42	1136	0.16	0.36	0.16
4	-1.33	0.35	-1.33	1137	-0.78	0.26	-0.78
5	-1.92	-0.24	1.92	1138	0.65	0.00	0.66
6	-0.24	-0.15	-0.23	1139	0.57	1.28	0.57
7	-0.90	0.84	-0.90	1140	1.21	1.67	1.22
8	-1.09	-0.14	-1.09	1141	1.11	1.46	1.12
9	-1.16	1.03	-1.16	1142	1.46	0.61	1.46
10	-1.16	-1.31	-1.16	1143	0.21	1.57	0.21
11	-0.69	-0.79	-0.69	1144	0.87	-1.04	0.87
12	-1.38	-0.80	-1.38	1145	2.21	1.57	2.22
13	-1.62	-1.00	-1.62	1146	-0.36	0.78	-0.36
14	-1.68	-0.21	-1.68	1147	-0.06	-1.09	-0.06
15	-1.83	-0.38	-1.83	1148	2.21	-1.51	2.22
16	-0.90	0.02	-0.90	1149	1.21	0.36	1.22
17	-1.85	0.06	-1.85	1150	1.11	0.26	1.12
18	-0.16	0.00	-0.15	1151	2.21	0.00	2.22
19	-0.89	1.18	-0.89	**Mean**	**0.001**	**0.004**	**0.004**
20	-0.36	-1.16	-0.36	**SD**	**0.838**	**0.840**	**0.840**
+	+	+	+				
+	+	+	+				
+	+	+	+				
1132	2.21	0.90	2.22				
1133	-0.36	1.13	-0.36				

Table 4.22 revealed that when students ability scores in WAEC and NECO were placed on a common scale the WAEC adjusted ability gave almost the same estimate (M = 0.004; SD = 0.840) with the NECO ability estimate, (M = 0.001 SD= 0.838). It can be inferred that the NECO and WAEC tests measured relatively the same construct and could be stated to have equal status. Similarly, the comparability level of the two tests was also confirmed using Person Product Movement Correlation (PPMC) as presented in table 4.23.

Table 4.23: Analysis of the Correlation between Students Ability in WAEC and NECO Tests (IRT Estimate)

Achievement Tests	Number	Mean	Std. Dev.	*r*	p-value	Remark
NECO	1151	0.001	0.838	0.07	0.02*	Significant
WAEC	1151	0.004	0.840			

*=Significant at p<0.05

Table 4.23 presents the result of the Pearson Product Movement Correlational coefficient showing the relationship between students' ability in WAEC and NECO Mathematics tests. The result shows that there is a positive, low relationship between student's ability in the two achievement tests ($r = 0.07$; $p = 0.02$). This implies that a rise in students' ability in the WAEC test will cause a corresponding positive rise in students' ability in the NECO test. In addition, the result shows that the two tests were positively correlated.

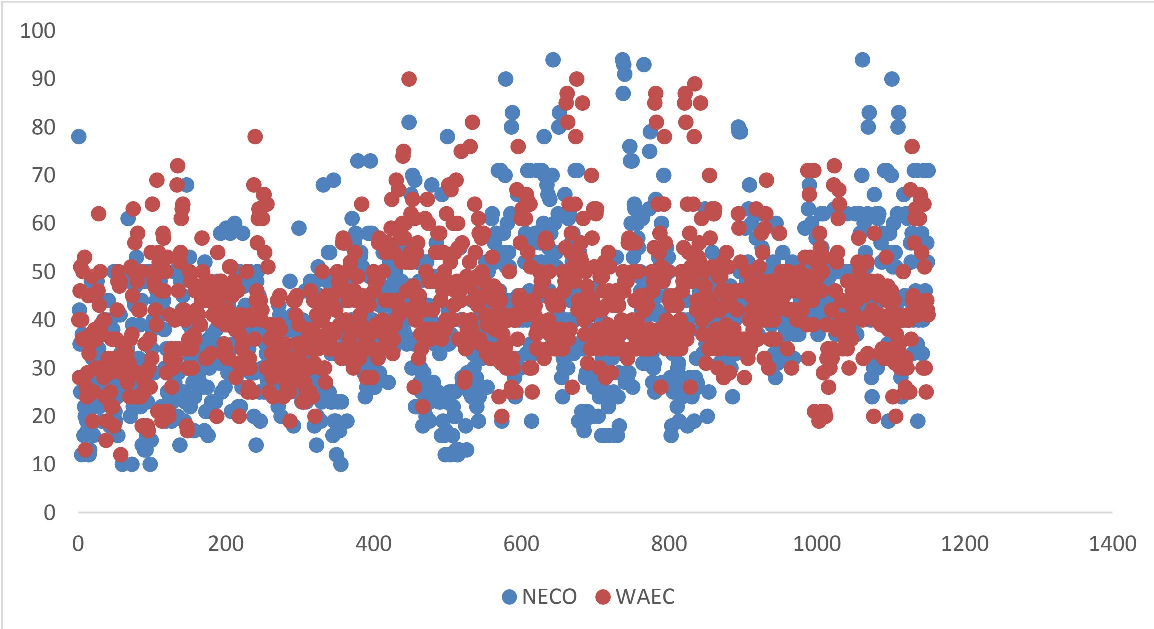


Figure 4.4: Scatter plot showing the pattern of correlation of student's ability in WAEC and NECO Mathematics constructed-response tests (IRTestimate)

From Figure 4.4, it could be deduced that the relationship between the students' abilities in WAEC and NECO Mathematics constructed-response tests under the IRT framework are relatively similar and this shows that student achievement in WAEC Mathematics would likely be the same in NECO Mathematics.

The result in Table 4.22 showed that student's ability in the two tests was equal. This is evident in the result when the tests were placed on a common scale the WAEC adjusted ability gave almost the same estimate as to the NECO ability estimate. This result supported the findings of Olonode (2017) that concluded that the equated scores of the examinee in 2014 WAEC and NECO Mathematics yielded a perfect relationship when all methods of IRT equating were used. Also. The finding agrees with the study of Ogbebor (2017) that confirmed 50-Mock Examination Test items and 2014-WAEC tests in Economics to measure the same construct using the linear equating method.

However, the results of this study disagrees with Metibemu (2016) that equated Physics self-developed achievement tests and WAEC 2014 which measures the same construct but are equal. The disagreement could be due to the peculiarity and differences in the subjects being equated. Furthermore, the study affirmed the findings of Adewale (2015) who reported using the linear method of equating for the 2013 and 2014 WAEC examinations and revealed that the variations among scores of the two examinations started from the score of 18. This implies that the linear method is somehow stable in producing equal results for a large number of different scores. The finding is in tandem with Olatunji (2015) that the linear equating methods equated better some of the scores obtained in WAEC and NECO compared to other methods of equating. Similarly, this study supported the findings of Bandele and Adewale (2013) that WAEC, NECO and NABTEB Mathematics items are comparable. Thus, this study further confirmed that not only WAEC and NECO selected-response items are equal but also the constructed-response items. It can be concluded that successes recorded by students in Mathematics in both NECO and WAEC are correlated. That is student abilities in WAEC and NECO are similar and both examinations are parallel.

4.1.5 Research Question 5(i): Which of the 15-item of WAEC and NECO Mathematics constructed-response test function differentially between male and female students using the framework of?

i. Classical Test Theory (CTT)?

ii. Item Response Theory (IRT)?

To answer this research question CTT and IRT differential item functioning frameworks were employed as follows:

Table 4.24: CTT Analysis of DIF on WAEC MCRAT concerning Gender

Item	Mantel	L-A LOR	LOR SE	LOR Z	DIF	Remark
1	4.03	0.24	0.12	1.98	No	
2	4.10	-0.24	0.12	-1.98	No	
3	0.51	-0.09	0.12	-0.69	No	
4	4.51	-0.24	0.11	-2.09	Yes	DIF favour female
5	8.08	0.33	0.12	2.79	Yes	DIF favour male
6	5.34	0.27	0.12	2.22	Yes	DIF favour male
7	0.08	-0.03	0.13	-0.27	No	
8	0.20	-0.06	0.13	-0.43	No	
9	0.00	0 .00	0.13	0.00	No	
10	0.91	0.12	0.13	0.92	No	
11	0.00	-0.00	0.14	0.02	No	
12	0.04	0.03	0.13	0.19	No	
13	2.49	-0.21	0.14	-1.53	No	
14	0.18	-0.05	0.13	-0.41	No	
15	0.12	-0.04	0.12	-0.34	No	

Table 4.24 presents the CTT-DIF results on student's item performance concerning Gender (Male and Female). The table gives the adjusted difficulty parameters estimate of the 15-items of WAEC Mathematics constructed-response items for the sub-groups, and also gave differences and probability on the items. Following the rules, the Mantel Chi-square statistic is distributed as Chi-square with one degree of freedom. Critical values of this statistic are 3.84 for a Type I error rate of 0.05 and 6.63 for a Type I error rate of 0.01. The Liu-Agresti cumulative common log-odds ratio (Liu and Agresti, 1996; Penfield and Algina, 2003) is asymptotically normally distributed.

Positive values indicate DIF in favour of the reference group (Male), and negative values indicate DIF in favour of the focal group (female). In addition, the Liu-Agresti cumulative common log-odds ratio is divided by the estimated standard error. A value greater than 2.0 or less than –2.0 may be considered evidence of the presence of DIF. From the results, out of the 15 items, 9 exhibits the presence of DIF (4, 5 and 6). Two items (4 and 5) were in favour of male students and one item (6) was in favour of female students. This implies that the WAEC Mathematics tests are fair to both boys and girls and there were few gender biases in the test items concerning CTT-DIF detect framework.

Table 4.25: IRT Analysis of DIF on WAEC MCRAT concerning Gender (Using Graded Response Model)

S/N	Male	Female	DIF Differences	Remarks
1	-0.83	-1.13	-1.96	DIF favour female
2	-0.44	-1.22	-1.66	DIF favour female
3	-1.23	-1.88	-3.11	DIF favour female
4	0.19	-0.33	-0.14	No DIF
5	0.69	0.00	0.69	DIF favour male
6	13.25	1.01	12.24	DIF favour male
7	1.53	0.58	0.95	DIF favour male
8	-155.87	1.40	-154.47	DIF favour male
9	8.00	1.02	6.98	DIF favour male
10	6.60	1.21	5.39	DIF favour male
11	1.99	1.41	0.58	DIF favour male
12	2.86	1.40	1.46	DIF favour male
13	7.51	6.04	1.47	DIF favour male
14	-3.51	-33.17	- 36.68	DIF favour female
15	-3.51	-4.17	-7.68	DIF favour female

Table 4.25 presents the IRT-DIF results of students' performance on items concerning Gender (Male and Female). The table gives the adjusted difficulty parameters estimate of the 15 WAEC Mathematics constructed-response items for the sub-groups, and also gave differences and probability on the items. The result shows that 14 items exhibit DIF because the difference in the difficulty parameters of the subgroups is greater than 0.50, while only one item did not exhibit DIF. Out of the 14 items with DIF 5 items (1,2,3,14 and 15) were in favour of female students while 10 items (4,5,6,7,8,9,10,11,12 and13) were in favour of the male students. This implies that the WAEC Mathematics tests were fair to boys than the girls. Also, there were many gender biases according to IRT-DIF detect framework compared to CTT-DIF detect framework.

Research Question 5(ii): Which of the 15-items of NECO Mathematics constructed-response test items function differentially between male and female students using:

i. Item Response Theory (IRT) framework?

ii. Classical Test Theory (CTT) framework?

To answer this research question IRT and CTT differential item functioning frameworks were employed as follows:

Table 4.26: CTT Analysis of DIF on NECO MCRAT concerning Gender

Item	Mantel	L-A LOR	LOR SE	LOR Z	DIF
1	0.85	0.11	0.12	0.92	No
2	0.07	-0.04	0.14	-0.27	No
3	0.67	-0.10	0.12	-0.81	No
4	0.04	-0.02	0.12	-0.20	No
5	3.51	0.22	0.12	1.86	No
6	0.13	-0.05	0.13	-0.35	No
7	0.78	-0.11	0.12	-0.88	No
8	1.54	-0.15	0.12	-1.22	No
9	0.02	-0.02	0.12	-0.13	No
10	0.23	0.06	0.13	0.49	No
11	1.38	0.16	0.14	1.18	No
12	0.29	-0.07	0.14	-0.54	No
13	2.28	0.23	0.15	1.49	No
14	0.21	-0.06	0.13	-0.46	No
15	0.02	0.02	0.13	0.14	No

Table 4.26 presents the CTT-DIF results on student's item performance concerning Gender (Male and Female). The table gives the adjusted difficulty parameters estimate of the 15 NECO Mathematics constructed-response items for the sub-groups, and also gave differences and probability on the items. Following the rules, positive values indicate DIF in favour of the reference group (Male), and negative values indicate DIF in favour of the focal group (female). In addition, the Liu-Agresti cumulative common log-odds ratio divided by the estimated standard error. A value greater than 2.0 or less than –2.0 may be considered evidence of the presence of DIF. From the results all 15- items did not show any evidence of the presence of DIF. This implies that the NECO Mathematics tests items were fair to both boys and girls and there were no gender biases in the test items concerning CTT-DIF detect framework.

Table 4.27: IRT Analysis of DIF on NECO MCRAT concerning Gender (Using Graded Response Model)

S/N	Male	Female	DIF Differences	Remarks
1	-1.08	-0.14	-0.95	DIF favour male
2	-0.67	-0.08	-0.59	DIF favour male
3	0.06	0.01	0.05	No DIF
4	-0.03	-0.00	-0.03	No DIF
5	0.84	0.10	0.73	DIF favour male
6	0.46	0.06	0.40	No DIF
7	-0.63	-0.08	-0.55	DIF favour male
8	-1.37	-0.17	1.20	DIF favour male
9	-0.26	-0.03	-0.23	No DIF
10	0.34	0.04	0.30	No DIF
11	0.61	0.08	0.54	DIF favour male
12	2.79	0.35	2.44	No DIF
13	7.60	0.95	6.65	DIF favour male
14	-13.23	-1.65	-11.57	DIF favour male
15	-2.82	-0.35	-2.46	DIF favour male

Table 4.27 presents the IRT-DIF results on student's item performance for Gender (Male and Female). The table gives the adjusted difficulty parameters estimate of the 15 NECO Mathematics constructed-response items for the sub-groups, and also gave differences and probability on the items. The result shows that 9 items exhibit DIF because the difference in the difficulty parameters of the subgroups is greater than 0.50, while 6 items did not exhibit DIF. Out of the 9 items with the DIF (1,2,3,5,7,8,11,13,14 and 15) non were in favour of female students. This implies that the NECO Mathematics tests were fairer to boys than the girls. Also, there were many gender biases according to IRT-DIF detect framework compared to CTT-DIF detect framework.

From the results from Table 4.24 to Table 4.27, it was detected that both WAEC and NECO Mathematics tests item exhibited DIF under CTT and IRT frameworks. It reveled that more than 50% of the WAEC test items are gender-biased under IRT while less than 20% exhibited gender bias under CTT. The DIF displayed favour male students than the female students. These items showed the DIF could have resulted in the structure of the questions and stem, which are not familiar to one of the subgroups. Thus, these could be the characteristics that affected the examinee's estimated ability scores.

More so, the result supported Enu (2014) that calibrated the developed Mathematics and Geography test items which showed significant DIF as regards gender. Similarly, the finding corroborates the work of Madu (2012) that out of 50 WAEC Mathematics multiple-choice items administered to students, 39 items displayed DIF in regards to gender. It also confirms the findings of Ogbebor (2017) that used CTT and IRT to detect DIF in 50 constructed Mock Economics tests and 2014 WAEC Economics exhibiting a significant DIF for the gender of the examinees.

Similarly, the findings are in tandem with the work of Metibemu (2015) that DIF existed in the PAT items constructed about school type, with 28 items out of the 100 items subjected for analysis. Thus, it has been further confirmed by the findings of this study that both WAEC and NECO constructed-response items usually contain some elements of DIF. The findings also agree with the work of Gierl, Rogers and Klingers (1999), a study on DIF in Alberta which examined 30 education Social Studies Diploma students.

Similarly, the study disagrees with Barati and Ahmadi (2010) whose study revealed some general gender DIF patterns across the subject area. Females were favoured on the three sections of grammar, language function, and the cloze test, whereas males were favoured on the vocabulary and word order sections. The reading comprehension section favoured males and females equally. From all these findings and the present study, it can be concluded that most public examinations are filled with the presence of DIF. In this study, there was confirmation on the item biases of public examining bodies which revealed that most items were filled with biases especially the WAEC test items.

CHAPTER FIVE

SUMMARY, CONCLUSION AND RECOMMENDATIONS

This chapter outlines the summary of findings, the implication of the study, recommendations, and limitations of the study, conclusion, and suggestions for further research as follows:

5.1 Summary

The application of test models in the calibration of the two public examination bodies Mathematics constructed-response test items revealed that the test items are multidimensional. The test items measured more than one trait of Mathematics ability and each item in the tests did not give a clue to the students in responding to another item. Both test theories models established that the tests had more than one dimension., The test reliability coefficient of both tests was high under Item Response Theory (IRT) but was low under Classical Test Theory (CTT) and shows that the IRT framework is valid and reliable compare to CTT, NECO Mathematics constructed-response test item parameters produced better estimates when compared to WAEC, There was a significant mean difference between the CTT and IRT scoring frameworks. Nevertheless, the IRT scoring framework produced different scores for students who have the same scores under the classical test theory framework.

Also, there was a significant mean difference in the observed scores of the students in NECO and WAEC tests using the CTT scoring framework while there was no significant mean difference in the observed scores of the students in NECO and WAEC tests using the IRT scoring framework The students' ability score of the two tests tend to be higher under IRT scoring models compare to CTT. The Pearson moment-correlation coefficient shows that the relationship between students' ability scores was significantly positively with a low relationship value. This showed that the two test forms can be used concurrently and interchangeably. In terms of comparison of both theories, there is a significant difference in the parameter estimates for CTT and IRT**.** The item parameters produced for WAEC constructed-response test items under IRT (GPCM) Model showed that high ability is required from students to score one while low ability is required from students under IRT (GRM) to score one. The item parameters produced for NECO constructed-response test items under both IRT (GPCM) and IRT (GRM) Models showed that high ability is required from students to score one.

However, In terms of item difficulty, NECO Mathematics constructed-response test items seemed difficult compared to WAEC Mathematics constructed-response test items. The linear equating method revealed that WAEC and NECO Mathematics constructed-response test items are equal and parallel, thus, measuring the virtually same construct. Both WAEC and NECO test items undergone DIF analysis using IRT and CTT frameworks and it was revealed that some of the items displayed DIF in respect to gender under the IRT for both examinations but only revealed complete evidence of the presence of DIF for WAEC. The DIF analysis of both CTT and IRT revealed that most of the Mathematics constructed-response test items of WAEC and NECO with evidence of the presence of DIF favour the male students compared to the female counterparts.

5.2 Conclusion

The focus of this study was to estimate and compare the ability of students in mathematics constructed-response test items of WAEC and NECO using Classical Test Theory (CTT) and Item Response Theory (IRT) frameworks. Findings from this study revealed that the mathematics achievement items of both public examining bodies were multidimensional and not conditional independence respectively. There was a significant difference between item parameters and observe scores under the two contrasting frameworks; the IRT framework gave better estimates than the CTT framework in the calibration and analysis of test items.

Further, CTT and IRT methods of scoring did not produce similar results on average. However, the IRT score method gave a better estimation of each of the students based on the student's ability. Moreover, the result of equating the two tests revealed that the tests are equal. Therefore, the two combined test forms cannot be used concurrently. The DIF analysis conducted for the two tests revealed that some of the items are gender-biased.

5.3 Recommendations

Based on the findings and implications of this study, the following recommendations were made: Examination bodies presently working within the CTT framework should incorporate the IRT framework into all the activities that involve students' assessment because of its efficiency. For content validity of a test, examination bodies such as WAEC and NECO should ensure that the test used is equated with an existing test measuring the same construct. Public examining bodies should endeavour to determine the psychometric properties of constructed-response test items using the IRT framework. Public examining bodies should shift from the paradigm of estimating students ability through the CTT approach to the IRT approach of scoring students ability in not only constructed-response tests but in all forms of tests. The Compilation and administration of past tests of public examining bodies should be used to prepare Senior Secondary School students for internal and external examinations.

5.4 Implication

The findings of this study have implications for test developers, school evaluators, teachers, learners and public examinations bodies. The use of classical test theory and item response theory is very essential in item analysis to establish psychometric properties. Item response theory framework gave better estimates in the calibration process of test items and also giving adequate information concerning the behaviour of an item as well as the students.

All stakeholders must involve in assessment procedures, especially, the examining bodies ensure constructed-response test items are properly scored using an effective test theory framework like IRT. In addition, IRT scoring frameworks gave almost the same student's ability score in WAEC and NECO. By implication, WAEC and NECO test items are equivalent and parallel. Another implication of the findings is that both CTT and IRT showed to be good methods for measuring DIF in a test, thus helping to identify the item that favours group membership about gender.

5.5 Limitations

The study was confronted with the following limitations:

i. This study would have sampled a larger population but the cost of printing test booklets and answer sheets was high and thus this sample was used.
ii. Time spent on the manual marking of the students answer booklet was much and this extended the timeframe allocated for marking.
iii. Some of the software employed for data analysis was expensive and complex to understand. So the researcher took a lot of time to study them.

5.6 Contributions to Knowledge

The study outcome contributed to knowledge in the following areas:

i. It filled the gap on the application of IRT frameworks in the establishment of psychometric properties and scoring of constructed-response items.

i. It brings to bear the level of validity and reliability of constructed-response items of the two famous public examination bodies for senior secondary students in Nigeria.
ii. It further confirmed that WAEC and NECO Mathematics tests are equivalent and parallel and debunk the speculation that one test is inferior to another.
iii. It proved that the models used in the study are reliable models for estimating and comparing the ability levels of students in the constructed-response test item. However, IRT models are the most efficient.

5.7 Suggestions for Further Studies

This study has investigated the comparability status of WAEC and NECO Mathematics constructed-response test items in terms of their equivalence. Also, the establishment of psychometric properties of the test item was examined and as well the fairness of constructed-response test items was considered. The following are suggested:

i. Further studies could be conducted on the use of the computerised scoring model for constructed-response test items.
ii. Studies similar to this could be conducted and researched in other subject areas.
iii. Another study could be carried out using another method of data collection design apart from the counterbalance design used in this study.
iv. Some other studies can be carried out using other multidimensional IRT software to establish item statistics of the data structure.

REFERENCES

Achor E., Imoko, B. I., and Ajai, J. 2010. Sex differentials in student's achievement and interest in geometry games and simulations technique, *Necatibey Faculty of Education Electronic Journal of Science and Mathematics Education* 4: 1-10.

Adedoyin, C. 2010. Investigating the invariance of person parameter estimates based on classical test and item response theories. An *international journal on education science* 2 (2); 107-113

Adegbuyi, J.Y. 2018. Construction and use of student mathematics engagement scale in predicting mathematics achievement among senior secondary school students in Ekiti-State, Nigeria A PhD Post-field presented at a seminar of the International Centre for Educational Evaluation, Institute of Education, University of Ibadan.

Adegoke, B.A. 2003. Teacher influence as a determinant of dependent prone students' learning outcomes in Geometry in Ibadan South East, Nigeria. An Unpublished PhD Thesis, University of Ibadan

Adegoke, B. A. 2013. Effects of item-pattern scoring method on senior secondary school student's ability scores in physics achievement test. *West African Journal of Education* 24: 181-190.

Adegoke, B. A. 2016. Estimation of students ability in physics essay test under two IRT polytomous models. In B.A. Adegoke, O. Popoola, and O.E. Babatunde (Eds.), *Public Examining in Sub-Saharan Africa: Issue, Challenges and Prospects*.229-240. Abuja: Marvelous Mike Press Ltd.

Adegoke, B. A. and Obot, P. F 2020. Comparison of Content Validity of Mathematics Test Items of Public Examining Bodies in Nigeria. *West African Journal of Education*

Adeleke, J.O. 2007. Identification and effect of cognitive entry characteristics of students learning outcomes in bearing mathematics. An Unpublished PhD Thesis, University of Ibadan

Adepoju, T.L .2002. Location factors as correlates of private cost and academic performance of secondary school students in Oyo State, Nigeria. Unpublished PhD. Thesis, University of Ibadan.

Adewale J. G 2015. Equating two Year BECE Results in Basic Science and Technology in Oyo State, Nigeria.

Akindele B. P, 2004. The development of an item bank for selection test into Nigerian universities. An exploratory study. Unpublished Ph.D. thesis, University of Ibadan.

Angoff; W H. 1971. Scales, norms, and equivalent scores In R L. Thorndike (Ed.), Educational measurement. Washington. American Council on Education.

Ajayi, I.S. and Awogbemi, A, C, 2012. Correlational analysis of students' achievement in WAEC and NECO Mathematics. *Journal of Educational and Practice*, 3(1):23-36

Akaike, H. 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19(6): 716–723.

Akinoso, S.O. 2011. Correlates of some factors affecting students' achievement in secondary school mathematics in Osun State. *In International Journal of education, science, mathematics and environmental studies* 30:83-95.

Alagoz, C. 2005. Scoring Tests with dichotomous and polytomous items. A Master of Arts Published Thesis, University of Georgia.

Ale, S. O. and Adetula, L. O. 2010. The national mathematical centre and the mathematics improvement project in nation-building. *Journal of Mathematical Sciences Education*.
1(1)1-19

Alu, A. O. and Adediwura. A. A. 2019. Assessment of dimensionality of Mathematics examination by National Examinations Council. *Journal of Educational Research and Development* (AJERD). 7, 210 - 215

Anderson, L.W and KrathwonhI, D. R. 2001. A Taxonomy for learning, teaching and assessing Revised December 2006 by Georgetown University School of Medicine.

Ani, E.N. 2014. Application of item response theory I the development and validation of multiple-choice tests in economics. (Master's Thesis). University of Nigeria, Nsukka

Anigbo, L. C. 2018. Comparison of 2017 West Africa Examination Council (WAEC) and National Examination Council (NECO) Senior Secondary School Certificate Examination Mathematics Questions I*nternational Journal of Education and Evaluation* ISSN 2489-0073 4(3):117-131.

Ariyo, S.O. 2017 Diagnostic assessment and effect of after-school programmes on the mathematics learning outcomes of low achieving students in Oyo State secondary schools. A PhD Post field presented at a seminar of the International Centre for Educational Evaluation, Institute of Education, University of Ibadan.

Assessment Systems Cooperation [ASC], 2014. Helping Your Organization Develop And Deliver Higher Quality Assessments Using Advanced Psychometrics.

Ayanwale, M.A. 2017. Efficacy of item response theory in score ranking and concurrent Validity of dichotomous and polytomous response mathematics achievement test in Osun State, Nigeria. A PhD Pre-field presented at a seminar of the International Centre for Educational Evaluation, Institute of Education, University of Ibadan.

Ayanwale, M.A. 2018. Efficacy of item response theory in score ranking and concurrent Validity of dichotomous and polytomous response mathematics achievement test in Osun State, Nigeria. A PhD Post-field presented at a seminar of the International Centre for Educational Evaluation, Institute of Education, University of Ibadan.

Ayanwale, M.A. 2019. Efficacy of item response theory in the validation and score ranking of dichotomous and polytomous responses Mathematics achievement tests in Osun State, Nigeria *An unpublished Ph.D Thesis,* University of Ibadan.

Barati, H. and Ahmadi A. R. 2010. Gender-based DIF across the Subject Area: A Study of the Iranian National University Entrance Exam *The Journal of Teaching Language Skills* 2(3) 61- 64

Bejar, I.I. 1977. An application of the continuous response level model to personality measurement. *Applied Psychological Measurement* 1*:* 509−521.

Bethany, M.C., and Michael, K., 2012. Seeing privilege Where It Isn't: Marginalized Masculinities and the Intersectionality of Privilege. Journal of Social Issue. 68 (1): 97-111.

Bichi, A. A. 2016. Differential Item Functioning Analysis: Evidence of Gender-Based Items in Chemistry Achievements. In B.A. Adegoke, O. Popoola, and O.E. Babatunde (Eds.), *Public Examining in Sub-Saharan Africa: Issue, Challenges and Prospects*.229-240.

Baker, F.B. 2001. The basics of item response theory. Test calibration. ERIC clearing House on Assessment and Evaluation. University of Maryland, College Park, MD. 136-330.

Baker F.B., and Kim, S.H. 2004. Item response theory: Parameter estimation techniques (2nd ed.). New York: Marce

Bandele, S.O. and Adewale A. E. 2013. Comparative analysis of the reliability and validity coefficient of WAEC, NECO and NATEB Constructed Examination. *Journal of Education and Social Research* 3(2): 397-402

Birnbaum, A. 1968. Some latent trait models and their use in inferring a student's ability. In F. M. Lord, and M. R. Novick, Statistical theories of mental test scores. Reading, MA: Addison-Wesley. 397-479.

Bock, R. D. 1972. Estimating item parameters and latent ability when responses are scored in two or more nominal categories. Psychometrika. 37: 29-51.

Bock, R. D. 1997.The nominal categories model. In W. J. van der Linden and R. K. Hambleton (Eds), *Handbook of modern item response theory* 33-49.

Cappelleri, J. C. Lundy, J. J. and Hays, R. D. 2014. Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient-reported outcomes measures. Elsevier HS Journals, Inc. Published by Elsevier Inc.

Chalmers, R. P. 2012. MIRT: a multidimensional item response theory package for the R environment. *J. Stat. Software* 48: 1–29.

Chen, S. 1996. A comparison of maximum likelihood estimation and expected a posteriori estimation in computerised adaptive testing using the generalised partial credit model. Unpublished doctoral dissertation, the University of Texas at Austin.

Chinason, M. M. 2011. Effect of cooperative learning strategy on student's retention in circle geometry in secondary school students in Benue State, Nigeria, American *Journal of Science and Industrial Research*, 2(1)33-36

Courville, T.R. 2004. An empirical comparison of item response theory and classical test

Daniel, F. 2014. Discriminating powers of the WASSCE and NECO SSCE general Mathematics Objective Question Papers. *Research Journal of Education* 2(9): 1-8.

Dawadi, B. R. 1999. Robustness of the polytomous IRT model to violations of the unidimensionality assumption. A paper delivered at the annual meeting of the American Educational Research Association, Montreal, Quebec, Canada.

De Ayala, R. J. 1989. Computerised adaptive testing: A comparison of the nominal response model and the three-parameter model. *Educational and Psychological Measurement*, *49,* 789-805.

De Ayala, R. J., Dodd, B. G., and Koch, W. R.1992. A comparison of partial credit and graded response models in computerised adaptive testing. *Applied Measurement in Education, 5(1):* 17-34.

De Ayala, R. J. 2009. The Theory and Practice of Item Response Theory. New York, NY: Guilford Publications.

De Boeck, P., and Wilson, M. 2004. Explanatory item response models: A Generalised Linear and Nonlinear Approach. New York: Springer.

DeMars, C.E. 2002. Recovery of graded response and partial credit parameters in MULTILOG and PARSCALE. A paper delivered at the annual meeting of the American Educational Research *Association*, Chicago, IL.

Donoghue, J.R. 1993. An empirical examination of the IRT information in polytomous scored reading items under the generalised partial credit model. *Journal of Educational Measurement* 31: 295-311.

Dorans, N.J. 2004. Equating, concordance and expectation. *Applied Psychological Measurement* 28(4) :227-246..

Dorans, N.J., Pommerich, M., and Holland, P.W. 2007. Linking *and aligning scores and scales.* New York, NY: *Springer*-Verlag.

Downing, S.M. 2006. The effects of violating standard item writing principles on tests and students: The consequences of using flawed test items on achievement examinations in medical education. *Advances in Health Sciences Education,* 10: 133-143.

Educational Testing Service 2010. Linking TOEFL iBT Scores to IELTS Scores*:* A Research Report*.* Princeton, NJ: *Educational Testing Service*.

Edward, A.L. and Thurstone, L.L. 1952. An internal consistency check for scale values determined by the method of successive intervals. *Psychometrika* 17:169-180.

Embretson, S. E., and Reise, S. P. 2000. Item response theory for psychologists. Mahwah, NJ: Lawrence Erlbaum.

Enu, V.O. 2014. The use of item response theory in the validation and calibration of mathematics and geography items for joint command school's promotion examination in Nigeria. banks for A PhD Pre-field presented at a seminar of the International Centre for Educational Evaluation, Institute of Education, University of Ibadan.

Ercikan, K., Schwarz, R.D., Julian, M.W.; Burket, G. R.;Weber, M. M. and Link, V. 1998. Calibration and scoring of tests with multiple-choice and constructed-response item types. *Journal of Educational Measurement 35:* 137-154.

Fakayode, O.T. 2017. Relative effectiveness of the classical test and item response theories in equating west African examinations council mathematics test scores for June and November 2015. A PhD Post-field presented at a seminar of the International Centre for Educational Evaluation, Institute of Education, University of Ibadan.

Fan, X. 1998. Item response theory and classical test theory: An empirical comparison of their item/person statistics. Educational and Psychological Measurement.

Federal Republic of Nigeria. 2013. Federal ministry of education. Abuja.

Federal Republic of Nigeria. 2013. National mathematics centre. Abuja

Fox, J.-P. 2010. Bayesian item response modeling: Theory and applications. Springer Science and Business Media.

Goolsby, L. 2013. *School Interest.* Boston: Allyn and Bacon.

Gierl, M. J., Rogers, W. T., Klinger, D. A. 1999. Using statistical and judgmental reviews to identify and interpret translation DIF. Alberta Journal of Educational Research, 45: 353-376.

Grima, A. M., and Weichun, W. M. 2002. Test scoring: multiple-choice and constructed-response items. A paper delivered at the annual meeting of the American Educational Research Association, New Orleans.

Hambleton, R. K. and Swaminathan, H. 1985. Item response theory: Principles and applications. Boston, MA: Kluwer-Nijhoff.

Hambleton, R. K., Swaminathan, H. and Rogers, H. J. 1991. Fundamentals of item response theory. Newbury Park, CA: Sage Press.

Hambleton, R. K. 1997. Principles and selected applications of item response theory. In R. L. Linn (Ed.), *Educational measurement* 3:147-200. Pheonix, Arizona: American Council on Education and the Oryx Press.

Hambleton, R. K. and Jones, R. W. 1993. An NCME instructional module on comparison of classical test theory and item response theory and their applications to test development. *Educ. Meas. Issues Pract* 12: 38–47. doi: 10.1111/j.1745-3992.

Hulin, C. L., Drasgow, F., and Parsons, C. K. 1983. Item response theory; Application to psychology www. Brain bench .com

Idigo, E.C. 2010. An effective method of retaining students' interest in mathematics in secondary Schools in Enugu East Local Government Area of Enugu State, Unpublished UG Thesis, Institute of Ecumenical Education, Thinker's Corner, Enugu, in Affiliation with (ESUT), Enugu.

Ifamuyiwa, A.S. 2005. Effects of self and cooperative instructional strategies on senior secondary school students. Learning outcomes in Mathematics. An unpublished PhD Thesis, University of Ibadan. Ibadan.

Johnson, T. R. and K. M. Kuhn 2015. Simulation-based Bayesian inference for latent traits of item response models: Introduction to the ltbayes package for R. Behavior Research Methods 47(4):1309 –1327.

Joint Admissions Matriculation Board. 2007. Abuja. Nigeria

Kolen, M. J. 1988. Smoothing methods for estimating test-score distributions. *Journal of Educational Measurement 28*: 257-282.

Kolen, M. J., and Brennan, R. L. 2014. Test equating, scaling, and linking*: Methods and practices* (3rd ed.). New York, NY: Springer-Verlag

Kurumeh, M., Agogo, P.,and Usman, K. O. 2010. Effect of Montessori method of teaching on junior secondary students achievement in mensuration in education zone B, Benue State of Nigeria. *American Journal of Scientific and Industrial Research* 1(2):375-379

Lau, C. A., and Wang, T. 1998. Comparing and combining dichotomous and polytomous items with SPRT procedure in computerized classification testing. Paper presented at the annual meeting of the American Educational Research Association, San Diego.

Lee, A and von Davier 2011. Equating through alternative kernels. In A. von Davier (ed.),StatisticalModelsforTestEquating,Scaling,andLinking,Springer,New York, 159 -173.

Livingston, S. A., and Rupp, S. L. 2004. Performance of men and women on multiple-choice and constructed-response tests for beginning teachers (ETS Research Report No. RR-04-48). Princeton, NJ: Educational Testing Service.

Liu, J. 2007. Comparing multidimensional and unidimensional computer adaptive strategies in psychological and health assessment. Unpublished Doctoral Dissertation, University of Pittsurg.

Lotfi, F. H., Dehkordi, S. S., and Vaez-Ghasem, M., 2012. Teaching Method Survey Based on Constructivist Approach in Permanent Learning Geometry. Applied Mathematical Sciences 6(6): 259 – 267.

Lord, F.M. 1980. Application of Item Response Theory to Practical Testing Problems. Hillsdale, NJ: Lawrence Erlbaum.20

Lord, F. N., and Novick, M. R. 1968. Statistical theories of mental test scores. Reading, MA: Addison-Wesley.

Lukhele, R., Thissen, D., and Wainer, H. 1994. On the relative value of multiple-choice, constructed-response, and students-selected items on two achievement tests. *Journal of Educational Measurement* 31:234–250.

Madu, B. C. 2012. Using transformed item difficulty procedure to assess gender–related differential item functioning of multiple-choice mathematics items administered in Nigeria. Research on Humanities and Social Sciences, 2 (6)41-55

Maduabum, M.A., and Odili, G.A. 2006. Analysis of students' performance in general mathematics at SSCE level in Nigeria 1991–2002.*Journal of Research and Curriculum Teaching* 1 (1*)* 64 – 68*.*

Mamman, M. and Eya, S.D. 2014. Trend analysis of students' mathematics performance in west African senior secondary certificate examinations from 2004 to 2013. Implication for Nigeria's Vision 20:2020

Mair P., and Hatzinger R., 2007. CML based estimation of extended Rasch models with the eRm package in R *Psychology Science* 49 (1): 26-43.

Masters, G. N. 1982. A Rasch model for partial credit scoring. Psychometrika 47: 149-174.

Masters, G. N. 1988a. The analysis of partial credit scoring. *Applied Measurement in Education* 1:*279−297*.

Masters, G. N. 1988b. Item discrimination: When more is worse. *Journal of Educational Measurement* 25: 15−29.

Masters, G. N. 1988c. Measurement models for ordered response categories. In R. Langeheine and J. Rost (Eds.), *Latent traits and latent class models* 11−29. New York: Plenum.

Mazzeo, J., Schmitt, A., and Bleistein, C. A. 1992.Sex-related performance differences on constructed-response and multiple-choice sections of Advanced Placement examinations (College Board Research Rep. No. 92-7; ETS Research Report No. RR-93-05). New York: College Entrance Examination Board

Martin, A. D., Quinn, K. M., and Park, J. H. 2010. Package 'MCMCpack'. Available at: http://cran.r-project.org/web/packages/MCMCpack/MCMCpack.pdf

Mellenbergh, G. J., 1995. Conceptual notes on models for discrete polytomous item responses. *Applied Psychological Measurement* 19:91–100,

Metibemu, M.A. 2015. Comparison of classical test theory and item response theory in the development and scoring of senior school physics tests in Ondo State, Nigeria. A PhD Pre-field presented at a seminar of the International Centre for Educational Evaluation, Institute of education, University of Ibadan

Metibemu, M.A 2016. Comparison of classical test theory and item response theory in the development and scoring of senior school physics tests in Ondo State, Nigeria. A PhD Post-field presented at a seminar of the International Centre for Educational Evaluation, Institute of education, University of Ibadan

Metibemu, M.A 2020. Assessment of the dimensionality and psychometric quality of the West African Examination Council (WASSCE) May/June 2014 Physics. In A, O. U. Onuka. *Public examining in Nigeria: A dynamic leadership*. 115-137

Michaelides, M. P., and Haertel, E H. 2004. Sampling of Common Items: An Unrecognized Source of Error in Test Equating. CSE Report 636: 33

Mislevy, R. J. 1986. Bayes modal estimation in item response models. Psychometrika 51: 177- 195.

Moyinoluwa, T. D. 2015. Analysing the psychometric properties of mathematics in public examinations in Nigeria Research on Humanities and Social Sciences. Retrieved from www.iiste.org

Muraki, E. 1990. Fitting a polytomous item response model to Likert-type data. Applied Psychological Measurement 14: 59-71.

Muraki, E. 1992. A generalised partial credit model: Application of EM algorithm. applied psychological measurement 16: 159-176.

Muraki, E., Hombo C. M., and Lee Y. 2000. Equating and linking of performance assessments. Educational Testing Service (ETS) DOI:10.1177/01466210022031787 Applied Psychological Measurement 24: 325.

National Council of Teachers of Mathematics. 2000. Teachers perspective on teaching mathematics. NCTM. Abuja

Nenty, H.J. 2000. Possible factors that influence student’s pattern of responses to mathematics examination items. *Journal of Educational Assessment* 17: 47-58.

Nigerian Educational Research and Development Council, 2013. Senior secondary school mathematics curriculum for SS one to three, Abuja. Nigeria.

Nosa, N.E. and Ohenhen, 1998. Science technology and mathematics. The tools for technological development in Nigeria. Ekiadolor. *Journal of Education* 3(1): 350 – 356

Obot, P. F 2019. Assessing the comparability of 2018 WAEC and NECO Mathematics Test Items. M.Ed. Project Submitted to International Centre for Educational Evaluation, Institute of Education, University of Ibadan, Nigeria

Ogbebor, U.C and Onuka A.O.U. 2013. Differential item functioning as an item bias indicator. Journal of International Education Research.367-373.http://www.interesiials.org/ER

Ogbebor, U.C. 2017. Mock economics test construction and parameter estimation using classical test and item response theories for senior secondary school students in Delta State, Unpublished PhD Post-Field Seminar paper presented at International Centre for Educational Evaluation, Institute of Education, University of Ibadan.

Ogunyomi, K. O. 2010. The effect of information and communication technology on teaching and learning Mathematics in selected Lagos State secondary school. M.Ed project. Department of Educational Foundation, University of Lagos. Ix-98

Ojerinde 2013. Classical test theory (CTT) VS item response theory (IRT): An evaluation of the comparability of item analysis results. A guest lecture presented at the Institute of Education, University of Ibadan on 23rd May.

Ojumba, D.P. 2012. Strategies for teaching and sustaining mathematics as an indispensable tool for technological development in Nigeria. Retrieved on 19th October 2004 from http/www.mcset.org/images/stories/muss-special.

Okafor, C. F., and Anaduaka, U. S., 2013. Consistent poor performance of Nigerian students in mathematics in senior secondary certificate examination (SSCE): What is not working? Journal of Research in National Development 1 (7): 247-251.

Oladele, B. K. and Adegoke, B. A. 2020. Using test theories models to assess senior secondary students ability in constructed-response mathematics tests. *Journal of Education and Practice.*11: 46 -55 www.iiste.org.

Oladele, B. K., Adegoke, B. A. and Longjohn, D. A. 2020. Assessment of differential item functioning in public examinations mathematics constructed-response tests, *Journal of Positive Psychology and Counselling* 4: 221 – 233

Olatunji, D. S. 2015. Analysis of linear and equipercentile equating of senior school certificate examination Economics multiple-choice papers in Kwara State, Nigeria. Unpublished Ph.D. Thesis, University of Ilorin.

Olonade, P. O. 2017. Equating and scaling 2014 senior school certificate mathematics examinations of west African examinations council and national examinations council in Lagos State, Nigeria. A PhD Post-field presented at a seminar of the International Centre for Educational Evaluation, Institute of Education, University of Ibadan. Ibadan.

Omotayo, S.A. 2016. Effects of Dynamic Geometry Software and 5e Instructional Model on Students' Achievement, Interest and Retention in Senior Secondary SchoolGeometry in Ibadan. A PhD Post-field presented at a seminar of the International Centre for Educational Evaluation, Institute of Education, University of Ibadan.

Ostini, R. and Nering, M. L. 2006.Polytomous item response theory models. Thousand Oaks: Sage Publication.

Oyo State Ministry of Education. 2017. List of senior secondary schools in Oyo State. Oyo State Government. Nigeria.

Penfield, R.D. 2001. Assessing differential item functioning among multiple groups: a comparison of three Mantel-Haenszel procedures. *Applied Measurement in Education 14*: 235-259.

Penfield, R. D., and Algina, J. 2003. Applying the Liu-Agresti Estimator of the Cumulative Common Odds Ratio to DIF Detection in Polytomous Items. *Journal of Educational Measurement* 40:353-370.

Pido, S., 2012.Comparison of item analysis results obtained using item response theory and classical test theory approaches. Journal of educational assessment in Africa, 7: 192-209

Rasch, G. 1977. On specific objectivity: An attempt at formalising the request for generality and validity of scientific statements. *Danish Yearbook of Philosophy 14:* 58–94.

Rasch, G., 1960.Probabilistic models for some intelligence and attainment tests. Danish Institute for Educational Research, Copenhagen, Denmark.

Reckase, M. D. 1985. The difficulty of items that measures more than one ability/ *Applied Psychological Measurement* 9:401-412.

Reckase, M. 2009. Multidimensional item response theory. New York, NY: Springer.

Rizopoulos, D. 2006. ltm: An R Package for latent variable modelling and item response Theory Analyses. Journal of Statistical Software 17: 1-25.

Rosa, K., Swygert, K.A., Nelson, L., and Thissen D. 2001. Item response theory applied to combinations of multiple-choice and constructed-response items-Scale scores for patterns of summed scores. In D. Thissen and H. Wainer (Eds.), Test Scoring pp. 253-292. Mahwah, NJ: Lawrence Erlbaum.

Sama, R. 2017. Washback effects of the West African senior secondary certificate mathematics examination on teaching and learning in secondary schools in Ebonyi State, Nigeria. A Ph.D Post-field presented at a seminar of the International Centre for Educational Evaluation, Institute of Education, University of Ibadan.

Samejima, F. 1969. Estimation of latent ability using a response pattern of graded scores. Psychometric Monograph, No.17.

Samejima, F., 1979. A new family of models for the multiple-choice item (Research Report No. 79-4). Knoxville, TN: University of Tennessee, Department of Psychology.

Schipke C. S., Knoll L. J., Friederici A. D., Oberecker R. 2012. Preschool children's interpretation of object-initial sentences: neural correlates of their behavioral performance. *Dev. Sci.* 15, 762–774. 10.1111/j.1467-7687.2012.01167.

Schumacker, R.E. Si C. B R. and Mount R. 2003.Ability estimation under different item parameterisation and scoring models. *American Educational Research Association* Chicago, Illinois USA.

Schwarz, G.E. 1978. Estimating the dimension of a model. Annals of Statistics 6(2): 461–464.

Si, Ching-Fung B. 2002. Ability estimation under different item parameterisation and scoring models. An Unpublished Doctoral Thesis, University of Texas 107

Steven, J.P. 2012. Applied Multivariate Statistics for the Social Sciences, Fifth Edition. 664

Stone, C. A., and Yeh, C. C., 2006. Assessing the Dimensionality and Factor Structure of Multiple-Choice Exams. *Educational and Psychological Measurement* 66(2): 193 - 214.

Svetina, D., and Levy, R., 2014. A Framework for dimensionality assessment for multidimensional item response models. *Educational Assessment.* 19(1) 35-57.

Swaminathan, H. and Roger, H. I. 1990. Detecting Item Differential Functioning using Logistic Regression Procedures. *Journal of Educational Measurement.* 27: 361-370

Thissen, D. 1991. The MULTILOGTM (Version 6) [Computer software]. Mooresville, IN: Scientific Software International.

Thissen, D., Chen, W-H, and Bock, R. D. 2003. MULTILOG 7 for Windows: Multiple category item analysis and test scoring using item response theory [Computer software]. Lincolnwood, IL: Scientific Software International, Inc.

Thissen, D., and Steinberg, L. 1984. A response model for multiple-choice items. Psychometrika 49:501-519.

Thissen, D., and Steinberg, L. 1986. A taxonomy of item response models. Psychometrika, 51, 567-577.

Thissen, D., Steinberg, L., and Fitzpatrick, A. R. 1989. Multiple-choice models: The distractors are also part of the item. Journal of Educational Measurement, 26, 161-176.

Udofia, Nsikak-Abasi and Udo, K. I. 2017.Comparative analysis of WAEC and NECO senior secondary school Mathematics examination. *Mathematical Theory and Modelling* www.iiste.org ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)

University of North Carolina at Greensboro.

Usman, K.O. and Nwoye, M.N. 2010. Effect of graphical symbol approach on the pupils' achievement in ratio at upper Primary school level in Nsukka central L.G.A. *Journal of Mathematical Centre, Abuja,* I (1), 123-132.

Usman, K. O. 2002. A review of studies on process errors by students in solving mathematical problems. *Journal of Nigerian Education Research Association*. 15 (1) 76-83

Usman, K. O. 2002 and Nwabueze, M. U., 2011. Promoting students' interest in quadratic equations in area: Tiles approach in mathematics instruction. ABACUS: *The Journal of the Mathematics Association of Nigeria* 36(1): 74-84.

Vermunt, J. K, and Magidson, J. 1996. Graphical displays for identified and unidentified latent class cluster and factor Models.

van der Linden, W. J. and R. K. Hambleton 199. Handbook of modern item response theory 33-49.

von Davier, A. A., Holland, P. W. and Thayer, D. T. 2004. The kernel method of test equating. New York. NY: Springer.

Watkins, D. S., 2005. *Product eigenvalue problems* (.pdf), *SIAM Review* 47:3-40. epubs.siam.org

Wainer, H., 1995. How is reliability related to the quality of test scores? What is the effect of local dependence on reliability? *Educational Measurement: Issues and Practic*e, 15(1): 22-29.

Welch, C. 2006. Item/prompt development in performance testing. In S.M. Godschalk, F. I., Swineford, F., and Coffman, W. E. 1966.*The measurement of writing ability*. New York: College Entrance Examination Board.

West Africa Examinations Council, Nigeria. 2013. Chief examiners report. Available: http://waeconline.org.ng/e-learning/

West Africa Examination Council, Nigeria 2014. Chief examiners report. Available http://waeconline.org.ng/e-learning/

West Africa Examination Council, Nigeria 2015. Chief examiners report. Available http://waeconline.org.ng/e-learning/

Wiberg, M. 2004. Classical test theory vs item response theory: An evaluation of the theory test in the Swedish driving-license test. *Educational Measurement Working paper:* EM No 50, UMEA University.

Wiberg, M., and González, J. 2016. Statistical assessment of estimated transformations in observed-score equating. *J. Educ. Meas.* 53, 106–125. doi: 10.1111/jedm.12103

Wilson, M., and Hoskens, M. 2001.The rater bundle model. *Journal of Educational and Behavioral Statistics, 26*, 283–306.

Yen, T.Y. 1993.A comparison of three statistical procedures to identify clusters of items with Local dependency. Huynh University of Carolina

Zimowski, M., Muraki, E., Mislevy, R. J., and Bock, R. D. 2003. BILOG-MG 3: Item analysis and test scoring with binary logistic models. Chicago, IL: Scientific Software. [Computer software].

Zumbo, B. D. 1999. A handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modelling as a Unitary Framework for Binary and Likert-like (Ordinal) Item Scores. Ottawa, Canada: Directorate of Human Resources Research and Evaluation.

Zhang. M., 2016. Exploring the dimensionality of scores for mixed-format tests. PhD\ (Doctor of Philosophy) thesis, University of Iowa, https://doi.org/10.17077/etd.

APPENDIX I

INTERNATIONAL CENTRE FOR EDUCATIONAL EVALUATION, INSTITUTE OF EDUCATION UNIVERSITY OF IBADAN

MATHEMATICS ACHIEVEMENT TEST OF WEST AFRICAN EXAMINATIONS COUNCIL (WAEC) FOR SSS THREE

INSTRUCTION: Please tick as applicable to you and answer all the questions

TIME ALLOWED: 2 hours 30 minutes

NAME OF SCHOOL:

GENDER: Male ( ), Female ( )

AGE: 14 – 16 ( ), 17- 19 ( ), 20 – 22 ( )

NUMBER AND NUMERATION

1 (a) Simplify, without using tables or calculator:

$$\frac{\frac{3}{4}\left(3\frac{3}{8}+1\frac{5}{8}\right)}{2\frac{1}{8}-1\frac{1}{2}}$$

(b) Given that $\log_{10}{}^{2} = 0.3010$ and $\log_{10}{}^{3} = 0.4771$, evaluate correct to 2 significant figures and without using tables of calculator, $\log_{10} 1.125$.

2 (a) Simplify $3\sqrt{75}-\sqrt{12}+\sqrt{108}$, leaving the answer in surd form (radicals)

(b) Without using tables or calculator, simplify: $\frac{0.6x32x0.004}{1.2x0.008x0.16}$

Leaving the answer in standard form (scientific notation)

3. (a) Without using Mathematical tables or calculators, simplify: $3\frac{4}{9}+(5\frac{1}{3}-2\frac{3}{4})+5\frac{9}{10}$

(b) If 124, = 232_{five}, find n.

ALGEBRAIC PROCESS

4. (a) Solve the simultaneous equations:
$^1/_x + ^1/_y = 5$
$^1/_y - ^1/_x = 1$
(b) Solve the inequality: $4 + \frac{3}{4}(x + 2) \leq \frac{3}{8}x + 1$

5 (a) Solve: 7x + 4 $< \frac{1}{2}(4x + 3)$. (b) Salem, Sunday and Shaka shared a sum of ₦1,100.00. for every ₦2.00 that Salem gets, Sunday gets 50 kobo and for every N4.00. Sunday gets, Shaka gets ₦2.00. Find Shaka's share. (b) By how much is the sum $3\frac{2}{3}\, and\, 2\frac{2}{5}$ less than 7?

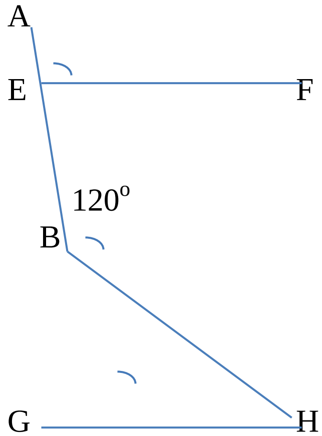
GEOMETRY AND MENSURATION

6 (a) The area of a circle is $154cm^2$. It is divided into three sectors such that two of the sectors are equal in size and the third sector is three times the size of the other two put together. Calculate the perimeter of the third sector. (Take $\pi = {}^{22}\!/_7$).

(b) A man drives from Ibadan to Oyo, a distance of 48km in 45 minutes. If he drives at 72km/h where the surface is good and 48km/h where it is bad, find the number of kilometres of good surface.

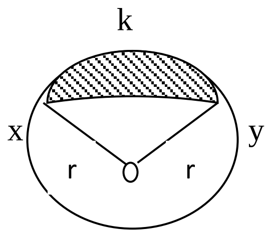
7. A boy 1.2m tall stands 6m away from the foot of a vertical lamp pole 4.2m long. If the lamp is at the tip of the pole, (a) represent this information in diagram (b) calculate the: length of the shadow of the boy cast by the lamp; (ii) angle of elevation of the lamp from the boy, correct to the nearest degree.

8. (a)



In the diagram, EF is parallel to GH, if ∠ AEF =3x°, ∠ ABC = 120o and CHG = 7xo, find the value of GHB

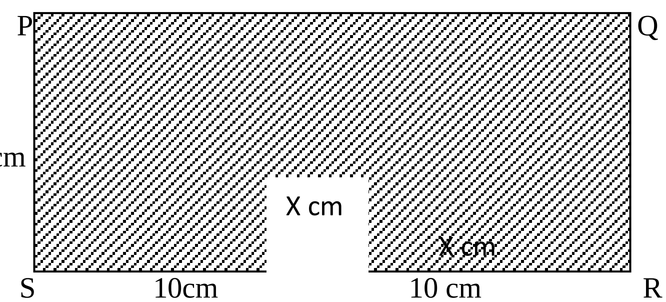
(b)



In the diagram, O is the centre of the circle radius r cm and XOY = 90°. If the area of the shaded part is 506 cm^2, calculate the value of r. (Take ∏= 22/7]

9. (a) Two isosceles triangles PQR and PQS are drawn on opposite sides of a common base PQ If PQR = 66o and PSQ = 109°, calculate the value of RQS.

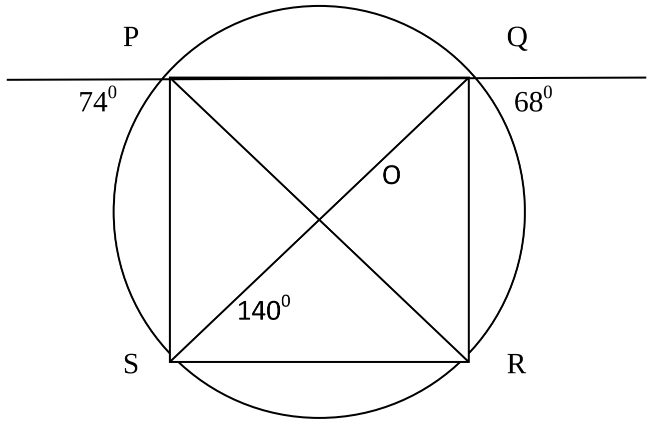
(b)



The diagram shows a rectangle PQRS from which a square of side x has been cut. If the area of the shaded portion is 484 cm^2, find the values of x.

10. (a) the ratio of the interior angle to the exterior angle of a regular polygon is 5:2. Find the number of sides of the polygon.

(b)



The diagram shows a circle PQRS with centre O, $\angle$UQR = 68^0, $\angle$TPS = 74^0 and $\angle$QSR 40^0. Calculate the value of $\angle$PRS.

11. (a) The height, h m, of a dock above sea level is given by h = 6+4 cos $(15p)^0$, 0<p<6. Find:

(b) The value of h when p = 4; correct to two significant figures, the value of p when h = 9m.

12. A trapezium PQRS is such that PQ/RS and the perpendicular from P to RS is 40cm. If IPQI = 20cm, ISRI – 60 cm, calculate, correct to 2 significant figures, the:

(a) Area of the trapezium;

(b) $\angle$QRS

STATISTICS

13. A building contractor tendered for two independent contracts, X and Y. The probabilities that he will win contact X is 0.5 and not win contract Y is 0.3. What is the probability that he will win: (a) both contracts; (b) exactly one of the contracts, (c) neither of the contracts?

14. (a) The present ages of a father and his son are in the ratio 10:3. If the son is 15years old now, in how many years will the ratio of their ages be 2:1?

(b) The arithmetic mean of x, y and z is 6 while that of x,y,z,l,u,v and w is 9. Calculate the arithmetic mean of I,u,v and w.

15. A number is selected at random from each of the sets (2, 3, 4,) and (1, 3,5). Find the probability that the sum of the two numbers is greater than 3 and less than 7.

APPENDIX II

INTERNATIONAL CENTRE FOR EDUCATIONAL EVALUATION, INSTITUTE OF EDUCATION UNIVERSITY OF IBADAN

MATHEMATICS ACHIEVEMENT TEST OF NATIONAL EXAMINATIONS COUNCIL (NECO) FOR SSS THREE

INSTRUCTION: Please tick as applicable to you and answer all the questions

TIME ALLOWED: 2 hours 30 minutes

NAME OF SCHOOL:

GENDER: Male ( ), Female ( )

AGE: 14 – 16 ( ), 17- 19 ( ), 20 – 22 ( )

NUMBERS AND NUMERATION

1(a) Simplify $\frac{a^2 - 4b^2}{a^2 - 5ab + 6b^2}$

(b) Simplify $\frac{3}{x+2} - \frac{6}{3x-1}$

2. The fourth and ninth terms of an Arithmetic Progression are -3 and 12 respectively. Find the

   (i) common difference;

   (ii) sum of the first seven terms

3. The 5^{th} term of a geometric progression (G.P). Is $\frac{2}{81}$ If the first term is 2, find the

   (i) Common ratio,

(ii) Sum of the first five terms of the G.P.

4 (a) Calculate the compound interest on ₦25, 000.00 for 3 years at 12% per annum.

(b) What number must be added to each term of the ratio 13:16 so that it becomes the ratio 4:5?

5. In a class of 40 Students, 15 like mangoes, 21 like pineapples and 6 like the two fruits.

(i) Represent the information in a Venn diagram.

(ii) How many do not like mangoes and pineapples?

(iii) What percentage of the class like mangoes only?

6. In a class of 55 students, 21 study Physics, 24 study Geography ad 23 study Economics, 6 study both Physics and Geography, 8 study both Geography and Economics and 5 study both Economics and Physics. If x study all the 3 subjects and 2x study none of the three subjects Find

(i) the value of x,

(ii) the number of students that study Physics only

(iii) the number of students that study only two subjects

ALGEBRAIC PROCESS

7. (a) Solve the inequality $\frac{5x}{8} - \frac{1}{6} \leq \frac{x}{3} + \frac{7}{24}$

(b) Solve the equation

$$\frac{(x+5)}{3} - \frac{(2x-1)}{5} - \frac{5}{6}$$

8.(a) Solve the equation: $\frac{1}{3}(2x+1) - \frac{2}{5}(x-2) = 3.$

(b) Use completing the square method to solve the equation $2x^2 - 5x + 3 = 0$

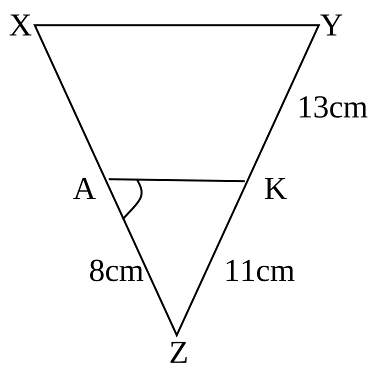
9. (a) Given that h = $\left(\frac{2p+q}{p-3q}\right)^{\frac{1}{2}}$, express p in terms of q and h.

(b) If $244_n = 1022_{four}$, find n.

GEOMETRY AND MENSURATION

10. (a) In the diagram below, XYZ=KAZ, /YZ/= 24cm, /KY/ = 13cm, /ZK/ = 11cm and /AZ/ =

8cm, Calculate /XZ/.



(b) If the volume of a cylindrical container of base radius 4cm is found to be $352cm^3$, calculate the depth of the container. (Take $\pi = \frac{22}{7}$)

11(a) A chord PQ of length 24cm is drawn in a circle of radius 37cm. If point R is the centre of the circle, find the area of △PRQ.

(b) The angle of a sector of a circle of a radius 9cm is 120^o. Calculate the perimeter of the sector, correct to three significant figures. (Take $\pi = \frac{22}{7}$)

12 (a) A both is on the same horizontal level as the foot of a cliff and the angle of depression of the boat from the top of the cliff is 60^o. If the boat is 150m away from the foot of the cliff, find the height of the cliff correct to three significant figures.

(b) A regular polygon with 12 sides is inscribed in a circle of radius 10cm. Find the area of the polygon.

13 (a) If the interior angle of a regular polygon is x times the exterior angle, express the number of sides of the polygon in terms of x.

(b) Two points C and P lie on a straight line such that C is due north of P. Another point T is 6km away from the line CP on a bearing of 150^0 from C and 40^0 from P. What is the length of the line CP correct to 3 significant figures?

STATISTICS

14 In a certain year, the government of a certain state bought 480 cars. 180 were Peugeot, 108 were Datsun, 72 were Ford, 56 were Toyota and the rest were Volkswagen.

(i) Represent the above information on a pie chart.

(ii) If one of the cars had a faulty engine on delivery, what is the probability that it is either a Toyota or a Volkswagen?

15 A basket contains 80 mangoes and 60 oranges. If the two fruits are picked one after the other without replacement, what is the probability that

(i) One of each fruit is picked?

(ii) One type of fruit is picked?

APPENDIX III
(COUNTERBALANCE ITEMS ADMINISTERED)

UNIVERSITY OF IBADAN
INSTITUTE OF EDUCATION
INTERNATIONAL CENTRE FOR EDUCATIONAL EVALUATION (ICEE)

MATHEMATICS ACHIEVEMENT TEST 1 FOR SSS THREE

INSTRUCTION: Please tick as applicable to you and answer all the questions

TIME ALLOWED: 2 hours 30 minutes

NAME OF SCHOOL:

GENDER: Male ( ), Female ( )

AGE: 14 – 16 ( ), 17- 19 ( ), 20 – 22 ( )

1 (a) Simplify, without using tables or calculator:

$$\frac{\frac{3}{4}\left(3\frac{3}{8}+1\frac{5}{8}\right)}{2\frac{1}{8}-1\frac{1}{2}}$$

(b) Given that $\log_{10}{}^{2} = 0.3010$ and $\log_{10}{}^{3} = 0.4771$, evaluate correct to 2 significant figures and without using tables of calculator, $\log_{10} 1.125$.

2. The fourth and ninth terms of an Arithmetic Progression are -3 and 12 respectively. Find the
   (i) common difference;
   (ii) sum of the first seven terms

3. (a) Without using Mathematical tables or calculators, simplify: $3\frac{4}{9}+(5\frac{1}{3}-2\frac{3}{4})+5\frac{9}{10}$
   (b) If 124, = 232_{five}, find n.

4 (a) Calculate the compound interest on ₦25, 000.00 for 3 years at 12% per annum
   (b) What number must be added to each term of the ratio 13:16 so that it becomes the ratio 4:5?

5 (a) Solve 7x + 4 $< \frac{1}{2}(4x+3)$. (b) Salem, Sunday and Shaka shared a sum of ₦1,100.00. for every ₦2.00 that Salem gets, Sunday gets 50 kobos and for every N4.00. Sunday gets, Shaka gets ₦2.00. Find Shaka's share.

(b) By how much is the sum $3\frac{2}{3}$ *and* $2\frac{2}{5}$ less than 7?

6. In a class of 55 students, 21 study Physics, 24 study Geography ad 23 study Economics, 6 study both Physics and Geography, 8 study both Geography and Economics and 5 study both Economics and Physics. If x study all the 3 subjects and 2x study none of the three subjects Find

(i) the value of x,

(ii) the number of students that study Physics only

(iii) the number of students that study only two subjects

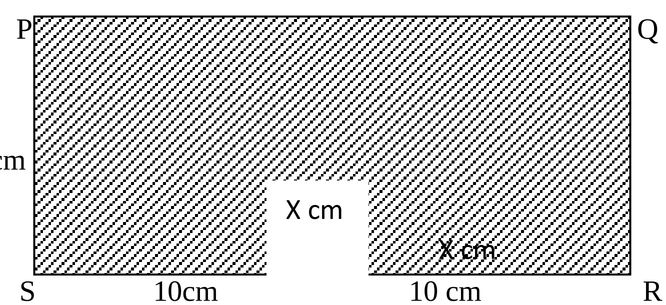
7. A boy 1.2m tall, stands 6m away from the foot of a vertical lamp pole 4.2m long. If the lamp is at the tip of the pole, (a) represent this information in diagram (b) calculate the: (i) length of the shadow of the boy cast by the lamp; (ii) angle of elevation of the lamp from the boy, correct to the nearest degree.

8. (a) Solve the equation: $\frac{1}{3}(2x+1)-\frac{2}{5}(x-2)=3.$

(b) Use completing the square method to solve the equation $2x^2 - 5x + 3 = 0$

9. (a) Two isosceles triangles PQR and PQS are drawn on opposite sides of a common base PQ If PQR = 66o and PSQ = 109^o, calculate the value of RQS.

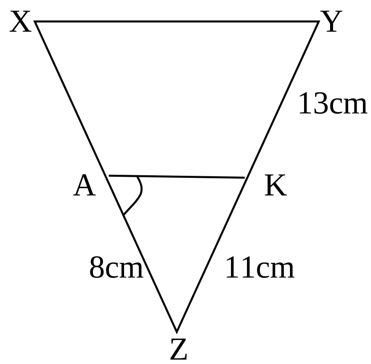
(b)



The diagram shows a rectangle PQRS from which a square of side x has been cut. If the area of the shaded portion is 484 cm^2, find the values of x.

10. (a) In the diagram below, XYZ=KAZ, /YZ/= 24cm, /KY/ = 13cm, /ZK/ = 11cm and /AZ/ =

8cm, Calculate /XZ/.



(b) If the volume of a cylindrical container of base radius 4cm is found to be $352cm^3$, calculate the depth of the container. (Take $\pi = \frac{22}{7}$)

11. (a) The height, h m, of a dock above sea level is given by h = 6+4 cos $(15p)^0$, 0<p<6. Find:

(b) The value of h when p = 4; correct to two significant figures, the value of p when h = 9m.

12 (a) A both is on the same horizontal level as the foot of a cliff and the angle of depression of the boat from the top of the cliff is 60^o. If the boat is 150m away from the foot of the cliff, find the height of the cliff correct to three significant figures.

(b) A regular polygon with 12 sides is inscribed in a circle of radius 10cm. Find the area of the polygon.

13. A building contractor tendered for two independent contracts, X and Y. The probabilities that he will win contact X is 0.5 and not win contract Y is 0.3. What is the probability that he will win: (a) both contracts; (b) exactly one of the contracts, (c) neither of the contracts?

14 In a certain year, the government of a certain state bought 480 cars. 180 were Peugeot, 108 were Datsun, 72 were Ford, 56 were Toyota and the rest were Volkswagen.

(i) Represent the above information on a pie chart.

(ii) If one of the cars had a faulty engine on delivery, what is the probability that it is either a Toyota or a Volkswagen?

15. A number is selected at random from each of the sets (2, 3, 4,) and (1, 3,5). Find the probability that the sum of the two numbers is greater than 3 and less than 7.

APPENDIX IV
(COUNTERBALANCE ITEMS ADMINISTERED)

UNIVERSITY OF IBADAN
INSTITUTE OF EDUCATION
INTERNATIONAL CENTRE FOR EDUCATIONAL EVALUATION,

MATHEMATICS ACHIEVEMENT TEST II FOR SSS THREE

INSTRUCTION: Please tick as applicable to you and answer all the questions

TIME ALLOWED: 2 hours 30 minutes

NAME OF SCHOOL:

GENDER: Male ( ), Female ( )

AGE: 14 – 16 ( ), 17- 19 ( ), 20 – 22 ( )

1(a) Simplify $\frac{a^2 - 4b^2}{a^2 - 5ab + 6b^2}$

(b) Simplify $\frac{3}{x+2} - \frac{6}{3x-1}$

2 (a) Simplify $3\sqrt{75}-\sqrt{12}+\sqrt{108}$, leaving the answer in surd from (radicals)

(b) Without using tables or calculator, simplify: $\frac{0.6x32x0.004}{1.2x0.008x0.16}$

leaving the answer in standard form (scientific notation)

3. The 5th term of a geometric progression (G.P). Is $\frac{2}{81}$ If the first term is 2, find the

(i) Common ratio,
(ii) Sum of the first five terms of the G.P.

4 (a) Solve the simultaneous equations:

$^1/_x + ^1/_y = 5$

$^1/_y - ^1/_x = 1$

(b) Solve the inequality: $4+\frac{3}{4}(x+2)\le\frac{3}{8}x+1$

5. In a class of 40 Students, 15 like mangoes, 21 like pineapples and 6 like the two fruits.

(i) Represent the information in a Venn diagram.

(ii) How many do not like mangoes and pineapples?

(iii) What percentage of the class like mangoes only?

6 (a) The area of a circle is 154cm^2. It is divided into three sectors such that two of the sectors are equal in size and the third sector is three times the size of the other two put together. Calculate the perimeter of the third sector. [Take $\pi = 22/7$).

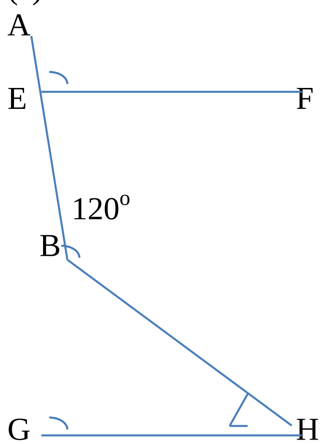
(b) A man drives from Ibadan to Oyo, a distance of 48km in 45 minutes. If he drives at 72km/h where the surface is good and 48km/h where it is bad, find the number of kilometres of good surface.

7. (a) Solve the inequality $\frac{5x}{8}-\frac{1}{6}\le\frac{x}{3}+\frac{7}{24}$

(b) Solve the equation

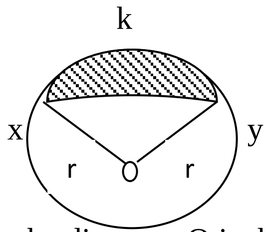
$$\frac{(x+5)}{3}-\frac{(2x-1)}{5}-\frac{5}{6}$$

8. (a)



In the diagram, EF is parallel to GH, if AEF =3x°, ABC = 120o and CHG = 7xo, find the value of GHB

(b)



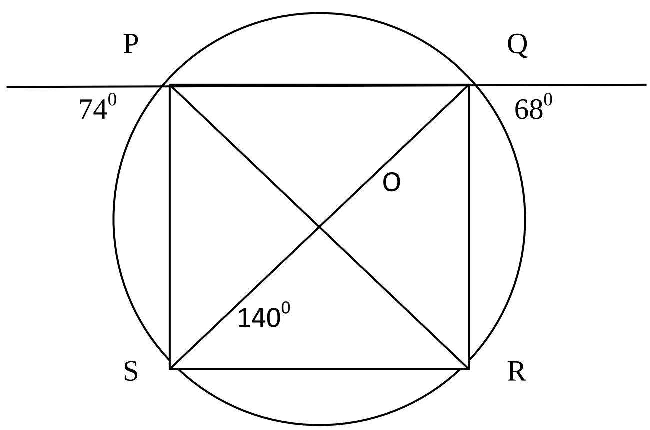
In the diagram, O is the centre of the circle radius r cm and XOY = 90^o. If the area of the shaded part is 506 cm^2, calculate the value of r. (Take ∏= 22/7]

9. (a) Given that h = $\left(\frac{2p+q}{p-3q}\right)^{\frac{1}{2}}$, express p in terms of q and h.

(b) If $244_n = 1022_{four}$, find n.

10. (a) the ratio of the interior angle to the exterior angle of a regular polygon is 5:2. Find the number of sides of the polygon.

(b)



The diagram shows a circle PQRS with centre O, ∠UQR = 68^0, ∠TPS = 74^0 and ∠QSR 40^0. Calculate the value of ∠PRS.

11 (a) A chord PQ of length 24cm is drawn in a circle of radius 37cm. If point R is the centre of the circle, fi△d the area of PRQ.

(b) The angle of a sector of a circle of a radius 9cm is 120^o. Calculate the perimeter of the

sector, correct to three significant figures. (Take $\pi = \frac{22}{7}$)

12. A trapezium PQRS is such that PQ/RS and the perpendicular from P to RS is 40cm. If IPQI = 20cm, ISRI – 60 cm, calculate, correct to 2 significant figures, the:

(a) Area of the trapezium;

(b) $\angle QRS$

13 (a) If the interior angle of a regular polygon is x times the exterior angle, express the number of sides of the polygon in terms of x.

(b) Two points C and P lie on a straight line such that C is due north of P. Another point T is 6km away from the line CP on a bearing of 150^0 from C and 40^0 from P. What is the length of the line CP correct to 3 significant figures?

14. (a) The present ages of a father and his son are in the ratio 10:3. If the son is 15years old now, in how many years will the ratio of their ages be 2:1?

(b) The arithmetic mean of x, y and z is 6 while that of x,y,z,l,u,v and w is 9. Calculate the arithmetic mean of I,u,v and w.

15. A basket contains 80 mangoes and 60 oranges. If the two fruits are picked one after the other without replacement, what is the probability that

(i) One of each fruit is picked?

(ii) One type of fruit is picked?

APPENDIX V

DEVELOPED MARKING GUIDE ADAPTED FROM WEAC BY THE RESEARCHER

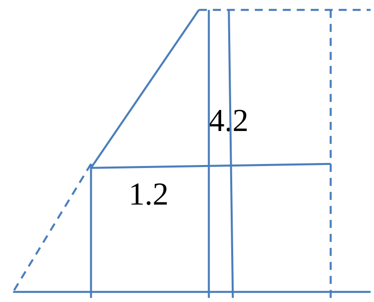
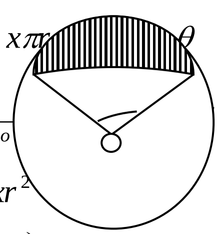
No.	Answer	Details	Marks
1a.	**6**	$\frac{\left(\frac{3}{4}3\frac{3}{8}+1+\frac{5}{8}\right)}{2\frac{1}{8}-1\frac{1}{2}}$ $= \frac{\frac{3}{4}\left(\frac{27}{8}+\frac{13}{8}\right)}{\frac{17}{8}-\frac{3}{2}}$ $= \frac{\frac{3}{4}\left(\frac{27+13}{8}\right)}{\frac{17-3}{8}}$ $= \frac{3}{4}\left(\frac{40}{8}\right) \div \frac{5}{8}$ $= \frac{120}{32} x5$ = 6	8marks
1b.	0.0512	$Log_{10}{}^{2}$ = 0.3020 $Log_{10}{}^{3}$ = 0.4771 $Log_{10}{}^{1.125}$ 10 = Log_{10} 1125/1000 Putting $\frac{1125}{1000} = \frac{225}{200} = \frac{45}{40} = \frac{9}{8}$ $Log_{10}{}^{9/8}$ = $Log_{10}{}^{9}$-$Log_{10}{}^{8}$ = 2$Log_{10}{}^{3}$-3$Log_{10}{}^{2}$ = 2(0.4771)-3(0.3010) = 0.9542 – 0.903 = 0.0512	
2a.	$19\sqrt{3}$	$3\sqrt{75}$ - $\sqrt{12}$ + $\sqrt{108}$ = $3\sqrt{25x3}$ - $\sqrt{4x3}+\sqrt{36x3}$ = 3 x $5\sqrt{3}$ - $2\sqrt{3}$ + $6\sqrt{3}$ = $15\sqrt{3}$ - $2\sqrt{3}$ + $6\sqrt{3}$ = $\sqrt{3}$ (19) = $19\sqrt{3}$	8
2b.	5 x 10^{1}		

		$\frac{0.6x32x0.004}{1.2x0.008x016}$ $= \frac{6x10^{-1}x32^4x4x10^{-3}}{12_2x10^{-1}x8x10^3x16_4x10^{=2}}$ $= \frac{4}{2x4x10^{=2}}$ $= \frac{1}{2}x\frac{1}{10^{-2}}$ $= 0.5 \text{ x } (10^{-2})^{-1}$ $= 5 \text{ x } 10^{2-1}$ $= 5 \text{ x } 10^1$	
3a.	$7\frac{7}{30}$	$3\frac{4}{9} \div \left(5\frac{1}{3} - 2\frac{3}{4}\right) + 5\frac{9}{10}$ $= \frac{31}{9} \div \left(\frac{16}{3} - \frac{11}{4}\right) + \frac{59}{10}$ $= \frac{31}{9} \div \left(\frac{64-33}{12}\right) + \frac{59}{10}$ $= \frac{31}{9} \div \left(\frac{31}{12}\right) + \frac{59}{10}$ $= \left(\frac{31}{9} \div \frac{31}{12}\right) + \frac{59}{10}$ $= \frac{31}{9} \times \frac{12}{31} + \frac{59}{10}$ $= \frac{4}{3} + \frac{59}{10}$ $= \frac{40+177}{30} = \frac{217}{30}$ Ans = $7\frac{7}{30}$	
3b.	-7 or +9	$124_n = 232_5$ $= 1 \text{ x } n^2 + 2 \text{ x } n^1 + 4 \text{ x } n^o \quad = 2 \text{ x } 5^2 + 3 \text{ x } 5^1 + 2 \text{ x } 5^o$ $= n^2 + 2n + 4 \text{ x } 1 \quad = 2 \text{ x } 25 + 3 \text{ x } 5 + 2 \text{ x } 1$ $= n^2 + 2n + 4 \quad = 50 + 15 + 2$ $= n^2 + 2n + 4 \quad = 67$ $= n^2 + 2n + 4 - 67 \quad = 0$ $= n^2 + 2n - 63 \quad = 0$ $= n^2 + 9n - 7n - 63 \quad = 0$ $= n(n+9) - 7(n+9) \quad = 0$ $= (n-7)(n+9) \quad = 0$ Either $n = -7 = 0$ or $n + 9 \quad = 0$	8marks
4a.	x = ½ and	$/_x + {}^1/_y = 5$ …………. (1)	

	$y = {}^{1}/_{3}$	${}^{1}/_{y} - {}^{1}/_{x} = 1$ …………. (2) Substitute p and q for ${}^{1}/_{x}$ and ${}^{1}/_{y}$ respectively $p + q = 5$ .... (3) $q - p = 5$ …………. (4) From (4), $q = 1 + p$ ………… (5) Substitute 1+ P for q in (3) $p + 1 + p = 5$ $2p + 1 = 5$ $2p = 5 - 1$ $2p = 4$ $p = ½$ $= 2$ Using (5), $q = 1 + 2 = 3$ But $p = {}^{1}/_{x} = 2$ ${}^{1}/_{x} = 2$ $x = ½$ also, $q = {}^{1}/_{y} = 3$ ${}^{1}/_{y} = 3$ $Y = ½$ Hence, x = ½ and $y = {}^{1}/_{3}$	
4b.	$\Rightarrow x \le -12$	$4 + \frac{3}{4}(x+2) \le \frac{3}{8}x + 1$ Multiply through by the LCM of 4 and 8, to clear fractions: $8x4 + 8x\frac{3}{4}(x+2) \le 8x\frac{3}{8}x + 8$ $\Rightarrow 32 + 6(x+2) \le 3x + 8$ $\Rightarrow 32 + 6x + 12 \le 3x + 8$ $\Rightarrow 6x - 3x \le 8 - 44$ $\Rightarrow 3x \le -36$ $\Rightarrow \frac{3x}{3} \le \frac{-36}{3}$ $\Rightarrow x \le -12$	8marks
5a.	₦100	$7x+4<1/2(4x+3)$ $7x+4<2x+3/2$ $7x-2x<3/2-4$ $7x-2x<3-8/2$ $5x<-5/2$ $x<-5/2:-5$ $x<-5/2x1/8$ $x<-1/2$ when Salam gets ₦2 --→ Sunday gets 50k. :- since ₦1 = 100k :- Salam get 200k --→Sunday gets 50k Also	

		Sunday gets ₦4 --→Shaka gets N2 Since ₦1 = 100k Sunday gets 400k -→Shaka gets 20k whensundy gets 50k -→ Salam gets 200k thenwhen Sunday gets 400 --→ Salam gets 400k then Shaka gets 200k Thus Salam Sunday Shaka 1600k 400k 200k Since N1 = 100k Salam Sunday Shaka ₦16 ₦4 ₦2 8 2 1 Hence Shaka's share $\frac{1}{11} xN1100$ = ₦100	8marks
5b.	$1\frac{2}{5}$	Let the sum of $3\frac{2}{3}$ and $2\frac{1}{5}$ be less than 7 by x. Then x = $7-\left(3\frac{2}{3}\right)+\left(2\frac{1}{5}\right)$ $7-\left(\frac{11}{3}+\frac{11}{5}\right)$ $7-\frac{88}{15}$ $\frac{17}{15}=1\frac{2}{15}$ Hence the sum of $3\frac{2}{3}$ and $2\frac{1}{5}$ is less than Ans = $1\frac{2}{5}$	8marks
6a.	47cm	Area of a circle = πr^2 154 = πr^2 $r^2 = \frac{154}{\pi}$ $r^2 = \frac{154}{22} x7$ $r^2 \sqrt{49}$ R = 49 = 7cm Since it is divided into 3 sectors then $8x^o = 360$ xo = 360/8 xo = 45 The 3rd sector is 3 times	

		The size of the other 2 put together i.e. 3(45+45) $3(90) = 270^o$ Perimeter of sector = 2r +1 where L = length of the arc L = $\frac{\Theta}{360} x 2\pi r$ = $\frac{270}{360} x 2 x \frac{22}{7} x 7$ = 33cm So, perimeter 2r+L = 2(7) +33 = 14+33 = 47cm	8marks
6b.	36km	Total distance travelled = 48km Total time taken = 45 minutes = $\frac{45}{60}$ = 0.75n Let the man travel a distance of x km on a good surface. Then he travels a distance of (4-x) km on a bad surface. Let the time taken to travel on good surface be th. Then the time taken to travel on a bad surface = (0.75 - t)h. Using speed = $\frac{dis\tan ce}{time}$ In each case: on a good surface; 72 = $^x/_1$ x = (72t) km ………….. (1) On bad surface; 48 = $\frac{48-x}{0.75-1}$ 48 (0.75 - t) = 48 - x 38 – 4t = 48 – x …………. (2) Substitute 72t for x in (2) 36 – 4t = 4 – 72t 72t – 48t = 48 – 36 24t = 12 t = $^{12}/_{24}$ = 0.5n From (i), x = 72 x 0.5 = 36 km	8marks

7a. 7b. i and ii	$\tan\theta = 0.5$ $\theta = 27^o$ x= 2.4m	 θ x6 $\tan\theta = \frac{4.2}{6+x}$ $\tan = \frac{1.2}{x}$ tanθ = tan θ (ii) $\frac{4.2}{6+x} = \frac{1.2}{x}$ 4.2x = 1.2 (6+x) tan θ = 1 4.2x -1.2x = 7.2 tan θ = 0.5 3.0 x = 7.2 $\theta = \tan^{-1} 0.5$ $x = \frac{7.2}{3.0}$ $\theta = 27^o$ (angle of elevation) x= 2.4m. (length of the shadow of the boy cast by the lamp)	8marks
8a 8b.	75 42cm	Then $3x^o + (180^o - 7x^o) = 120^o$ $3x^o - 7x^o + 180^o = 120^o$ $180^o - 120^o = 7x^o - 3x^o$ $60^o = 4x^o$ $x^o = \frac{60^o}{4}$ $= 15^0$ Hence, ∠GHB = $180^o - 7 \times 15^o$ $= 180^o - 105^o$ $= 75^o$ Area of the shaded segment = area of sector XOY – area of OXY $504 = \frac{\theta}{360} x\pi r$ θ  $504 = \frac{90}{360^o}$ $\sin 90^o$ $= \frac{1}{4} x \frac{22}{7} x r^2$ $= r^2\left(\frac{22}{28} - \frac{1}{2}\right)$ $= r^2\left(\frac{11}{14} - \frac{1}{2}\right)$	

		$= r^2\left(\frac{11-7}{14}\right)$ $= \quad 504 = r^2 x \frac{4}{14}$ $r^2 \quad = \quad \frac{504x14}{4}$ = 126 x 14 = 1764 r = 42cm	8marks
9a. 9b.	101.5^o $\Rightarrow$ x = 14 or x = 6	In the diagram above. R $\quad a_1 = a_2$ (base of angles of isosceles.):. $a_1 + a_2 = 2a_1$ $2a_1 + 109_o = 180^o$ (sum of angles of) $2a_1 = 180^o - 109^o$ P= 71^o $a_1 = 71^o/2$ = 35.5^o Hence ∠ RQS = 66^o = $66^o + 35.5^o$ = 101.5 Area A, of rectangle PQRS = 20 (20 + x) cm^2 Area A^2 of square = xXx cm^2 Area of shaded portion = A1 – A^2 = 484 $\Rightarrow$ 20 (20 + x) – xXx = 484 $\Rightarrow 400 + 20x - x^2 = 484$ $\Rightarrow 0 = x^2 - 20x + 484 - 400$ $\Rightarrow x^2 - 20x + 84 = 0$ $\Rightarrow x^2 - 6x - 14x + 84 = 0$ $\Rightarrow x(x-6) - 14(x-6) = 0$ $\Rightarrow (x-14)(x-6) = 0$ $\Rightarrow x - 14 = 0$ or $x - 6 = 0$ $\Rightarrow$ x = 14 or x = 6	8marks
10a.	7	Let the diagram below represent a section of the polygon Also, let: i represent the size of an interior angle; ℓ represent the size of an exterior angle Then $\frac{i}{e} = \frac{5}{2}$ $\Rightarrow$ i = $\frac{5}{2}\ell$	

		But i + e = 180^0 (sum of angles on a straight line) Substitute $\frac{5}{2}\ell$ for i in the equation i + e = 180^0 $\Rightarrow \frac{5}{2}\ell + e = 180^0$ $\Rightarrow \frac{7}{2}\ell = 180^0$ $\Rightarrow \ell = \frac{180^0 x 2}{7}$ $\Rightarrow \frac{360^0}{7}$ Number of sides of the polygon $= \frac{360^0}{\text{side of one exterior angle}}$ $= 360 \div \frac{360}{7}$ $360x\frac{7}{360} = 7$	
10b.	46^0	In the diagram above, <PSR 68^0(interior angle of a cyclic quad = opp. exterior angle) <PSQ - 68^0 - 40^0 - 28^0 <PRQ = 28^0 (angles in same segment) <SRQ - 74^0(interior angle of a cyclic quad = opp. exterior angle) Hence, <PRS = 74^0 - 28^0 = 46^0	8marks
11a.	8	h = 6 + 4 cos $(15p)^0$ (given) (i) When p = 4 h = 6 = 4 cos $(15 \times 4)^0$ = 6 + 4 cos 60^0 = 6 + 4 x 0.5 = 6 + 2 = 8	
11b.	13	When h = 9m, the given equation becomes 9 = 6 + 4 cos $(15p)^0$ $\Rightarrow$ 9 -6 = 4 cos $(15p)^0$ $\Rightarrow$ 3 = 4 cos $(15p)^0$ $\Rightarrow \frac{3}{4} = \cos (15p)^0$ $\Rightarrow (15p)^0 = \cos^{-1}\left(\frac{3}{4}\right) = 41.4^0$	

		$\Rightarrow (15p)^0 = 414^0$ $\Rightarrow P = \frac{41.4^0}{15} = 2.76^0$ $\approx 2.8^0 (2s.f)$	8marks
12a.	$1600cm^2$	Area A of trapezium PQRS is given by $A = \frac{1}{2}(PQ + RS)x40$ $= \frac{1}{2}(20+60)x40$ $= \frac{1}{2}x80x40$ $= 1600cm^2$	
12b.	76^0 (2 s.f.)	In ▲ SPT, $/SP/^2 = /ST/^2 + /TP/^2$ (Pythagoras' theorem) $/ST/^2 = /SP/^2 - /TP/^2$ $= 50^2 - 40^2$ $= 2500 – 1600$ $= 900$ $/ST/ = \sqrt{900}$ $/ST/ = \sqrt{900} = 30cm$ $\Rightarrow$ /UR/ = 60 – (ISTI + 20)cm = 60-(30+20)cm = 10cm Let <QRS = α^0 Then tan $\alpha^0 = \frac{40}{10} = 4$ $\Rightarrow = \tan^{-1}(4)$ $\approx 75.96^0$ Hence, <QRS = 76^0 (2 s.f.)	8marks
13a 13i	0.35 0.5 & 0.15	Let A and B denote the events that the man wins contracts X and Y respectively. Then P(A) = 0.5 and P(A) = 1 – 0.5 = 0.5; P(B) = 0.3 and P(B) = 1 – 0.3 = 0.7 The probability that the man wins both contract is P(A) = P(B) = 0.5 x 0.7 = 0.35 The probability that the man wins exactly one of the contracts is	

and ii		P(A) x P(B') + P(A') x P(B) = 0.5 x 0.3 + 0.5 x 0.7 = 0.15 + 0.35 = 0.5 Neither of the contracts means not X and Y. Hence the probability that the man wins neither of the contracts is P(A') + P(B) = 0.5 x 0.3 = 0.15	8marks
14a.	20	Present ages of a father and his son are in ratio 10:3 Son age is 15years Fathers age is unknown say x then Father to Son ratio $\frac{10}{5} = \frac{x}{15}$ 3x = 150 x = 150/3 x = 30 so fathers age = 50years Because age cannot reduce, in y years' time Father's age = 50+y So in y years' time $\frac{2}{1} = \frac{50+y}{15+y}$ 2(115+y)=50+y 30+2y=50+y 2y-y = 50-30 y = 20 so in 20years time the ratio of their age will be 2:1	
14b.	11.25	$\frac{x+y+z+t+u+v+w}{7} = 9$...........(2) $\frac{t+u+v+w}{4} = x$...........(3) from equation (1) x+y+z=18 ………… (4) putting this in equation (2) $\frac{18+t+u+v+w}{7} = 9$ 18+t+u+v+w = 63 t+u+v+w = 63-18 t+u+v+w = 45 …….(3) fromequation (3) $\frac{t+u+v+w}{4} = x$	8marks

<table>
<tr><td></td><td></td><td>putting equation (3) in (2)
$\frac{45}{4} = 4$
$x = 11.25$
So the arithmetic means of t,u,v and w is
11.25 i.e $\frac{t+u+v+w}{4} = 11.25$</td><td></td></tr>
<tr><td>15.</td><td>$\frac{4}{9}$</td><td>

+		1	3	5
2		3	5	7
3		4	6	8
4		5	7	9

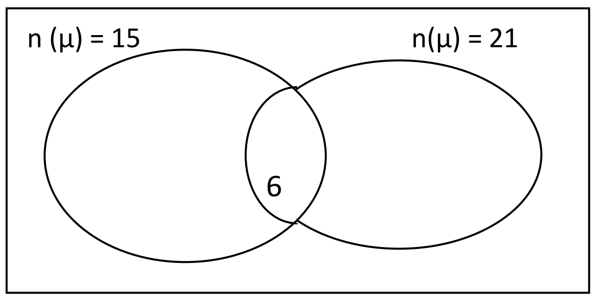
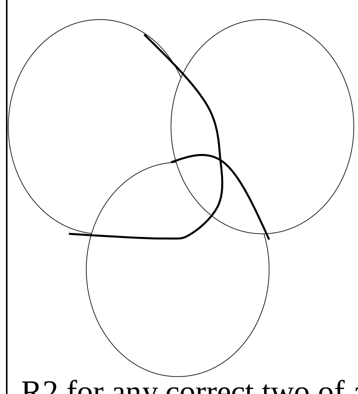
Let: E denote the event that a sum is greater than 3 but less than 7;
S denote the sample space; P(E), the probability that the sum of two numbers is greater than 3 and less than 7.
Then n(E) = 4 and n (S) = 9
Hence, P (E) = $\frac{n(E)}{n(S)} = \frac{4}{9}$</td><td>8marks</td></tr>
</table>

APPENDIX VI

DEVELOPED MARKING GUIDE ADAPTED FROM NECO BY THE RESEARCHER

No.	**Answer**	**Details**	**Marks**
1a	$\frac{a+2b}{a-3b}$	M1 for factorization and simplifying $= \frac{(a-2b)(a+2b)}{(a-2b)(a-3b)}$ A1 for $\frac{a+2b}{a-3b}$	
1b.	$\frac{3(x-5)}{(x+2)(3x-1)}$	Simplify $\frac{3}{x+2} - \frac{6}{3x-1}$ LCM= (x+2)(3x-1) B_1 $= \frac{3(3x-1)-6(x+2)}{(x+2)(3x-1)}$ $= \frac{9x-3-6x-12}{(x+2)(3x-1)}$ M_1 $= \frac{3(x-5)}{(x+2)(3x-1)}$ A_1	8marks
2(i)	-12	a + 3d = -3 …………………..(i) a + 8d = 12 …………………..(ii) B1 for either equation (i) or (ii) Subtract equation (i) from (ii) A1 for d = 3 A1 for a = -12	
ii.	24	$S_7 = \frac{7}{2}[2x-12+(7-1)3]$ M1 for finding S_7 $= \frac{7}{2}(-24+18)$	8marks

		A1 for = -21	
3a.	$\frac{1}{3}$	$T_5 = ar^4$ $2r^4 = \frac{20}{81}$ 0 $r^4 = \frac{1}{81}$ M_1 1 $r^4 = \left(\frac{1}{3}\right)^4$ $r = \frac{1}{3}$ A_1 2	8marks
3b.	$2\frac{80}{81}$	$S_s = \frac{a(1-r^3)}{(1-r)}$ $= \frac{2\left[1-\left(\frac{1}{3}\right)^5\right]}{1-\frac{1}{3}}$ M_1 1 $= \frac{2\left(1-\frac{1}{243}\right)}{\frac{2}{3}}$ M_1 2 $= 2\left(\frac{242}{243}\right)\frac{3}{2}$ $= \frac{242}{81} or 2\frac{80}{81}$ $A_1$3	
4a.	₦ 10,123.20	$A=P\left(1+\frac{r}{100}\right)^n$ $=25000\left(1+\frac{=12}{100}\right)^3$ M_1 = ₦ 25000 $(1+0.12)^3$ M_1 = 25000 $(1.12)^3$ = 35123.20 A_1 C.I= A – P = 35123.20 – 25000.00 M_1 = ₦ 10,123.20 A_1 ALTER $1_1 = \frac{N25{,}000}{100} \times 12 = N3{,}000$	

		$1_2 = \frac{N28,000}{100} \times 12 = N3,360$ $1_3 = \frac{N31,360}{100} \times 12 = N3,763.20$ $\therefore 1 = 1_1 + 1_2 + 1_3 = N10,123.20$ Accept any other correct method that will lead to the correct answer	
4b.	-1	Let the number be x $\frac{13+x}{16+x} = \frac{4}{5}$ M1 for the equation 65 + 5x = 64 + 4x M1 for solving A1 for x = -1	8marks
5i,ii and iii	30 and 22.5% or equivalent	i.  ii. a+6+15+x = 40 M1 for simplifying x = 40 – 30 A1 for x = 10 iii. M1 for finding percentage $\frac{9}{40} x100$ A1 for 22.5% or equivalent	8marks
6i, ii and iii		B2 for correct diagram  R2 for any correct two of a, b, and c.	

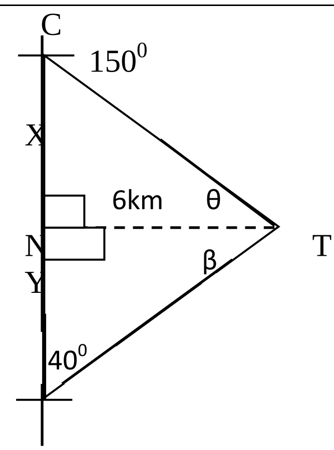
		6-x+x+5-x+a = 21 a = 10 + x 6-x+x-8-x+6 = 24 5 = x+x+8 = x+c = 23 c = 10 + x 30+x+6+8+5-2x+2x = 55 x = 2 M1 for solving A1 for x = 2 ii. Physics only = 10 + 2 B1 for 12 students iii. Only two subjects = (6-2) + (8-2) + (5-2) B1 for 1 students	8marks
7a.	$1\frac{4}{7}$	$\frac{5}{8}x-\frac{1}{6}\leq\frac{1}{3}x+\frac{7}{24}$ L.C.M. = 24 Multiply through by 24. $24\times\frac{5}{8}x-\frac{1}{6}\leq\frac{1}{3}x+\frac{7}{24}\times 24$ $15x - 4 \leq 8x+7$ M_1 $15x - 8x \leq 11$ M_1 $7x\leq 11$ $x\leq\frac{11}{7}$ $x\leq 1\frac{4}{7}$ A_1	8marks
7b.	15.5	$\frac{10(x+5)-6(2x-1)}{30}=\frac{5}{6}$ 10x + 50-12x+16=25 MI for mult by LCM = 30 MI for opening bracket -2x=25-56 MI for solving x = $\frac{31}{2}$ AI for x = $15\frac{1}{2}$ or 15.5	

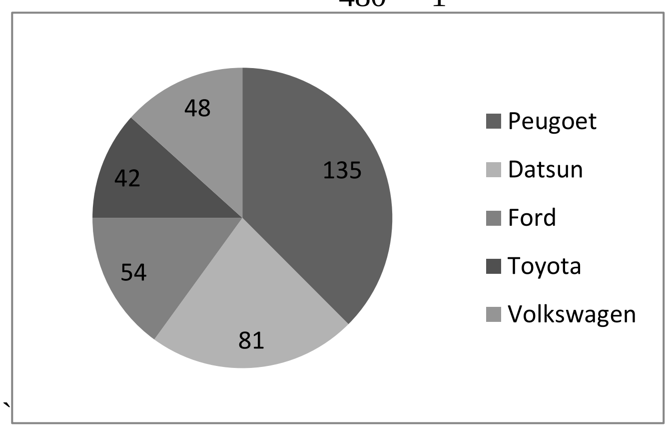
8a 8b.	7 $\frac{3}{2}$ and 1	$\frac{1}{3}(2x+1)-\frac{2}{5}(x-2)=3$ 5 (2x+1)-6(x-2)=45 M1 10x+5-6x+12=45 4x-45-17 M1 4x=28 X=7 $2x^2 - 5x + 3 = 0$ $x^2-\frac{5x}{2}+\frac{3}{2}=0$ M1 M1 for dividing by 2. M1 $x^2-\frac{5x}{2}+\left(-\frac{5}{4}\right)^2=\left(-\frac{5}{4}\right)^2-\frac{3}{2}$ MI for adding $\left(\frac{-5}{4}\right)^2$ $\frac{25}{16}-\frac{3}{2}$ $\left(x-\frac{5}{2}\right)^2=\frac{1}{16}$ $x-\frac{5}{4}=\pm\sqrt{\frac{1}{16}}=\pm\frac{1}{4}$ M1 for taking the square root $x=\frac{5}{4}-\frac{1}{4}$ or $\frac{5}{4}+\frac{1}{4}$ M1 for simplifying A1 for x = $\frac{3}{2}$ and 1	 8marks
9a.	$\frac{q(3h^2+1)}{h^2-2}$	$h^2=\frac{2p+q}{p-3q}$ MI for squaring both sides MI for cross multiplying $h^2(p-3q)=2p+q$ $ph^2-3qh^2=2p+q$ $P(h^2-2)=q(3h^2+1)$ MI for simplifying A1 for P = $\frac{q(3h^2+1)}{h^2-2}$	

9b.	5	$244_n = 1022_{four}$ find n, $2 \times n^2 + 4 \times n^1 + 4 \times n^0$ $1 \times 4^3 + 0 \times 4^2 + 2 \times 4^1 + 2 \times 4^0$ M1 for LHS or RHS in base 10 $2n^2 + 4n + 4 = 64 + 8 + 2$ M_1 A1 for the equation $2n^2 + 4n + 4 – 74 = 0$ $2n^2 + 4n -70 = 0$ $n^2 + 2n – 35 = 0$ M_1 M1 for solving equation $(n - 5) (n + 7) = 0$ A1 for n = 5 or -7, M_1 n cannot be negative A1 for n = 5	8marks
10a.	33cm	$\frac{YZ}{AZ} = \frac{XZ}{KZ}$ M1 for correct ratio $\frac{24}{8} = \frac{XZ}{11}$ M_1 $XZ = \frac{24x11}{8}$ M1 for substitutes and simplifying M_1 Al for XZ = 33cm	8marks
10b.	7cm	Volume = $\pi r^2 h$ M_1 352 = $\frac{22}{7} x4^2 xh$ M_1 M1 for substitution h = $\frac{352x7}{22x16}$ M1 for simplifying = $\frac{2464}{352}$ M_2 A1 for h = 7cm	

11a.	$420cm^2$	h2 = $37^2 - 12^2$ (Pythagoras) M1 for using Pythagoras correctly M$_1$ = 1369 – 144 h^2 = 1225 h = $\sqrt{1225}$ A1 for h = 35cm M$_1$ Area of ▲PQR = $\frac{1}{2}$ bh = ½ x 24 x 35 MI for substituting A1 for $420cm^2$ alternative Areaof △PRQ $\sqrt{5(5-a)(5-b)}$ = $\frac{1}{2}(a+b+c)$ $\frac{1}{2}(37+37+24)$ 49 B1 for 5 = 49 Area = $\sqrt{49(49-37)(49-37(49-24)}$ MI for Substituting = $\sqrt{49+12+12+25}$ = $\sqrt{176400}$ MI for simplifying A1 for $420cm^2$ h2 = $37^2 - 12^2$ (Pythagoras) M1 for using Pythagoras correctly = 1369 – 144 h^2 = 1225 h = $\sqrt{1225}$ A1 for h = 35cm Area of ▲PQR = $\frac{1}{2}$ bh = ½ x 24 x 35 MI for substituting A1 for $420cm^2$ alternative Areaof △PRQ	

		$\sqrt{5(5-a)(5-b)}$ $= \frac{1}{2}(a+b+c)$ $\frac{1}{2}(37+37+24)$ 49 B1 for 5 = 49 Area = $\sqrt{49(49-37)(49-37(49-24)}$ MI for Substituting $= \sqrt{49+12+12+25}$ $= \sqrt{176400}$ MI for simplifying A1 for 420cm^2	
11b.	36.9cm	The perimeter of a sector $= \frac{\Theta}{360} x 2\pi r + 2r$ $= \frac{120}{360} x 2 x \frac{22}{7} x 9 + (2x9)$ Mi For Correct Substitution MI for simplifying = 18.857+18 A1 for 36.857cm A1 for 36.9cm	8marks
12a.	260m	$60 = \frac{h}{150}$ an 60 x 150 MI for correct trig ratio MI for simplifying A1 for 259.808 or 259.8 A1 for h = 260m	

12b.	300cm^2	For a polygon with 12 regular sides, the angle at the centre will be $\frac{360^0}{12} = 30^0$ B$_1$ Area of 1 triangle $= = \frac{1}{2} r^2 \sin\theta$ $= \frac{1}{2} x 10 x 10 x \sin 30^0$ M$_1$ = 25cm^2 A$_1$ Area of the polygon = 25 x 12 M$_1$ =300cm^2 A$_1$	8marks
13a.	2(x+1)	Each interior angle of a regular polygon $= \frac{(n-2)180}{n}$ Each exterior angle = $\frac{360}{n}$ MI for equating $\frac{(n-2)180}{n} = \left(\frac{360}{n}\right)x$ (n-2) 180^o = 360x 180^{o}n = 360 = 360x n-2 = 2x MI for solving A1 for n=2(x+1) (accept any other correct method)	8marks
14b.	17.5km to 3 sig. fig	 B$_2$ Let $\overline{CN}$ = x and $\overline{NP}$ = y. From ΔCNT, $\angle$ NCT = 180^0 – 150^0 = 30^0 $\angle$ CTN = 150^0 – 90^0 = 60^0 = θ Also, NTP = 90^0 – 40^0 = 50^0 = β	

		$\therefore \tan 60^0 = \frac{x}{6}$ x= 6 tan 60^0 M_1 x= 10.39km A_1 Also tan 50^0 = $\frac{y}{6}$ y= 6 tan 500 M_1 y= 7.15 km A_1 Between C and P = x + y M_1 = 10.39 + 7.15 = 17.54km = 17.5km to 3 sig. fig. A_1	
14i 14ii.		Peugeot 180 = $\frac{180}{480} x \frac{360}{1} = 135$ Datsun 108 = $\frac{108}{480} x \frac{360}{1} = 81$ Ford 72 = $\frac{72}{480} x \frac{360}{1} = 54$ Toyota 56 = $\frac{56}{480} x \frac{360}{1} = 42$ Volkswagen 64 = $\frac{64}{480} x \frac{360}{1} = 48$  ` B1 for Volkswagen = 64 M1 for finding an angle A2 for any correct four ($-\frac{1}{2} e.e$) B2 for correct pie chart $\left(\frac{1}{2} e.e\right)$ M1 for Aob of faulty engine to be Toyota or Volkswagen $\frac{56}{480} + \frac{64}{480} = 0.25 \text{ or } \frac{1}{4}$	8marks

15i 15ii	0.25 $\frac{40}{973}$ $\frac{493}{973}$	A1 for 0.25 or $\frac{1}{4}$ Alternative $\frac{42}{360}+\frac{48}{360}=0.25 \text{ or } \frac{1}{4}$ Total number of fruits = 60 + 80 = 140 Prob. (mango 1st) = $\frac{80}{140}=\frac{4}{7}$ Prob. (mango 2nd) = $\frac{79}{139}$ B_1 Prob. (orange 1st) = $\frac{60}{140}=\frac{3}{7}$ B_1 Prob. (mango 2nd) = $\frac{60}{139}$ B_1Prob. (one of each fruit) $=\frac{4}{7}\times\frac{60}{139}+\frac{3}{7}\times\frac{80}{139}$ M_1 $=\frac{240+240}{973}=\frac{40}{973}$ A_1 Prob. (only one type of fruit) $=\frac{4}{7}\times\frac{79}{139}+\frac{3}{7}\times\frac{59}{139}$ M_1 $=\frac{316+177}{973}$ $=\frac{493}{973}$ M_1	8marks

APENDIX VII

Students Ability Estimates

S/N	WAEC_CTT_SCORE	NECO_CTT_SCORE	WAEC_GPCM_ZSCORE	NECO_GPCM_ZSCORE	WAEC_GRM_ZSCORE	NECO_GRM_ZSCORE	GPCM_WAEC_TSCORE	GPCM_NECO_TSCORE	GRM_WAEC_TSCORE	GRM_NECO_TSCORE
1	33	65	-0.547	1.350	-0.175	1.594	45	64	48	66
2	23	35	-0.917	-0.220	-0.915	-0.197	41	48	41	48
3	38	29	0.030	-0.423	0.171	-0.502	50	46	52	45
4	43	21	0.349	-1.327	0.342	-1.179	53	37	53	38
5	33	10	-0.239	-1.921	-0.42	-1.978	48	31	46	30
6	30	31	-0.152	-0.235	-0.246	-0.213	48	48	48	48
7	42	21	0.844	-0.898	0.857	-0.59	58	41	59	44
8	30	13	-0.143	-1.086	-0.161	-0.936	49	39	48	41
9	44	18	1.032	-1.162	1.046	-0.569	60	38	60	44
10	11	17	-1.306	-1.164	-1.074	-0.862	37	38	39	41
11	22	21	-0.789	-0.694	-0.765	-0.801	42	43	42	42
12	20	16	-0.796	-1.382	-0.629	-1.213	42	36	44	38
13	24	14	-0.997	-1.621	-0.818	-1.33	40	34	42	37
14	38	10	-0.214	-1.681	-0.134	-1.604	48	33	49	34
15	28	10	-0.382	-1.828	-0.448	-1.857	46	32	46	31
16	29	11	0.015	-0.899	-0.024	-0.695	50	41	50	43
17	31	14	0.060	-1.848	0.064	-1.513	51	32	51	35
18	30	40	0.000	-0.155	-0.121	0.007	50	48	49	50
19	41	23	1.184	-0.887	1.245	-0.948	62	41	62	41
20	16	27	-1.159	-0.364	-1.2	-0.404	38	46	38	46
21	21	13	-0.739	-1.171	-0.842	-1.501	43	38	42	35
22	31	18	-0.192	-0.780	-0.135	-0.904	48	42	49	41
23	32	28	0.339	-0.109	0.361	-0.037	53	49	54	50
24	25	21	-0.632	-0.527	-0.526	-0.561	44	45	45	44
25	30	27	-0.238	-0.968	-0.563	-0.895	48	40	44	41
26	32	40	-0.037	-0.155	-0.235	0.007	50	48	48	50
27	38	27	0.515	-0.748	0.239	-0.76	55	43	52	42
28	52	28	1.830	-0.416	1.927	-0.349	68	46	69	47
29	36	18	0.968	-1.711	0.985	-1.81	60	33	60	32
30	42	25	1.143	-0.298	1.088	-0.261	61	47	61	47
31	33	26	-0.017	-0.444	-0.377	-0.49	50	46	46	45
32	31	16	-0.299	-0.871	-0.371	-0.818	47	41	46	42

33	25	23	-0.487	-0.692	-0.544	-0.353	45	43	45	46
34	25	32	-0.133	-0.322	-0.158	-0.258	49	47	48	47
35	23	33	-0.446	-0.255	-0.472	-0.12	46	47	45	49
36	33	20	-0.381	-0.854	-0.296	-0.982	46	41	47	40
37	20	22	-1.510	-0.616	-1.366	-0.616	35	44	36	44
38	13	33	-1.687	-0.433	-1.553	-0.457	33	46	34	45
39	16	31	-1.151	0.145	-1.059	-0.092	38	51	39	49
40	16	32	1.279	0.333	-1.314	0.176	63	53	37	52
41	27	37	-0.420	-0.262	-0.621	-0.159	46	47	44	48
42	25	32	-0.582	-0.010	-0.853	0.081	44	50	41	51
43	24	15	-0.897	-1.302	-0.933	-1.541	41	37	41	35
44	21	33	-0.601	-0.075	-0.805	0.003	44	49	42	50
45	18	32	-1.015	-0.186	-1.146	-0.169	40	48	39	48
46	30	23	0.157	-0.299	0	-0.335	52	47	50	47
47	18	32	-1.089	-0.211	-1.312	-0.115	39	48	37	49
48	25	22	-0.191	-0.938	-0.207	-1.12	48	41	48	39
49	15	42	-1.253	-0.060	-1.134	0.111	37	49	39	51
50	27	18	0.066	-0.894	0	-1.234	51	41	50	38
51	25	16	-0.538	-0.956	-0.777	-1.101	45	40	42	39
52	30	26	-0.464	-0.055	-0.637	-0.379	45	49	44	46
53	35	20	0.215	-0.608	-0.107	-0.895	52	44	49	41
54	40	13	1.135	-1.169	1.219	-1.345	61	38	62	37
55	42	24	0.431	-0.021	0.432	-0.094	54	50	54	49
56	39	23	1.127	-0.077	1.202	-0.102	61	49	62	49
57	40	34	1.169	0.540	1.243	0.212	62	55	62	52
58	10	28	-1.885	0.030	-1.882	-0.255	31	50	31	47
59	24	26	-0.614	-0.034	-0.708	-0.379	44	50	43	46
60	29	8	0.146	-1.065	0.059	-1.49	51	39	51	35
61	29	21	0.146	-0.358	0.059	-0.621	51	46	51	44
62	26	40	-0.442	-0.155	-0.501	0.007	46	48	45	50
63	28	23	-0.106	0.232	-0.369	0.046	49	52	46	50
64	25	22	-0.383	-0.118	-0.308	-0.333	46	49	47	47
65	26	27	-0.403	-0.147	-0.404	-0.487	46	49	46	45
66	28	28	-0.219	-0.308	-0.167	-0.349	48	47	48	47
67	23	28	-0.783	-0.040	-0.761	-0.445	42	50	42	46
68	24	51	-0.736	0.910	-0.688	0.893	43	59	43	59
69	26	42	-0.646	-0.207	-0.696	-0.009	44	48	43	50
70	20	40	-0.557	-0.155	-0.61	0.007	44	48	44	50
71	30	17	0.151	-0.828	-0.193	-1.04	52	42	48	40
72	43	22	1.375	-0.006	1.451	-0.434	64	50	65	46
73	31	8	0.480	-1.429	0.438	-1.41	55	36	54	36
74	38	33	0.751	0.599	0.744	0.277	58	56	57	53
75	53	18	0.433	-0.256	0.399	-0.457	54	47	54	45

76	42	37	0.036	0.769	0	0.725	50	58	50	57
77	47	36	0.122	0.879	0.024	0.711	51	59	50	57
78	20	39	-1.385	0.948	-1.478	0.809	36	59	35	58
79	24	44	-1.442	1.311	-1.579	1.12	36	63	34	61
80	40	23	-0.709	-0.340	-0.757	-0.775	43	47	42	42
81	48	26	-0.417	-0.377	-0.57	-0.364	46	46	44	46
82	35	27	-0.222	-0.620	-0.178	-0.831	48	44	48	42
83	20	40	-1.152	-0.155	-1.1	0.007	38	48	39	50
84	35	33	0.270	0.599	0.347	0.277	53	56	53	53
85	20	18	-0.848	-0.256	-0.681	-0.457	42	47	43	45
86	42	37	0.469	0.769	0.508	0.725	55	58	55	57
87	15	12	-1.292	-0.855	-1.413	-0.665	37	41	36	43
88	23	42	-0.468	-0.207	-0.704	-0.009	45	48	43	50
89	21	15	-0.739	-0.845	-0.901	-0.815	43	42	41	42
90	23	11	-0.627	-0.935	-0.856	-0.831	44	41	41	42
91	29	40	0.009	-0.155	0	0.007	50	48	50	50
92	24	11	-0.238	-0.823	-0.153	-0.785	48	42	48	42
93	24	13	-0.490	-0.629	-0.45	-0.739	45	44	46	43
94	15	19	-1.419	-1.211	-1.515	-1.075	36	38	35	39
95	14	42	-1.427	-0.207	-1.55	-0.009	36	48	35	50
96	25	42	-0.690	-0.207	-0.761	-0.009	43	48	42	50
97	30	42	-0.485	0.222	-0.619	0.368	45	52	44	54
98	22	8	-0.646	-1.304	-0.559	-1.164	44	37	44	38
99	45	13	1.355	1.311	1.158	1.12	64	63	62	61
100	27	28	-1.175	-0.340	1.391	-0.775	38	47	64	42
101	53	32	1.053	-0.377	0.934	-0.364	61	46	59	46
102	42	33	-0.016	-0.620	-0.113	-0.831	50	44	49	42
103	40	24	-0.032	-0.155	-0.209	0.007	50	48	48	50
104	39	42	-0.235	0.599	-0.244	0.277	48	56	48	53
105	35	35	-0.563	-0.256	-0.709	-0.457	44	47	43	45
106	33	37	-0.758	0.769	-0.889	0.725	42	58	41	57
107	58	38	1.274	-0.855	1.117	-0.665	63	41	61	43
108	18	38	-0.852	-0.207	-0.811	-0.009	41	48	42	50
109	45	38	0.176	-1.848	0.097	-1.513	52	32	51	35
110	43	43	0.084	0.072	-0.049	0.076	51	51	50	51
111	44	40	0.035	0.187	-0.042	0.238	50	52	50	52
112	38	34	-0.478	-0.071	-0.59	-0.178	45	49	44	48
113	42	18	-0.002	-0.361	-0.143	-0.518	50	46	49	45
114	16	42	-2.015	0.720	-2.109	0.752	30	57	29	58
115	48	35	-0.126	0.435	-0.279	0.381	49	54	47	54
116	48	37	0.203	0.656	0	0.735	52	57	50	57
117	38	32	0.466	0.318	0.363	0.319	55	53	54	53
118	18	25	-2.393	-1.100	-2.459	-0.724	26	39	25	43

119	42	16	-0.032	-1.979	-0.179	-1.989	50	30	48	30
120	24	28	-0.884	-1.089	-0.836	-0.881	41	39	42	41
121	44	21	0.057	-1.301	-0.037	-0.978	51	37	50	40
122	18	17	-1.538	-1.312	-1.555	-1.117	35	37	34	39
123	17	19	-1.862	-1.018	-1.909	-0.837	31	40	31	42
124	34	24	-0.305	-0.788	-0.512	-0.745	47	42	45	43
125	27	20	-1.019	-1.301	-1.294	-1.117	40	37	37	39
126	28	17	-1.255	-1.549	-1.48	-1.444	37	35	35	36
127	22	16	-1.484	-1.877	-1.61	-1.72	35	31	34	33
128	27	19	-0.140	-1.527	-0.138	-1.267	49	35	49	37
129	25	21	-0.299	-1.509	-0.348	-1.158	47	35	47	38
130	28	42	-0.006	-0.060	-0.07	0.111	50	49	49	51
131	33	23	-0.124	-0.486	-0.228	-0.438	49	45	48	46
132	27	27	-0.119	0.166	-0.171	-0.065	49	52	48	49
133	28	21	-0.821	-0.342	-1.037	-0.485	42	47	40	45
134	57	24	2.117	-0.404	2.216	-0.596	71	46	72	44
135	60	17	2.227	-0.871	2.341	-0.81	72	41	73	42
136	38	24	0.231	-0.694	0.226	-0.774	52	43	52	42
137	28	42	-0.602	-0.060	-0.722	0.111	44	49	43	51
138	43	12	0.484	-1.039	0.525	-1.07	55	40	55	39
139	45	40	0.685	-0.155	0.445	0.007	57	48	54	50
140	51	40	1.127	-0.155	1.044	0.007	61	48	60	50
141	53	35	2.152	0.851	2.102	0.688	72	59	71	57
142	53	38	1.970	0.498	1.901	0.291	70	55	69	53
143	35	21	0.298	-0.389	0.146	-0.617	53	46	51	44
144	35	16	-0.082	-0.871	-0.119	-0.926	49	41	49	41
145	33	19	-0.088	-0.465	-0.089	-0.574	49	45	49	44
146	28	33	0.800	0.335	0.826	0.061	58	53	58	51
147	15	57	-1.419	1.532	-1.515	1.48	36	65	35	65
148	14	25	-1.427	-0.172	-1.55	-0.465	36	48	35	45
149	25	29	-0.690	-0.700	-0.761	-0.504	43	43	42	45
150	30	33	-0.383	-0.063	-0.391	0.076	46	49	46	51
151	42	44	0.094	0.629	0.042	0.803	51	56	50	58
152	28	18	-0.263	-1.644	-0.143	-1.603	47	34	49	34
153	40	27	0.640	-0.506	0.635	-0.669	56	45	56	43
154	28	32	-0.286	-0.402	-0.189	-0.302	47	46	48	47
155	25	29	-0.665	-0.608	-0.588	-0.519	43	44	44	45
156	30	23	-0.223	-1.004	-0.338	-0.723	48	40	47	43
157	31	14	0.024	-1.984	0	-1.808	50	30	50	32
158	30	20	-0.567	-1.355	-0.486	-1.377	44	36	45	36
159	38	24	-0.071	-1.099	-0.181	-0.715	49	39	48	43
160	35	28	0.135	-0.701	0.384	-0.674	51	43	54	43
161	39	23	0.077	-1.262	-0.037	-0.88	51	37	50	41

162	34	28	-0.121	-0.701	-0.141	-0.674	49	43	49	43
163	34	23	-0.173	-1.282	-0.267	-0.982	48	37	47	40
164	38	28	0.604	-0.773	0.531	-0.491	56	42	55	45
165	33	23	-0.058	-1.227	-0.061	-0.81	49	38	49	42
166	33	26	-0.310	-1.150	-0.36	-0.998	47	39	46	40
167	37	26	-0.067	-1.330	-0.09	-0.876	49	37	49	41
168	48	26	0.630	-1.262	0.634	-0.883	56	37	56	41
169	42	22	0.491	-1.604	0.611	-1.338	55	34	56	37
170	39	25	0.478	-1.090	0.456	-0.775	55	39	55	42
171	38	14	0.492	-2.043	0.435	-2.059	55	30	54	29
172	38	43	0.327	-0.007	0.403	0.116	53	50	54	51
173	27	18	-0.934	-1.682	-0.864	-1.759	41	33	41	32
174	27	29	-0.703	-0.798	-0.78	-0.548	43	42	42	45
175	37	22	-0.072	-0.790	0.112	-0.826	49	42	51	42
176	38	13	0.268	-0.568	0.275	-0.46	53	44	53	45
177	38	30	-0.087	-0.582	-0.044	-0.268	49	44	50	47
178	38	25	0.492	-1.344	0.435	-0.864	55	37	54	41
179	40	18	0.288	-1.730	0.279	-1.345	53	33	53	37
180	28	18	-0.696	-1.350	-0.745	-1.018	43	37	43	40
181	38	19	0.516	-1.192	0.486	-0.846	55	38	55	42
182	37	31	0.017	-0.676	0	-0.627	50	43	50	44
183	34	27	0.018	-0.789	-0.059	-0.558	50	42	49	44
184	38	37	0.344	0.313	0.33	0.454	53	53	53	55
185	37	26	0.288	-1.019	0.287	-0.945	53	40	53	41
186	35	19	0.032	-1.486	0	-1.467	50	35	50	35
187	36	32	-0.044	0.146	0.033	0.162	50	51	50	52
188	17	38	-1.210	0.195	-1.221	0.34	38	52	38	53
189	45	40	0.839	0.150	0.852	0.088	58	52	59	51
190	33	38	-0.163	0.326	-0.246	0.376	48	53	48	54
191	38	29	0.222	-0.706	0.205	-0.544	52	43	52	45
192	33	26	-0.076	-0.505	0	-0.454	49	45	50	45
193	40	48	-0.032	0.964	-0.122	1.044	50	60	49	60
194	40	33	0.416	-0.412	0.39	-0.438	54	46	54	46
195	38	21	0.085	-1.552	0.08	-1.616	51	34	51	34
196	34	34	-0.213	0.020	-0.315	-0.085	48	50	47	49
197	28	29	-0.589	0.533	-0.654	0.446	44	55	43	54
198	29	29	-0.586	-0.696	-0.606	-0.597	44	43	44	44
199	33	22	-0.575	-1.184	-0.536	-0.796	44	38	45	42
200	28	36	-0.319	-0.295	-0.357	-0.091	47	47	46	49
201	34	49	-0.280	0.263	-0.131	0.413	47	53	49	54
202	27	33	-0.746	-0.473	-0.831	-0.408	43	45	42	46
203	35	24	-0.577	-1.234	-0.813	-1.049	44	38	42	40
204	27	36	-0.521	0.081	-0.473	0.11	45	51	45	51

205	43	33	0.412	-0.329	0.346	-0.253	54	47	53	47
206	38	48	-0.054	0.460	-0.02	0.639	49	55	50	56
207	43	28	0.198	-0.835	0.224	-0.672	52	42	52	43
208	33	18	-0.177	-1.353	-0.143	-1.002	48	36	49	40
209	38	38	0.120	-0.733	0.16	-0.81	51	43	52	42
210	33	29	-0.236	-0.520	-0.278	-0.488	48	45	47	45
211	40	49	-0.079	0.263	-0.176	0.413	49	53	48	54
212	40	50	0.183	0.747	0.066	0.983	52	57	51	60
213	32	24	-0.631	-0.007	-0.75	0.026	44	50	43	50
214	23	33	-0.702	-0.017	-0.5	0.044	43	50	45	50
215	32	32	-0.427	-0.345	-0.594	-0.323	46	47	44	47
216	38	24	0.569	-0.568	0.409	-0.625	56	44	54	44
217	26	18	-0.804	-0.856	-0.901	-0.957	42	41	41	40
218	17	25	-0.948	-0.280	-0.94	-0.307	41	47	41	47
219	31	42	-0.141	0.799	-0.202	0.794	49	58	48	58
220	34	18	-0.121	-0.961	-0.266	-1.141	49	40	47	39
221	34	30	-0.297	-0.703	-0.38	-0.638	47	43	46	44
222	28	22	-0.209	-1.081	-0.321	-1.192	48	39	47	38
223	29	48	-0.301	0.501	-0.45	0.351	47	55	46	54
224	25	35	-0.591	-0.154	-0.627	-0.09	44	48	44	49
225	33	39	-0.659	-0.003	-0.639	0.091	43	50	44	51
226	33	32	0.017	-0.529	0.024	-0.488	50	45	50	45
227	33	36	-0.185	-0.410	-0.168	-0.388	48	46	48	46
228	42	27	-0.030	-0.558	-0.059	-0.618	50	44	49	44
229	25	33	-1.091	-0.347	-1.014	-0.273	39	47	40	47
230	21	33	-1.506	-0.289	-1.379	-0.201	35	47	36	48
231	22	28	-1.477	-0.478	-1.623	-0.445	35	45	34	46
232	27	32	-0.140	-0.013	-0.138	-0.016	49	50	49	50
233	25	42	-0.299	-0.207	-0.348	-0.009	47	48	47	50
234	34	24	-0.012	-0.544	-0.13	-0.483	50	45	49	45
235	33	25	-0.124	0.214	-0.228	-0.017	49	52	48	50
236	33	21	-0.124	-0.342	-0.228	-0.485	49	47	48	45
237	21	24	-0.813	-0.404	-0.989	-0.596	42	46	40	44
238	57	17	2.123	-0.871	2.235	-0.81	71	41	72	42
239	30	24	-0.620	-0.694	-0.786	-0.774	44	43	42	42
240	65	40	2.267	-0.155	2.366	0.007	73	48	74	50
241	37	12	0.233	-1.039	0.253	-1.07	52	40	53	39
242	38	40	0.762	-0.155	0.658	0.007	58	48	57	50
243	47	42	0.902	-0.060	0.773	0.111	59	49	58	51
244	51	33	1.127	0.896	1.044	0.748	61	59	60	57
245	53	38	2.108	0.498	2.103	0.291	71	55	71	53
246	52	21	1.860	-0.389	1.798	-0.617	69	46	68	44
247	37	16	0.362	-0.871	0.262	-0.926	54	41	53	41

248	33	32	0.036	0.051	0.069	-0.204	50	51	51	48
249	35	35	0.115	0.185	0.128	-0.182	51	52	51	48
250	51	25	1.278	-0.402	1.256	-0.528	63	46	63	45
251	55	31	1.670	-0.062	1.635	-0.215	67	49	66	48
252	55	21	1.583	-0.531	1.493	-0.579	66	45	65	44
253	45	25	0.607	-1.066	0.488	-1.202	56	39	55	38
254	53	29	1.567	-0.558	1.55	-0.568	66	44	66	44
255	25	28	-1.040	-0.549	-1.179	-0.552	40	45	38	44
256	53	23	1.567	-0.790	1.55	-0.73	66	42	66	43
257	43	23	0.784	-0.804	0.607	-0.962	58	42	56	40
258	25	26	-1.091	-0.558	-1.014	-0.568	39	44	40	44
259	21	24	-1.506	-1.083	-1.379	-0.988	35	39	36	40
260	30	24	-1.401	-0.629	-1.506	-0.692	36	44	35	43
261	33	31	-0.803	-0.415	-0.811	-0.397	42	46	42	46
262	28	22	-1.547	-0.763	-1.383	-1.008	35	42	36	40
263	23	25	-1.988	-0.700	-1.824	-0.432	30	43	32	46
264	20	23	-2.102	-0.254	-1.982	-0.307	29	47	30	47
265	28	23	-1.450	-1.119	-1.264	-0.628	36	39	37	44
266	30	23	-1.856	-1.160	-1.585	-0.923	31	38	34	41
267	24	20	-1.641	-0.513	-1.603	-0.416	34	45	34	46
268	26	27	-1.372	-0.487	-1.297	-0.527	36	45	37	45
269	30	23	-1.119	-0.338	-1.172	-0.173	39	47	38	48
270	37	23	-0.470	-0.318	-0.577	-0.331	45	47	44	47
271	33	18	-0.327	-1.188	-0.328	-1.077	47	38	47	39
272	29	36	-0.389	0.157	-0.397	0.132	46	52	46	51
273	38	34	-0.184	0.023	-0.202	-0.067	48	50	48	49
274	32	28	-0.702	-0.542	-0.614	-0.779	43	45	44	42
275	28	17	-1.588	-1.214	-1.602	-1.293	34	38	34	37
276	25	23	-1.393	-0.075	-1.501	-0.181	36	49	35	48
277	28	22	-1.265	-0.402	-1.15	-0.461	37	46	39	45
278	20	23	-1.660	-0.341	-1.649	-0.488	33	47	34	45
279	23	33	-1.336	0.149	-1.451	0.225	37	51	35	52
280	25	23	-1.264	-0.922	-1.246	-0.899	37	41	38	41
281	29	22	-1.082	-0.439	-1.065	-0.246	39	46	39	48
282	23	28	-1.435	0.035	-1.466	0.19	36	50	35	52
283	22	28	-1.497	-0.424	-1.525	-0.451	35	46	35	45
284	22	33	-1.721	-0.112	-1.734	-0.036	33	49	33	50
285	23	32	-1.360	-0.180	-1.34	-0.061	36	48	37	49
286	21	23	-1.524	0.089	-1.496	0.083	35	51	35	51
287	16	40	-2.283	-0.155	-2.231	0.007	27	48	28	50
288	23	23	-1.284	-0.922	-1.312	-0.899	37	41	37	41
289	27	22	-0.946	-0.439	-0.844	-0.246	41	46	42	48
290	25	28	-1.035	0.035	-1.117	0.19	40	50	39	52

291	26	34	-0.976	-0.509	-0.921	-0.476	40	45	41	45
292	37	15	-0.150	-0.643	-0.22	-0.71	49	44	48	43
293	23	33	-1.212	-0.240	-1.177	-0.465	38	48	38	45
294	37	22	0.010	-0.439	0.017	-0.246	50	46	50	48
295	34	28	-0.284	-0.221	-0.272	-0.462	47	48	47	45
296	38	21	0.101	-0.727	0.089	-0.78	51	43	51	42
297	29	31	-0.644	-0.024	-0.78	-0.17	44	50	42	48
298	29	24	-0.620	-0.365	-0.611	-0.638	44	46	44	44
299	33	49	-0.253	0.263	-0.183	0.413	47	53	48	54
300	27	26	-0.938	-0.316	-1.021	-0.46	41	47	40	45
301	25	29	-1.112	0.239	-1.139	0.04	39	52	39	50
302	29	28	-0.780	0.151	-0.832	-0.012	42	52	42	50
303	23	19	-1.264	-0.167	-1.338	-0.34	37	48	37	47
304	19	33	-1.771	0.013	-1.677	0.168	32	50	33	52
305	23	30	-1.466	0.439	-1.458	0.446	35	54	35	54
306	19	26	-1.802	0.287	-1.763	0.053	32	53	32	51
307	23	33	-1.466	0.604	-1.458	0.458	35	56	35	55
308	25	38	-1.091	0.670	-1.014	0.647	39	57	40	56
309	21	19	-1.506	-0.616	-1.379	-0.616	35	44	36	44
310	30	28	-1.226	-0.089	-1.318	-0.161	38	49	37	48
311	21	27	-1.661	-0.166	-1.494	-0.225	33	48	35	48
312	32	20	-1.272	-0.249	-1.369	-0.588	37	48	36	44
313	25	22	-1.091	-0.337	-1.014	-0.509	39	47	40	45
314	21	28	-1.506	-0.254	-1.379	-0.554	35	47	36	44
315	38	40	0.223	-0.155	0.116	0.007	52	48	51	50
316	28	23	-0.554	-0.486	-0.651	-0.438	44	45	43	46
317	20	25	-0.625	0.214	-0.384	-0.017	44	52	46	50
318	32	21	-0.427	-0.342	-0.594	-0.485	46	47	44	45
319	37	24	0.606	-0.404	0.475	-0.596	56	46	55	44
320	26	15	-0.804	-0.808	-0.901	-0.759	42	42	41	42
321	17	24	-0.948	-0.694	-0.94	-0.774	41	43	41	42
322	31	40	-0.141	-0.155	-0.202	0.007	49	48	48	50
323	34	12	-0.121	-1.039	-0.266	-1.07	49	40	47	39
324	34	17	-0.297	-0.534	-0.38	-0.64	47	45	46	44
325	28	43	-0.209	-0.014	-0.321	0.139	48	50	47	51
326	29	33	-0.301	0.896	-0.45	0.748	47	59	46	57
327	25	38	-0.591	0.498	-0.627	0.291	44	55	44	53
328	33	21	-0.659	-0.389	-0.639	-0.617	43	46	44	44
329	33	16	0.017	-0.871	0.024	-0.926	50	41	50	41
330	33	19	-0.185	-0.465	-0.168	-0.574	48	45	48	44
331	42	33	-0.030	0.335	-0.059	0.061	50	53	49	51
332	35	57	-0.487	1.532	-0.461	1.48	45	65	45	65
333	34	22	-0.785	-0.170	-0.702	-0.496	42	48	43	45

334	25	23	-0.887	0.126	-0.887	-0.255	41	51	41	47
335	23	41	-1.063	0.973	-1.059	0.872	39	60	39	59
336	38	29	0.425	0.033	0.363	0.065	54	50	54	51
337	36	41	-0.435	0.637	-0.498	0.674	46	56	45	57
338	31	40	-0.723	-0.155	-0.817	0.007	43	48	42	50
339	28	45	-0.843	0.775	-0.902	0.834	42	58	41	58
340	32	21	-0.745	-0.120	-0.696	-0.486	43	49	43	45
341	32	45	-0.346	1.071	-0.293	1.132	47	61	47	61
342	28	19	-0.932	-0.112	-0.633	-0.356	41	49	44	46
343	36	28	-0.470	-0.172	-0.293	-0.462	45	48	47	45
344	38	19	0.425	-0.537	0.363	-0.587	54	45	54	44
345	36	13	-0.435	-1.115	-0.498	-1.228	46	39	45	38
346	31	58	-0.723	1.413	-0.817	1.373	43	64	42	64
347	28	16	-0.843	-0.481	-0.902	-0.647	42	45	41	44
348	32	39	-0.745	1.247	-0.696	1.079	43	62	43	61
349	35	16	-0.797	-1.289	-0.667	-1.457	42	37	43	35
350	32	10	-0.833	-1.168	-0.568	-1.215	42	38	44	38
351	42	14	-0.081	-0.493	0.065	-0.488	49	45	51	45
352	38	19	-0.070	-0.604	0	-0.696	49	44	50	43
353	40	28	-0.054	-0.157	-0.029	-0.41	49	48	50	46
354	33	14	-0.178	-1.205	-0.123	-1.095	48	38	49	39
355	33	41	0.049	0.780	0.126	0.526	50	58	51	55
356	40	8	-0.174	-1.445	0	-1.558	48	36	50	34
357	27	19	-1.071	-0.632	-1.096	-0.713	39	44	39	43
358	47	40	-0.256	-0.155	-0.186	0.007	47	48	48	50
359	48	37	0.615	0.599	0.508	0.352	56	56	55	54
360	47	38	-0.186	0.621	-0.132	0.56	48	56	49	56
361	47	31	0.363	0.293	0.43	0.306	54	53	54	53
362	28	29	-0.328	0.388	-0.32	0.135	47	54	47	51
363	33	40	-0.684	-0.155	-0.747	0.007	43	48	43	50
364	42	16	0.819	-0.666	0.727	-0.796	58	43	57	42
365	38	27	-0.157	-0.721	-0.058	-0.545	48	43	49	45
366	37	38	-0.461	0.262	-0.488	0.222	45	53	45	52
367	30	28	-0.648	-0.547	-0.65	-0.601	44	45	44	44
368	27	32	-0.726	0.046	-0.527	0.163	43	50	45	52
369	43	31	0.169	-0.400	0.176	-0.55	52	46	52	45
370	42	33	-0.408	-0.227	-0.303	-0.155	46	48	47	48
371	32	51	-0.968	0.859	-0.862	1.004	40	59	41	60
372	38	28	-0.040	0.167	-0.031	0.19	50	52	50	52
373	25	46	-1.348	1.300	-1.394	1.178	37	63	36	62
374	38	47	0.175	1.337	0.229	1.441	52	63	52	64
375	45	48	-0.099	0.969	-0.122	1.118	49	60	49	61
376	37	48	-0.286	0.811	-0.228	0.888	47	58	48	59

377	37	33	-0.100	0.019	-0.036	0.257	49	50	50	53
378	41	41	-0.124	0.365	-0.053	0.333	49	54	49	53
379	41	61	0.331	1.636	0.358	1.592	53	66	54	66
380	28	26	-0.707	-0.218	-0.675	0.002	43	48	43	50
381	27	31	-0.806	0.059	-0.705	0.136	42	51	43	51
382	34	44	-0.337	0.356	-0.368	0.476	47	54	46	55
383	33	45	-0.655	0.470	-0.766	0.523	43	55	42	55
384	53	27	0.663	-0.547	0.625	-0.703	57	45	56	43
385	38	43	0.425	0.217	0.363	0.232	54	52	54	52
386	36	37	-0.435	0.634	-0.498	0.689	46	56	45	57
387	31	24	-0.723	-0.823	-0.817	-0.705	43	42	42	43
388	28	28	-0.843	-0.261	-0.902	-0.213	42	47	41	48
389	32	20	-0.745	-0.496	-0.696	-0.474	43	45	43	45
390	23	29	-1.360	-0.433	-1.285	-0.301	36	46	37	47
391	28	22	-0.932	0.090	-0.633	0.031	41	51	44	50
392	36	25	-0.470	-0.008	-0.293	0.171	45	50	47	52
393	38	48	0.425	0.811	0.363	0.888	54	58	54	59
394	36	33	-0.435	0.019	-0.498	0.257	46	50	45	53
395	31	41	-0.723	0.365	-0.817	0.333	43	54	42	53
396	28	61	-0.843	1.636	-0.902	1.592	42	66	41	66
397	32	26	-0.745	-0.218	-0.696	0.002	43	48	43	50
398	23	48	-1.293	0.811	-1.115	0.888	37	58	39	59
399	28	33	-0.932	0.019	-0.633	0.257	41	50	44	53
400	42	37	-0.081	0.116	0.065	0.057	49	51	51	51
401	38	27	-0.070	-0.471	0	-0.433	49	45	50	46
402	40	22	-0.054	-0.477	-0.029	-0.461	49	45	50	45
403	33	24	-0.178	-0.468	-0.123	-0.399	48	45	49	46
404	33	32	0.049	-0.597	0.126	-0.367	50	44	51	46
405	40	23	-0.174	-0.628	0	-0.514	48	44	50	45
406	27	27	-1.071	-0.184	-1.096	-0.208	39	48	39	48
407	47	30	-0.256	-0.141	-0.186	-0.03	47	49	48	50
408	48	24	0.615	-0.779	0.508	-0.737	56	42	55	43
409	47	36	-0.186	-0.211	-0.132	-0.115	48	48	49	49
410	47	37	0.363	0.428	0.43	0.526	54	54	54	55
411	28	36	-0.328	0.526	-0.32	0.363	47	55	47	54
412	33	39	-0.684	0.435	-0.747	0.362	43	54	43	54
413	42	45	0.819	0.170	0.727	0.233	58	52	57	52
414	31	38	-0.034	0.297	-0.057	0.105	50	53	49	51
415	32	36	0.002	0.273	-0.021	0.015	50	53	50	50
416	37	31	0.332	-0.427	0.239	-0.469	53	46	52	45
417	45	38	0.369	0.468	0.304	0.493	54	55	53	55
418	28	38	-0.144	0.807	0	0.754	49	58	50	58
419	47	37	1.048	0.724	1.329	0.683	60	57	63	57

420	28	23	-0.743	-0.715	-0.697	-0.913	43	43	43	41
421	32	33	0.140	0.350	0.194	0.187	51	54	52	52
422	30	43	-1.139	0.045	-1.082	0.218	39	50	39	52
423	37	37	-0.061	0.091	0.101	0.021	49	51	51	50
424	49	44	0.605	0.635	0.815	0.605	56	56	58	56
425	54	39	0.420	0.458	0.431	0.31	54	55	54	53
426	28	46	-0.485	1.040	-0.329	0.828	45	60	47	58
427	30	28	-0.643	-0.477	-0.842	-0.427	44	45	42	46
428	45	40	0.283	0.525	0.414	0.472	53	55	54	55
429	33	34	-0.381	-0.448	-0.457	-0.416	46	46	45	46
430	33	42	-0.339	0.448	-0.492	0.449	47	54	45	54
431	58	45	0.862	0.806	1.124	0.762	59	58	61	58
432	37	38	-0.431	0.167	-0.473	0.339	46	52	45	53
433	36	45	-0.078	0.562	-0.165	0.627	49	56	48	56
434	56	46	0.281	0.617	0.2	0.53	53	56	52	55
435	34	33	-0.044	-0.193	-0.053	-0.091	50	48	49	49
436	48	45	0.127	0.924	0.091	0.963	51	59	51	60
437	48	35	0.103	-0.035	0.151	0.01	51	50	52	50
438	43	40	0.054	0.297	-0.02	0.389	51	53	50	54
439	50	38	0.588	0.283	0.533	0.367	56	53	55	54
440	62	33	1.386	0.039	1.161	-0.054	64	50	62	49
441	63	32	0.910	-0.197	0.944	-0.223	59	48	59	48
442	48	44	0.975	0.902	0.978	0.684	60	59	60	57
443	50	31	0.777	-0.652	0.605	-0.569	58	43	56	44
444	43	35	0.582	-0.084	0.487	-0.014	56	49	55	50
445	38	29	0.092	-0.635	0.067	-0.54	51	44	51	45
446	44	35	0.056	0.247	0.036	0.244	51	52	50	52
447	35	37	0.118	0.227	0.172	0.157	51	52	52	52
448	75	68	2.299	1.878	2.264	1.961	73	69	73	70
449	43	39	0.308	0.036	0.381	0.053	53	50	54	51
450	45	46	0.283	0.863	0.52	0.71	53	59	55	57
451	52	55	0.984	1.470	1.115	1.596	60	65	61	66
452	43	25	-0.348	-0.640	-0.275	-0.618	47	44	47	44
453	54	58	1.351	1.628	1.286	1.685	64	66	63	67
454	39	39	0.153	0.578	0.045	0.533	52	56	50	55
455	22	29	-0.943	-0.311	-0.561	-0.257	41	47	44	47
456	51	58	1.184	1.549	1.558	1.691	62	65	66	67
457	33	18	-0.613	-0.990	-0.566	-0.995	44	40	44	40
458	45	38	0.710	0.276	0.646	0.134	57	53	56	51
459	37	43	-0.167	-0.117	-0.209	-0.178	48	49	48	48
460	47	18	0.448	-0.990	0.441	-0.995	54	40	54	40
461	27	38	-0.319	0.276	-0.451	0.134	47	53	45	51
462	38	40	0.425	-0.155	0.363	0.007	54	48	54	50

463	36	23	-0.435	-0.486	-0.498	-0.438	46	45	45	46
464	31	25	-0.723	0.214	-0.817	-0.017	43	52	42	50
465	28	21	-0.843	-0.342	-0.902	-0.485	42	47	41	45
466	35	24	-0.280	-0.404	-0.322	-0.596	47	46	47	44
467	18	15	-1.382	-0.808	-1.289	-0.759	36	42	37	42
468	35	24	-0.322	-0.694	-0.226	-0.774	47	43	48	42
469	32	43	0.428	-0.234	0.434	-0.02	54	48	54	50
470	51	18	1.192	-1.282	1.344	-1.201	62	37	63	38
471	42	17	0.953	-0.534	0.899	-0.64	60	45	59	44
472	32	43	0.252	0.032	0.256	0.168	53	50	53	52
473	54	33	0.924	0.896	0.981	0.748	59	59	60	57
474	50	36	1.913	0.499	2.058	0.282	69	55	71	53
475	40	21	0.779	-0.389	0.796	-0.617	58	46	58	44
476	32	16	-0.049	-0.871	-0.016	-0.926	50	41	50	41
477	30	19	-0.228	-0.465	-0.284	-0.574	48	45	47	44
478	40	33	1.078	0.335	1.077	0.061	61	53	61	51
479	32	57	-0.228	1.532	-0.184	1.48	48	65	48	65
480	35	22	0.285	-0.170	0.189	-0.496	53	48	52	45
481	40	23	0.731	0.126	0.723	-0.255	57	51	57	47
482	32	41	0.472	0.973	0.431	0.872	55	60	54	59
483	40	29	0.604	0.033	0.795	0.065	56	50	58	51
484	45	41	1.532	0.637	1.511	0.674	65	56	65	57
485	40	47	0.399	0.308	0.456	0.336	54	53	55	53
486	40	45	0.937	0.775	1.189	0.834	59	58	62	58
487	48	21	1.656	-0.120	1.755	-0.486	67	49	68	45
488	32	45	0.340	1.071	0.413	1.132	53	61	54	61
489	32	19	0.609	-0.112	0.566	-0.356	56	49	56	46
490	48	28	1.422	-0.172	1.441	-0.462	64	48	64	45
491	37	19	0.786	-0.537	0.674	-0.587	58	45	57	44
492	30	13	0.250	-1.115	0.245	-1.228	53	39	52	38
493	45	55	1.145	1.415	1.189	1.363	61	64	62	64
494	38	16	0.790	-0.481	0.73	-0.647	58	45	57	44
495	45	39	1.266	1.247	1.389	1.079	63	62	64	61
496	42	13	1.359	-1.286	1.456	-1.468	64	37	65	35
497	32	10	0.180	-1.168	0.144	-1.215	52	38	51	38
498	33	14	-0.212	-0.493	-0.202	-0.488	48	45	48	45
499	52	19	1.545	-0.604	1.574	-0.696	65	44	66	43
500	40	65	0.820	1.471	1.052	1.636	58	65	61	66
501	33	35	0.526	-0.220	0.508	-0.197	55	48	55	48
502	57	29	1.951	-0.280	2	-0.279	70	47	70	47
503	33	21	0.333	-1.142	0.373	-0.896	53	39	54	41
504	43	10	0.807	-1.704	0.79	-1.578	58	33	58	34
505	56	31	1.761	-0.235	1.817	-0.213	68	48	68	48

506	33	21	0.875	-0.898	0.935	-0.59	59	41	59	44
507	33	13	0.074	-1.086	0.056	-0.936	51	39	51	41
508	50	18	1.578	-1.162	1.64	-0.569	66	38	66	44
509	36	17	0.443	0.540	0.425	0.212	54	55	54	52
510	45	21	1.084	-0.694	1.075	-0.801	61	43	61	42
511	38	17	0.621	-1.290	0.606	-1.06	56	37	56	39
512	58	11	1.879	-1.466	1.827	-1.252	69	35	68	37
513	40	10	0.381	-1.681	0.395	-1.604	54	33	54	34
514	50	10	1.745	-1.828	1.769	-1.857	67	32	68	31
515	46	11	0.744	-0.899	0.825	-0.695	57	41	58	43
516	30	11	0.210	-1.684	0.23	-1.426	52	33	52	36
517	38	46	1.054	0.925	0.963	0.915	61	59	60	59
518	37	33	1.031	-0.058	0.97	0.048	60	49	60	50
519	63	33	1.485	-0.166	1.575	-0.103	65	48	66	49
520	47	23	1.246	-0.913	1.3	-0.725	62	41	63	43
521	39	23	1.185	-0.658	1.19	-0.498	62	43	62	45
522	33	30	-0.099	-0.551	-0.195	-0.397	49	44	48	46
523	23	30	-0.525	-0.666	-0.545	-0.499	45	43	45	45
524	23	39	-0.848	-0.399	-0.876	-0.175	42	46	41	48
525	28	15	0.297	-1.272	0.239	-0.84	53	37	52	42
526	41	11	0.793	-1.594	0.987	-1.281	58	34	60	37
527	32	24	0.322	-0.847	0.363	-0.477	53	42	54	45
528	29	22	0.591	-0.866	0.707	-0.55	56	41	57	45
529	39	40	1.117	-0.131	1.136	0.072	61	49	61	51
530	43	21	1.051	-1.506	1.019	-1.073	61	35	60	39
531	63	21	1.933	-0.882	1.883	-0.845	69	41	69	42
532	33	23	-0.071	-0.655	-0.065	-0.563	49	43	49	44
533	32	20	0.252	-1.209	0.242	-1.389	53	38	52	36
534	68	37	1.033	0.055	1.098	0.133	60	51	61	51
535	40	28	0.312	-0.555	0.606	-0.497	53	44	56	45
536	37	25	0.795	-0.969	0.896	-0.88	58	40	59	41
537	53	33	2.390	-0.232	2.546	-0.071	74	48	75	49
538	45	30	0.808	-0.535	0.998	-0.525	58	45	60	45
539	38	23	0.373	-0.859	0.36	-0.798	54	41	54	42
540	41	18	0.990	-1.711	1.027	-1.81	60	33	60	32
541	36	32	0.717	-0.421	0.722	-0.29	57	46	57	47
542	51	16	1.009	-0.871	0.97	-0.818	60	41	60	42
543	48	20	0.874	-0.809	0.769	-0.54	59	42	58	45
544	30	30	0.120	-0.150	0.117	-0.028	51	49	51	50
545	29	33	0.174	-0.255	0.057	-0.12	52	47	51	49
546	48	22	1.340	-1.266	1.312	-1.309	63	37	63	37
547	42	27	1.369	-0.703	1.449	-0.618	64	43	64	44
548	41	29	1.366	-0.610	1.369	-0.544	64	44	64	45

549	48	37	1.532	-0.472	1.651	-0.539	65	45	67	45
550	40	39	0.416	0.210	0.39	0.034	54	52	54	50
551	38	42	0.085	0.434	0.08	0.437	51	54	51	54
552	34	43	-0.213	-0.384	-0.315	-0.191	48	46	47	48
553	28	35	-0.589	-0.056	-0.654	0.008	44	49	43	50
554	33	22	-0.669	-1.337	-0.693	-1.461	43	37	43	35
555	33	42	-0.575	0.002	-0.536	0.137	44	50	45	51
556	28	32	-0.319	-0.398	-0.357	-0.316	47	46	46	47
557	33	37	-0.240	0.287	-0.07	0.401	48	53	49	54
558	27	32	-0.746	-0.211	-0.831	-0.115	43	48	42	49
559	33	30	-0.536	-0.970	-0.749	-1.05	45	40	43	40
560	27	51	-0.521	1.453	-0.473	1.456	45	65	45	65
561	44	52	0.372	1.663	0.318	1.579	54	67	53	66
562	37	33	-0.014	0.654	0.027	0.451	50	57	50	55
563	39	37	0.271	0.565	0.323	0.52	53	56	53	55
564	32	47	-0.137	1.213	-0.092	1.197	49	62	49	62
565	38	43	0.120	1.112	0.16	1.102	51	61	52	61
566	34	48	-0.279	1.458	-0.304	1.375	47	65	47	64
567	37	28	0.001	0.211	-0.092	0.316	50	52	49	53
568	38	33	0.223	0.868	0.116	0.875	52	59	51	59
569	28	59	-0.554	2.209	-0.651	2.27	44	72	43	73
570	20	38	-0.625	-0.363	-0.384	-0.419	44	46	46	46
571	32	38	-0.427	-0.064	-0.594	0.103	46	49	44	51
572	37	59	0.606	2.209	0.475	2.27	56	72	55	73
573	26	29	-0.804	0.161	-0.901	-0.038	42	52	41	50
574	17	16	-0.948	-0.782	-0.94	-0.989	41	42	41	40
575	31	45	-0.141	1.116	-0.202	0.842	49	61	48	58
576	34	48	-0.121	1.313	-0.266	1.315	49	63	47	63
577	34	43	-0.297	1.163	-0.38	1.303	47	62	46	63
578	28	58	-0.209	1.530	-0.321	1.654	48	65	47	67
579	29	75	-0.301	2.209	-0.45	2.555	47	72	46	76
580	25	43	-0.591	0.526	-0.627	0.49	44	55	44	55
581	33	50	-0.659	1.512	-0.639	1.454	43	65	44	65
582	33	37	0.017	0.294	0.024	0.314	50	53	50	53
583	33	45	-0.185	0.575	-0.168	0.789	48	56	48	58
584	42	43	-0.030	1.132	-0.059	1.119	50	61	49	61
585	25	35	-1.091	0.319	-1.014	0.248	39	53	40	52
586	21	52	-1.506	1.366	-1.379	1.302	35	64	36	63
587	22	67	-1.477	2.419	-1.623	2.367	35	74	34	74
588	27	69	-0.140	2.601	-0.138	2.669	49	76	49	77
589	25	37	-0.299	0.205	-0.348	0.305	47	52	47	53
590	34	33	-0.012	0.533	-0.13	0.424	50	55	49	54
591	33	23	-0.124	0.033	-0.228	0.156	49	50	48	52

592	33	20	-0.124	-0.430	-0.228	-0.481	49	46	48	45
593	21	27	-0.813	0.013	-0.989	-0.096	42	50	40	49
594	56	25	1.979	0.099	1.896	0.132	70	51	69	51
595	30	55	-0.620	1.666	-0.786	1.844	44	67	42	68
596	63	38	1.974	0.991	1.895	1.016	70	60	69	60
597	37	33	0.233	0.230	0.253	0.372	52	52	53	54
598	38	52	0.648	1.496	0.545	1.367	56	65	55	64
599	47	25	0.902	-0.138	0.773	0.037	59	49	58	50
600	51	51	1.127	1.453	1.044	1.456	61	65	60	65
601	53	52	2.108	1.663	2.103	1.579	71	67	71	66
602	51	33	1.724	0.654	1.569	0.451	67	57	66	55
603	37	37	0.362	0.565	0.262	0.52	54	56	53	55
604	35	47	0.260	1.213	0.27	1.197	53	62	53	62
605	34	43	0.003	1.112	0.021	1.102	50	61	50	61
606	51	48	1.278	1.458	1.256	1.375	63	65	63	64
607	55	28	1.670	0.211	1.635	0.316	67	52	66	53
608	54	33	1.455	0.868	1.381	0.875	65	59	64	59
609	45	59	0.607	-0.040	0.488	-0.445	56	50	55	46
610	53	38	1.567	-0.363	1.55	-0.419	66	46	66	46
611	25	38	-1.040	-0.064	-1.179	0.103	40	49	38	51
612	53	59	1.567	2.209	1.55	2.27	66	72	66	73
613	43	29	0.784	0.161	0.607	-0.038	58	52	56	50
614	25	16	-1.091	-0.782	-1.014	-0.989	39	42	40	40
615	21	33	-1.506	0.654	-1.379	0.451	35	57	36	55
616	37	37	0.362	0.565	0.262	0.52	54	56	53	55
617	35	47	0.260	1.213	0.27	1.197	53	62	53	62
618	34	43	0.003	1.112	0.021	1.102	50	61	50	61
619	36	48	-0.435	1.458	-0.498	1.375	46	65	45	64
620	31	28	-0.723	0.211	-0.817	0.316	43	52	42	53
621	28	33	-0.843	0.868	-0.902	0.875	42	59	41	59
622	32	59	-0.745	2.209	-0.696	2.27	43	72	43	73
623	33	38	-0.301	-0.363	-0.344	-0.419	47	46	47	46
624	28	38	-0.932	-0.064	-0.633	0.103	41	49	44	51
625	42	59	-0.081	2.209	0.065	2.27	49	72	51	73
626	38	47	-0.070	1.213	0	1.197	49	62	50	62
627	40	43	-0.054	1.112	-0.029	1.102	49	61	50	61
628	33	59	-0.178	2.209	-0.123	2.27	48	72	49	73
629	33	32	0.049	-0.074	0.126	0.026	50	49	51	50
630	40	50	-0.174	1.202	0	1.183	48	62	50	62
631	27	65	-1.071	1.555	-1.096	1.779	39	66	39	68
632	47	58	-0.256	1.741	-0.186	1.894	47	67	48	69
633	48	38	0.615	0.732	0.508	0.857	56	57	55	59

634	47	43	-0.186	1.129	-0.132	1.204	48	61	49	62
635	47	57	0.363	1.821	0.43	1.765	54	68	54	68
636	28	55	-0.328	1.665	-0.32	1.576	47	67	47	66
637	33	46	-0.684	1.523	-0.747	1.615	43	65	43	66
638	42	38	0.819	0.729	0.727	0.731	58	57	57	57
639	31	54	-0.034	1.294	-0.057	1.44	50	63	49	64
640	32	48	0.002	1.313	-0.021	1.315	50	63	50	63
641	37	43	0.332	1.163	0.239	1.303	53	62	52	63
642	38	58	0.425	1.530	0.363	1.654	54	65	54	67
643	36	78	-0.435	2.203	-0.498	2.573	46	72	45	76
644	31	43	-0.723	0.526	-0.817	0.49	43	55	42	55
645	28	50	-0.843	1.512	-0.902	1.454	42	65	41	65
646	32	37	-0.745	0.294	-0.696	0.314	43	53	43	53
647	32	45	-0.346	0.575	-0.293	0.789	47	56	47	58
648	28	43	-0.932	1.132	-0.633	1.119	41	61	44	61
649	36	35	-0.470	0.319	-0.293	0.248	45	53	47	52
650	38	52	0.425	1.366	0.363	1.302	54	64	54	63
651	36	67	-0.435	2.419	-0.498	2.367	46	74	45	74
652	31	69	-0.723	2.601	-0.817	2.669	43	76	42	77
653	28	37	-0.843	0.205	-0.902	0.305	42	52	41	53
654	32	33	-0.745	0.533	-0.696	0.424	43	55	43	54
655	32	23	-0.346	0.033	-0.293	0.156	47	50	47	52
656	28	20	-0.932	-0.430	-0.633	-0.481	41	46	44	45
657	42	27	-0.081	0.013	0.065	-0.096	49	50	51	49
658	38	25	-0.070	0.099	0	0.132	49	51	50	51
659	40	55	-0.054	1.666	-0.029	1.844	49	67	50	68
660	46	38	0.791	0.991	1.121	1.016	58	60	61	60
661	71	33	2.184	0.230	2.15	0.372	72	52	72	54
662	73	52	1.482	1.496	1.395	1.367	65	65	64	64
663	68	25	1.579	-0.138	1.505	0.037	66	49	65	50
664	30	51	-0.036	1.453	0	1.456	50	65	50	65
665	53	52	1.646	1.663	1.428	1.579	66	67	64	66
666	32	33	0.104	0.654	0.294	0.451	51	57	53	55
667	45	37	1.348	0.565	1.337	0.52	63	56	63	55
668	48	47	0.888	1.213	0.894	1.197	59	62	59	62
669	22	43	-1.154	1.112	-1.136	1.102	38	61	39	61
670	43	48	1.337	1.458	1.51	1.375	63	65	65	64
671	40	28	0.982	0.211	1.021	0.316	60	52	60	53
672	40	33	0.879	0.868	1.058	0.875	59	59	61	59
673	53	59	2.257	2.209	2.423	2.27	73	72	74	73
674	65	38	2.606	-0.363	2.637	-0.419	76	46	76	46
675	75	38	1.800	-0.064	1.995	0.103	68	49	70	51
676	43	59	0.411	2.209	0.322	2.27	54	72	53	73

677	35	29	0.537	0.161	0.798	-0.038	55	52	58	50
678	37	16	0.612	-0.782	0.639	-0.989	56	42	56	40
679	43	45	1.022	1.116	1.038	0.842	60	61	60	58
680	44	16	1.071	-0.611	1.07	-0.958	61	44	61	40
681	47	37	1.382	0.793	1.402	0.448	64	58	64	54
682	43	18	0.905	-0.670	0.863	-0.849	59	43	59	42
683	71	28	2.172	-0.012	2.372	-0.389	72	50	74	46
684	51	15	1.504	-0.636	1.574	-1.127	65	44	66	39
685	33	14	0.567	-0.757	0.602	-1.139	56	42	56	39
686	31	16	0.055	-0.830	0.176	-1.176	51	42	52	38
687	42	23	1.356	-0.057	1.447	-0.544	64	49	64	45
688	36	24	0.679	-0.200	0.64	-0.506	57	48	56	45
689	39	31	0.438	0.490	0.505	0.144	54	55	55	51
690	26	18	-0.337	-0.670	-0.123	-0.849	47	43	49	42
691	38	34	0.526	0.717	0.594	0.355	55	57	56	54
692	34	23	0.312	-0.321	0.299	-0.747	53	47	53	43
693	37	32	0.995	0.028	0.985	-0.112	60	50	60	49
694	28	23	0.228	0.012	0.223	-0.622	52	50	52	44
695	58	24	1.992	-0.286	2.003	-0.733	70	47	70	43
696	48	23	1.317	0.027	1.284	-0.354	63	50	63	46
697	37	17	0.577	-0.487	0.646	-1.03	56	45	56	40
698	53	22	1.736	-0.405	1.787	-0.844	67	46	68	42
699	43	30	1.025	0.171	1.052	-0.343	60	52	61	47
700	30	23	0.618	-0.134	0.635	-0.368	56	49	56	46
701	52	23	1.334	-0.156	1.325	-0.48	63	48	63	45
702	53	34	1.404	0.483	1.446	0.245	64	55	64	52
703	32	23	0.461	-0.263	0.519	-0.553	55	47	55	44
704	28	17	0.136	-0.547	0.165	-1.09	51	45	52	39
705	35	39	0.632	0.583	0.663	0.32	56	56	57	53
706	25	20	-0.203	-0.286	-0.23	-0.582	48	47	48	44
707	41	23	1.355	-0.340	1.425	-0.915	64	47	64	41
708	28	33	0.151	0.656	0.19	0.179	52	57	52	52
709	35	13	0.316	0.879	0.385	0.711	53	59	54	57
710	41	27	0.274	0.192	0.175	-0.427	53	52	52	46
711	38	25	0.302	-0.163	0.307	-0.477	53	48	53	45
712	30	28	-0.009	-0.014	-0.041	-0.412	50	50	50	46
713	43	13	1.198	-0.830	0.201	-1.176	62	42	52	38
714	23	23	0.302	-0.340	0.275	-0.915	53	47	53	41
715	43	43	0.779	1.038	0.812	0.754	58	60	58	58
716	36	29	-0.061	0.419	-0.022	0.029	49	54	50	50
717	33	21	-0.103	-0.415	0.041	-0.728	49	46	50	43
718	45	18	0.645	-0.517	0.675	-0.813	56	45	57	42
719	38	13	0.565	-0.767	0.396	-1.362	56	42	54	36

720	43	30	0.647	0.444	0.649	0.08	56	54	56	51
721	28	31	-0.494	0.474	-0.259	0.097	45	55	47	51
722	24	20	-0.299	-0.340	-0.136	-0.915	47	47	49	41
723	38	27	0.425	-0.111	0.363	-0.307	54	49	54	47
724	36	29	-0.435	0.115	-0.498	-0.193	46	51	45	48
725	31	21	-0.723	0.012	-0.817	-0.622	43	50	42	44
726	28	32	-0.843	0.580	-0.902	0.133	42	56	41	51
727	32	13	-0.745	-1.070	-0.696	-1.442	43	39	43	36
728	38	28	0.203	-0.237	0.319	-0.472	52	48	53	45
729	28	44	-0.932	0.927	-0.633	0.608	41	59	44	56
730	36	13	-0.470	-1.281	-0.293	-1.592	45	37	47	34
731	38	20	0.425	-0.959	0.363	-1.268	54	40	54	37
732	36	24	-0.435	-0.378	-0.498	-0.767	46	46	45	42
733	31	15	-0.723	-1.191	-0.817	-1.464	43	38	42	35
734	28	31	-0.843	0.208	-0.902	-0.182	42	52	41	48
735	32	28	-0.745	0.161	-0.696	-0.324	43	52	43	47
736	33	28	-0.212	-0.237	-0.254	-0.472	48	48	47	45
737	28	78	-0.932	2.370	-0.633	2.509	41	74	44	75
738	42	73	-0.081	2.353	0.065	2.631	49	74	51	76
739	38	78	-0.070	3.122	0	3.384	49	81	50	84
740	40	76	-0.054	2.529	-0.029	2.55	49	75	50	76
741	33	23	-0.178	-0.071	-0.123	-0.292	48	49	49	47
742	33	24	0.049	-0.824	0.126	-0.752	50	42	51	42
743	40	24	-0.174	-0.378	0	-0.767	48	46	50	42
744	27	33	-1.071	0.109	-1.096	-0.067	39	51	39	49
745	47	23	-0.256	-0.071	-0.186	-0.292	47	49	48	47
746	48	24	0.615	-0.824	0.508	-0.752	56	42	55	42
747	47	63	-0.186	1.743	-0.132	1.864	48	67	49	69
748	47	61	0.363	1.215	0.43	1.528	54	62	54	65
749	28	50	-0.328	1.152	-0.32	1.152	47	62	47	62
750	33	61	-0.684	1.499	-0.747	1.551	43	65	43	66
751	42	22	0.819	-0.181	0.727	-0.2	58	48	57	48
752	31	44	-0.034	0.932	-0.057	0.898	50	59	49	59
753	32	53	0.002	1.337	-0.021	1.487	50	63	50	65
754	37	53	0.332	1.710	0.239	1.731	53	67	52	67
755	47	28	0.363	-0.234	0.43	-0.155	54	48	54	48
756	28	31	-0.328	0.717	-0.32	0.812	47	57	47	58
757	33	50	-0.684	1.513	-0.747	1.346	43	65	43	63
758	42	48	0.819	1.387	0.727	1.206	58	64	57	62
759	31	46	-0.034	1.475	-0.057	1.565	50	65	49	66
760	32	41	0.002	0.787	-0.021	0.614	50	58	50	56
761	37	42	0.332	1.018	0.239	1.099	53	60	52	61
762	38	38	0.425	0.765	0.363	0.739	54	58	54	57

763	36	35	-0.435	0.893	-0.498	1.003	46	59	45	60
764	31	51	-0.723	1.409	-0.817	1.457	43	64	42	65
765	28	26	-0.843	0.114	-0.902	0.289	42	51	41	53
766	32	78	-0.745	2.003	-0.696	2.198	43	70	43	72
767	35	26	-0.830	-0.397	-0.76	-0.62	42	46	42	44
768	28	38	-0.932	0.640	-0.633	0.247	41	56	44	52
769	36	38	-0.470	0.452	-0.293	0.579	45	55	47	56
770	38	30	0.425	0.338	0.363	0.418	54	53	54	54
771	36	41	-0.435	1.177	-0.498	1.179	46	62	45	62
772	31	34	-0.723	0.325	-0.817	0.348	43	53	42	53
773	28	53	-0.843	1.604	-0.902	1.576	42	66	41	66
774	32	63	-0.745	2.302	-0.696	2.124	43	73	43	71
775	32	66	-0.346	2.610	-0.293	2.461	47	76	47	75
776	28	30	-0.932	-0.121	-0.633	-0.067	41	49	44	49
777	42	28	-0.081	0.285	0.065	0.181	49	53	51	52
778	38	23	-0.070	-0.013	0	-0.031	49	50	50	50
779	40	25	-0.054	0.122	-0.029	0.004	49	51	50	50
780	46	26	0.791	0.148	1.121	0.062	58	51	61	51
781	71	21	2.184	-0.251	2.15	-0.335	72	47	72	47
782	73	54	1.482	1.615	1.395	1.632	65	66	64	66
783	68	35	1.579	0.455	1.505	0.464	66	55	65	55
784	30	34	-0.036	0.277	0	0.425	50	53	50	54
785	53	38	1.646	0.549	1.428	0.375	66	55	64	54
786	32	44	0.104	1.145	0.294	1.003	51	61	53	60
787	45	32	1.348	0.511	1.337	0.237	63	55	63	52
788	48	32	0.888	0.432	0.894	0.283	59	54	59	53
789	22	43	-1.154	1.229	-1.136	1.17	38	62	39	62
790	43	50	1.337	1.539	1.51	1.44	63	65	65	64
791	40	23	0.982	0.023	1.021	-0.035	60	50	60	50
792	40	34	0.879	0.909	1.058	0.932	59	59	61	59
793	53	58	2.257	2.078	2.423	2.137	73	71	74	71
794	65	41	2.606	0.338	2.637	0.227	76	53	76	52
795	47	31	0.363	-0.118	0.43	-0.044	54	49	54	50
796	28	30	-0.328	0.329	-0.32	0.442	47	53	47	54
797	33	46	-0.684	1.342	-0.747	1.323	43	63	43	63
798	42	38	0.819	0.549	0.727	0.375	58	55	57	54
799	31	44	-0.034	1.145	-0.057	1.003	50	61	49	60
800	32	42	0.002	-0.060	-0.021	0.111	50	49	50	51
801	37	23	0.332	-0.887	0.239	-0.948	53	41	52	41
802	38	23	0.425	-0.255	0.363	-0.327	54	47	54	47
803	36	13	-0.435	-1.171	-0.498	-1.501	46	38	45	35
804	31	15	-0.723	-0.658	-0.817	-0.817	43	43	42	42
805	28	28	-0.843	-0.109	-0.902	-0.037	42	49	41	50

806	32	21	-0.745	-0.527	-0.696	-0.561	43	45	43	44
807	32	23	-0.346	-0.840	-0.293	-0.82	47	42	47	42
808	28	37	-0.932	-0.052	-0.633	0.078	41	49	44	51
809	36	23	-0.470	-0.855	-0.293	-0.665	45	41	47	43
810	38	28	0.425	-0.416	0.363	-0.349	54	46	54	47
811	36	18	-0.435	-1.711	-0.498	-1.81	46	33	45	32
812	31	25	-0.723	-0.298	-0.817	-0.261	43	47	42	47
813	28	26	-0.843	-0.444	-0.902	-0.49	42	46	41	45
814	32	16	-0.745	-0.871	-0.696	-0.818	43	41	43	42
815	33	23	-0.161	-0.692	-0.121	-0.353	48	43	49	46
816	28	32	-0.932	-0.322	-0.633	-0.258	41	47	44	47
817	42	33	-0.081	-0.255	0.065	-0.12	49	47	51	49
818	38	20	-0.070	-0.854	0	-0.982	49	41	50	40
819	40	22	-0.054	-0.616	-0.029	-0.616	49	44	50	44
820	46	33	0.791	-0.433	1.121	-0.457	58	46	61	45
821	71	31	2.184	0.145	2.15	-0.092	72	51	72	49
822	73	32	1.482	0.333	1.395	0.176	65	53	64	52
823	68	37	1.579	-0.262	1.505	-0.159	66	47	65	48
824	30	32	-0.036	-0.010	0	0.081	50	50	50	51
825	53	15	1.646	-1.302	1.428	-1.541	66	37	64	35
826	32	33	0.104	-0.075	0.294	0.003	51	49	53	50
827	45	32	1.348	-0.186	1.337	-0.169	63	48	63	48
828	48	23	0.888	-0.299	0.894	-0.335	59	47	59	47
829	22	32	-1.154	-0.211	-1.136	-0.115	38	48	39	49
830	43	22	1.337	-0.938	1.51	-1.12	63	41	65	39
831	40	20	0.982	-1.175	1.021	-0.992	60	38	60	40
832	40	21	0.879	-0.742	1.058	-0.845	59	43	61	42
833	53	23	2.257	-0.861	2.423	-0.78	73	41	74	42
834	65	16	2.606	-1.160	2.637	-0.91	76	38	76	41
835	74	23	1.250	-0.943	1.467	-1.028	63	41	65	40
836	43	23	0.411	-0.737	0.322	-0.35	54	43	53	47
837	35	27	0.537	-0.040	0.798	-0.06	55	50	58	49
838	37	23	0.612	-0.643	0.639	-0.454	56	44	56	45
839	43	22	1.022	-0.136	1.038	-0.301	60	49	60	47
840	44	40	1.071	0.419	1.07	0.553	61	54	61	56
841	47	38	1.382	-0.025	1.402	-0.075	64	50	64	49
842	43	38	0.905	0.486	0.863	0.478	59	55	59	55
843	71	35	2.172	0.292	2.372	0.254	72	53	74	53
844	51	35	1.504	0.238	1.574	0.378	65	52	66	54
845	33	32	0.567	0.252	0.602	0.347	56	53	56	53
846	31	28	0.055	-0.616	0.176	-0.628	51	44	52	44
847	42	31	1.356	-0.553	1.447	-0.486	64	44	64	45
848	36	40	0.679	-0.166	0.64	-0.187	57	48	56	48

849	39	53	0.438	1.433	0.505	1.413	54	64	55	64
850	26	29	-0.337	-0.388	-0.123	-0.23	47	46	49	48
851	38	34	0.526	0.547	0.594	0.484	55	55	56	55
852	34	17	0.312	-0.907	0.299	-1.046	53	41	53	40
853	37	33	0.995	-0.314	0.985	-0.088	60	47	60	49
854	28	21	0.228	-0.929	0.223	-0.808	52	41	52	42
855	58	40	1.992	0.323	2.003	0.358	70	53	70	54
856	48	39	1.317	0.421	1.284	0.538	63	54	63	55
857	37	37	0.577	-0.488	0.646	-0.333	56	45	56	47
858	53	41	1.736	-0.230	1.787	-0.044	67	48	68	50
859	43	45	1.025	0.275	1.052	0.46	60	53	61	55
860	30	52	0.618	0.457	0.635	0.406	56	55	56	54
861	52	31	1.334	-0.067	1.325	-0.083	63	49	63	49
862	53	37	1.404	0.703	1.446	0.628	64	57	64	56
863	32	39	0.461	0.686	0.519	0.46	55	57	55	55
864	28	44	0.136	0.763	0.165	0.778	51	58	52	58
865	35	33	0.632	-0.182	0.663	-0.084	56	48	57	49
866	25	34	-0.203	0.196	-0.23	0.415	48	52	48	54
867	41	34	1.355	0.276	1.425	0.426	64	53	64	54
868	28	42	0.151	0.266	0.19	0.243	52	53	52	52
869	35	26	0.316	-0.479	0.385	-0.55	53	45	54	45
870	41	33	0.274	0.074	0.175	0.026	53	51	52	50
871	38	44	0.302	0.550	0.307	0.756	53	56	53	58
872	30	32	-0.009	-0.002	-0.041	0.101	50	50	50	51
873	43	43	1.198	0.402	1.201	0.573	62	54	62	56
874	23	28	0.302	-0.378	0.275	-0.296	53	46	53	47
875	43	35	0.779	0.010	0.812	0.224	58	50	58	52
876	36	28	-0.061	-0.287	-0.022	-0.244	49	47	50	48
877	33	25	-0.103	-0.476	0.041	-0.333	49	45	50	47
878	45	26	0.645	-0.535	0.675	-0.416	56	45	57	46
879	38	35	0.565	0.153	0.396	0.147	56	52	54	51
880	43	28	0.647	-0.633	0.649	-0.486	56	44	56	45
881	28	31	-0.494	-0.470	-0.259	-0.244	45	45	47	48
882	24	27	-0.299	-0.211	-0.136	-0.147	47	48	49	49
883	38	39	0.425	0.307	0.363	0.262	54	53	54	53
884	36	25	-0.435	-0.570	-0.498	-0.545	46	44	45	45
885	31	35	-0.723	0.042	-0.817	0.034	43	50	42	50
886	28	20	-0.843	-0.834	-0.902	-0.685	42	42	41	43
887	32	32	-0.745	-0.241	-0.696	-0.175	43	48	43	48
888	32	33	-0.346	0.219	-0.293	0.217	47	52	47	52
889	28	33	-0.932	-0.319	-0.633	-0.194	41	47	44	48
890	30	28	-0.600	-0.282	-0.777	-0.37	44	47	42	46
891	38	28	-0.544	-0.533	-0.421	-0.437	45	45	46	46

892	35	28	0.054	-0.911	0.098	-0.706	51	41	51	43
893	28	37	-0.546	-0.179	-0.373	-0.101	45	48	46	49
894	52	67	1.340	1.436	1.366	1.708	63	64	64	67
895	49	66	1.431	1.697	1.477	1.947	64	67	65	69
896	38	42	-0.008	0.503	0.089	0.415	50	55	51	54
897	43	66	0.969	1.760	0.98	2.01	60	68	60	70
898	28	43	-0.546	0.426	-0.373	0.514	45	54	46	55
899	30	35	-0.499	0.290	-0.28	0.385	45	53	47	54
900	28	37	-0.825	-0.096	-0.522	-0.204	42	49	45	48
901	38	43	0.293	0.718	0.204	0.751	53	57	52	58
902	23	38	-0.826	0.451	-0.516	0.546	42	55	45	55
903	30	49	0.082	1.268	0.056	1.272	51	63	51	63
904	28	35	-0.546	-0.310	-0.373	-0.06	45	47	46	49
905	42	40	0.385	0.277	0.248	0.382	54	53	52	54
906	28	38	-0.546	-0.804	-0.373	-0.584	45	42	46	44
907	27	43	-1.070	0.533	-0.783	0.608	39	55	42	56
908	42	53	0.695	1.295	0.676	1.496	57	63	57	65
909	33	57	0.168	0.222	0.13	0.368	52	52	51	54
910	28	37	-0.546	0.291	-0.373	0.178	45	53	46	52
911	38	37	-0.377	0.299	-0.184	0.36	46	53	48	54
912	40	50	0.491	0.824	0.471	1.036	55	58	55	60
913	35	40	0.194	0.241	0.194	0.208	52	52	52	52
914	32	38	-0.312	-0.082	-0.215	0.147	47	49	48	51
915	38	35	-0.681	-0.178	-0.778	0.101	43	48	42	51
916	32	41	-0.452	0.404	-0.229	0.445	45	54	48	54
917	30	33	-0.499	-0.466	-0.28	-0.291	45	45	47	47
918	53	48	0.547	0.738	0.745	0.874	55	57	57	59
919	35	33	-0.342	0.099	-0.115	0.126	47	51	49	51
920	30	42	-0.499	0.228	-0.28	0.247	45	52	47	52
921	38	31	-0.359	-0.138	-0.308	-0.033	46	49	47	50
922	43	42	0.414	0.272	0.292	0.334	54	53	53	53
923	37	34	0.283	0.193	0.456	0.23	53	52	55	52
924	35	33	-0.640	0.533	-0.851	0.569	44	55	41	56
925	48	52	0.495	0.781	0.62	0.815	55	58	56	58
926	40	46	-0.041	0.600	-0.065	0.635	50	56	49	56
927	45	37	0.444	0.058	0.309	0.064	54	51	53	51
928	33	52	-0.127	0.702	-0.067	0.986	49	57	49	60
929	27	36	-1.074	-0.403	-0.954	-0.214	39	46	40	48
930	49	40	0.478	0.944	0.352	0.982	55	59	54	60
931	52	38	0.327	0.737	0.18	0.84	53	57	52	58
932	58	38	0.285	0.275	0.362	0.233	53	53	54	52
933	37	37	-0.572	0.021	-0.324	0.225	44	50	47	52
934	35	39	0.077	0.536	0.131	0.547	51	55	51	55

935	25	38	-0.870	-0.263	-0.84	-0.144	41	47	42	49
936	33	38	-0.266	0.075	-0.045	0.16	47	51	50	52
937	30	25	-0.499	-0.181	-0.28	-0.181	45	48	47	48
938	40	42	-0.519	0.358	-0.324	0.427	45	54	47	54
939	30	33	-0.499	0.122	-0.28	0.088	45	51	47	51
940	38	50	0.323	0.659	0.321	0.885	53	57	53	59
941	37	43	-0.482	0.144	-0.47	0.248	45	51	45	52
942	33	30	-0.059	-0.492	-0.033	-0.275	49	45	50	47
943	30	28	-0.499	-0.040	-0.28	-0.009	45	50	47	50
944	30	23	-0.499	-0.403	-0.28	-0.456	45	46	47	45
945	35	50	-0.232	-0.089	0	0.084	48	49	50	51
946	38	35	-0.587	-0.193	-0.492	-0.402	44	48	45	46
947	37	26	-0.477	0.177	-0.307	-0.038	45	52	47	50
948	38	29	-0.306	0.184	-0.095	0.122	47	52	49	51
949	37	48	-0.272	0.219	-0.021	0.469	47	52	50	55
950	48	30	0.741	0.372	0.733	0.278	57	54	57	53
951	37	35	-0.154	0.045	0.084	0.197	48	50	51	52
952	42	37	-0.049	0.485	0.118	0.457	50	55	51	55
953	37	27	-0.346	-0.043	-0.161	0.095	47	50	48	51
954	40	45	-0.291	0.398	0	0.404	47	54	50	54
955	38	27	0.039	-0.681	0.244	-0.694	50	43	52	43
956	40	38	0.221	0.396	0.289	0.473	52	54	53	55
957	42	27	-0.164	-0.147	-0.035	-0.341	48	49	50	47
958	42	32	-0.387	-0.097	-0.128	-0.023	46	49	49	50
959	38	32	-0.312	0.114	-0.027	0.115	47	51	50	51
960	28	33	-1.115	0.033	-1.134	-0.096	39	50	39	49
961	35	31	-0.232	0.108	0	0.153	48	51	50	52
962	42	33	0.320	-0.160	0.257	-0.074	53	48	53	49
963	35	33	-0.232	0.030	0	0.055	48	50	50	51
964	40	34	0.267	0.450	0.5	0.494	53	55	55	55
965	35	35	-0.301	-0.466	-0.125	-0.571	47	45	49	44
966	25	33	-0.754	-0.110	-0.593	-0.132	42	49	44	49
967	38	43	-0.306	-0.043	-0.095	0.148	47	50	49	51
968	40	33	-0.235	-0.573	-0.093	-0.702	48	44	49	43
969	35	27	-0.232	-0.634	0	-0.602	48	44	50	44
970	40	31	-0.349	-0.874	-0.089	-0.601	47	41	49	44
971	40	32	0.085	0.114	0.27	0.115	51	51	53	51
972	35	33	-0.232	0.033	0	-0.096	48	50	50	49
973	33	31	-0.317	0.108	-0.084	0.153	47	51	49	52
974	35	39	-0.232	-0.437	0	-0.148	48	46	50	49
975	38	35	0.126	-0.017	0.339	-0.124	51	50	53	49
976	38	31	-0.183	0.108	0.03	0.153	48	51	50	52
977	42	35	-0.292	-0.213	-0.571	-0.116	47	48	44	49

978	35	37	-0.232	-0.071	0	-0.018	48	49	50	50
979	38	36	-0.317	0.403	-0.09	0.445	47	54	49	54
980	35	37	-0.232	-0.523	0	-0.6	48	45	50	44
981	38	37	0.179	-0.096	0.202	-0.204	52	49	52	48
982	38	43	-0.180	0.718	0	0.751	48	57	50	58
983	38	38	0.015	0.451	0.219	0.546	50	55	52	55
984	44	49	-0.020	1.268	-0.124	1.272	50	63	49	63
985	39	35	-0.235	-0.310	-0.244	-0.06	48	47	48	49
986	37	40	-0.331	0.277	-0.353	0.382	47	53	46	54
987	33	38	-0.758	-0.804	-0.889	-0.584	42	42	41	44
988	59	43	0.946	0.533	0.823	0.608	59	55	58	56
989	27	53	-1.175	1.295	-1.391	1.496	38	63	36	65
990	55	57	1.011	1.522	0.911	1.635	60	65	59	66
991	41	37	-0.128	0.291	-0.208	0.178	49	53	48	52
992	44	37	-0.020	0.299	-0.124	0.36	50	53	49	54
993	39	50	-0.235	0.824	-0.244	1.036	48	58	48	60
994	37	40	-0.331	0.241	-0.353	0.208	47	52	46	52
995	33	38	-0.758	-0.082	-0.889	0.147	42	49	41	51
996	59	35	0.946	-0.178	0.823	0.101	59	48	58	51
997	18	41	-0.852	0.404	-0.811	0.445	41	54	42	54
998	44	33	0.065	-0.466	-0.067	-0.291	51	45	49	47
999	43	48	0.084	1.145	-0.049	1.003	51	61	50	60
1000	44	33	0.035	-0.060	-0.042	0.111	50	49	50	51
1001	38	42	-0.478	-0.887	-0.59	-0.948	45	41	44	41
1002	41	31	-0.205	-0.255	-0.324	-0.327	48	47	47	47
1003	16	42	-2.015	-1.171	-2.109	-1.501	30	38	29	35
1004	48	34	-0.126	-0.658	-0.279	-0.817	49	43	47	42
1005	46	33	0.021	-0.109	-0.31	-0.037	50	49	47	50
1006	38	52	0.466	-0.527	0.363	-0.561	55	45	54	44
1007	18	46	-2.393	-0.840	-2.459	-0.82	26	42	25	42
1008	42	37	-0.032	-0.052	-0.179	0.078	50	49	48	51
1009	24	52	-0.884	-0.855	-0.836	-0.665	41	41	42	43
1010	44	36	0.057	-0.416	-0.037	-0.349	51	46	50	47
1011	18	43	-1.538	-1.711	-1.555	-1.81	35	33	34	32
1012	17	38	-1.862	-0.298	-1.909	-0.261	31	47	31	47
1013	33	41	-0.265	-0.444	-0.474	-0.49	47	46	45	45
1014	27	37	1.019	-0.871	-1.294	-0.818	60	41	37	42
1015	28	39	-1.255	-0.692	-1.48	-0.353	37	43	35	46
1016	22	38	-1.484	-0.322	-1.61	-0.258	35	47	34	47
1017	27	41	-0.140	-0.255	-0.138	-0.12	49	47	49	49
1018	25	25	-0.299	-0.854	-0.348	-0.982	47	41	47	40
1019	28	42	-0.006	-0.616	-0.07	-0.616	50	44	49	44
1020	35	33	-0.164	-0.433	-0.257	-0.457	48	46	47	45

1021	27	42	-0.119	0.145	-0.171	-0.092	49	51	48	49
1022	28	34	-0.821	0.333	-1.037	0.176	42	53	40	52
1023	57	33	2.117	-0.262	2.216	-0.159	71	47	72	48
1024	60	53	2.227	-0.010	2.341	0.081	72	50	73	51
1025	38	43	0.231	-1.302	0.226	-1.541	52	37	52	35
1026	32	31	-0.686	-0.075	-0.817	0.003	43	49	42	50
1027	43	42	0.484	-0.186	0.525	-0.169	55	48	55	48
1028	45	34	0.685	-0.299	0.445	-0.335	57	47	54	47
1029	51	33	1.127	-0.211	1.044	-0.115	61	48	60	49
1030	56	42	2.049	-0.938	1.999	-1.12	70	41	70	39
1031	53	31	1.970	-1.175	1.901	-0.992	70	38	69	40
1032	33	42	0.335	-0.742	0.21	-0.845	53	43	52	42
1033	33	34	-0.045	-0.861	-0.056	-0.78	50	41	49	42
1034	37	33	-0.168	-1.160	-0.176	-0.91	48	38	48	41
1035	28	52	0.800	-0.943	0.826	-1.028	58	41	58	40
1036	32	42	-0.346	-0.737	-0.293	-0.35	47	43	47	47
1037	33	31	-0.231	-0.040	-0.056	-0.06	48	50	49	49
1038	33	42	-0.231	-0.643	-0.056	-0.454	48	44	49	45
1039	30	34	-0.383	-0.136	-0.391	-0.301	46	49	46	47
1040	32	33	0.020	0.419	0	0.553	50	54	50	56
1041	28	52	-0.263	-0.025	-0.143	-0.075	47	50	49	49
1042	40	42	0.640	0.486	0.635	0.478	56	55	56	55
1043	32	31	-0.367	0.292	-0.274	0.254	46	53	47	53
1044	25	42	-0.665	0.238	-0.588	0.378	43	52	44	54
1045	33	34	-0.304	0.252	-0.428	0.347	47	53	46	53
1046	36	33	-0.095	-0.616	-0.11	-0.628	49	44	49	44
1047	30	42	-0.567	0.553	-0.486	-0.486	44	56	45	45
1048	38	31	-0.071	-0.166	-0.181	-0.187	49	48	48	48
1049	37	42	0.096	1.433	0.325	1.413	51	64	53	64
1050	39	34	0.077	-0.388	-0.037	-0.23	51	46	50	48
1051	34	33	-0.121	0.547	-0.141	0.484	49	55	49	55
1052	34	52	-0.173	-0.907	-0.267	-1.046	48	41	47	40
1053	38	42	0.604	-0.314	0.531	-0.088	56	47	55	49
1054	37	31	-0.157	-0.929	-0.177	-0.808	48	41	48	42
1055	38	42	-0.431	0.323	-0.481	0.358	46	53	45	54
1056	43	34	-0.227	0.421	-0.235	0.538	48	54	48	55
1057	48	33	0.630	-0.488	0.634	-0.333	56	45	56	47
1058	42	52	0.491	-0.230	0.611	-0.044	55	48	56	50
1059	39	48	0.478	0.275	0.456	0.46	55	53	55	55
1060	38	43	0.492	0.457	0.435	0.406	55	55	54	54
1061	38	58	0.327	-0.067	0.403	-0.083	53	49	54	49
1062	33	78	-1.110	0.703	-1.046	0.628	39	57	40	56
1063	27	43	-0.703	0.686	-0.78	0.46	43	57	42	55

1064	40	50	-0.151	0.763	0.019	0.778	48	58	50	58
1065	38	37	0.268	-0.182	0.275	-0.084	53	48	53	49
1066	38	45	-0.087	0.196	-0.044	0.415	49	52	50	54
1067	38	43	0.492	0.276	0.435	0.426	55	53	54	54
1068	43	35	0.209	0.266	0.189	0.243	52	53	52	52
1069	28	52	-0.696	-0.479	-0.745	-0.55	43	45	43	45
1070	38	67	0.516	0.074	0.486	0.026	55	51	55	50
1071	40	69	-0.063	0.550	-0.09	0.756	49	56	49	58
1072	34	37	0.018	-0.002	-0.059	0.101	50	50	49	51
1073	38	33	0.344	0.402	0.33	0.573	53	54	53	56
1074	40	23	0.209	-0.378	0.191	-0.296	52	46	52	47
1075	35	20	0.032	0.010	0	0.224	50	50	50	52
1076	39	27	-0.124	-0.287	-0.065	-0.244	49	47	49	48
1077	17	25	-1.210	-0.476	-1.221	-0.333	38	45	38	47
1078	48	55	0.758	-0.535	0.776	-0.416	58	45	58	46
1079	33	38	-0.163	0.153	-0.246	0.147	48	52	48	51
1080	38	33	0.222	-0.633	0.205	-0.486	52	44	52	45
1081	33	52	-0.076	-0.470	0	-0.244	49	45	50	48
1082	40	25	-0.032	-0.211	-0.122	-0.147	50	48	49	49
1083	40	51	0.416	0.307	0.39	0.262	54	53	54	53
1084	38	52	0.085	-0.570	0.08	-0.545	51	44	51	45
1085	34	33	-0.213	0.042	-0.315	0.034	48	50	47	50
1086	28	37	-0.589	-0.834	-0.654	-0.685	44	42	43	43
1087	33	47	-0.669	-0.241	-0.693	-0.175	43	48	43	48
1088	33	43	-0.575	0.219	-0.536	0.217	44	52	45	52
1089	28	48	-0.319	-0.319	-0.357	0.194	47	47	46	52
1090	33	28	-0.240	-0.282	-0.07	-0.37	48	47	49	46
1091	27	33	-0.746	-0.533	-0.831	-0.437	43	45	42	46
1092	33	59	-0.536	-0.911	-0.749	-0.706	45	41	43	43
1093	27	38	-0.521	-0.179	-0.473	-0.101	45	48	45	49
1094	44	38	0.372	1.436	0.318	1.708	54	64	53	67
1095	37	59	-0.014	1.697	0.027	1.947	50	67	50	69
1096	39	29	0.271	0.503	0.323	0.415	53	55	53	54
1097	32	16	-0.137	1.760	-0.092	2.01	49	68	49	70
1098	38	45	0.120	0.426	0.16	0.514	51	54	52	55
1099	34	48	-0.279	-1.730	-0.304	-1.345	47	33	47	37
1100	37	43	0.001	-1.350	-0.092	-1.018	50	37	49	40
1101	38	58	0.223	-1.192	0.116	-0.846	52	38	51	42
1102	28	75	-0.554	-0.676	-0.651	-0.627	44	43	43	44
1103	20	43	-0.625	-0.789	-0.384	-0.558	44	42	46	44
1104	32	50	-0.427	0.313	-0.594	0.454	46	53	44	55
1105	37	37	0.606	-1.019	0.475	-0.945	56	40	55	41
1106	26	45	-0.804	-1.486	-0.901	-1.467	42	35	41	35

1107	17	43	-0.948	0.146	-0.94	0.162	41	51	41	52
1108	31	35	-0.141	0.195	-0.202	0.34	49	52	48	53
1109	34	52	-0.121	0.318	-0.266	0.319	49	53	47	53
1110	34	67	-0.297	2.419	-0.38	2.367	47	74	46	74
1111	28	69	-0.209	2.601	-0.321	2.669	48	76	47	77
1112	29	37	-0.301	0.205	-0.45	0.305	47	52	46	53
1113	25	33	-0.591	0.533	-0.627	0.424	44	55	44	54
1114	33	23	-0.659	0.033	-0.639	0.156	43	50	44	52
1115	33	20	0.017	-0.430	0.024	-0.481	50	46	50	45
1116	33	27	-0.185	0.013	-0.168	-0.096	48	50	48	49
1117	42	25	-0.030	0.099	-0.059	0.132	50	51	49	51
1118	25	55	-1.091	1.666	-1.014	1.844	39	67	40	68
1119	21	38	-1.506	0.991	-1.379	1.016	35	60	36	60
1120	22	33	-1.477	0.230	-1.623	0.372	35	52	34	54
1121	27	52	-0.140	1.496	-0.138	1.367	49	65	49	64
1122	25	25	-0.299	-0.138	-0.348	0.037	47	49	47	50
1123	34	51	-0.012	1.453	-0.13	1.456	50	65	49	65
1124	33	52	-0.124	1.663	-0.228	1.579	49	67	48	66
1125	33	33	-0.124	0.654	-0.228	0.451	49	57	48	55
1126	21	37	-0.813	0.565	-0.989	0.52	42	56	40	55
1127	56	47	1.979	1.213	1.896	1.197	70	62	69	62
1128	30	43	-0.620	1.112	-0.786	1.102	44	61	42	61
1129	63	48	1.974	1.458	1.895	1.375	70	65	69	64
1130	37	28	0.233	0.211	0.253	0.316	52	52	53	53
1131	38	33	0.648	0.868	0.545	0.875	56	59	55	59
1132	47	59	0.902	2.209	0.773	2.27	59	72	58	73
1133	51	38	1.127	-0.363	1.044	-0.419	61	46	60	46
1134	53	38	2.108	-0.064	2.103	0.103	71	49	71	51
1135	51	59	1.724	2.209	1.569	2.27	67	72	66	73
1136	37	29	0.362	0.161	0.262	-0.038	54	52	53	50
1137	35	16	0.260	-0.782	0.27	-0.989	53	42	53	40
1138	34	33	0.003	0.654	0.021	0.451	50	57	50	55
1139	51	37	1.278	0.565	1.256	0.52	63	56	63	55
1140	55	47	1.670	1.213	1.635	1.197	67	62	66	62
1141	54	43	1.455	1.112	1.381	1.102	65	61	64	61
1142	45	48	0.607	1.458	0.488	1.375	56	65	55	64
1143	53	28	1.567	0.211	1.55	0.316	66	52	66	53
1144	25	33	-1.040	0.868	-1.179	0.875	40	59	38	59
1145	53	59	1.567	2.209	1.55	2.27	66	72	66	73
1146	43	38	0.784	-0.363	0.607	-0.419	58	46	56	46
1147	25	38	-1.091	-0.064	-1.014	0.103	39	49	40	51
1148	21	59	-1.506	2.209	-1.379	2.27	35	72	36	73
1149	37	47	0.362	1.213	0.262	1.197	54	62	53	62

1150	35	43	0.260	1.112	0.27	1.102	53	61	53	61
1151	34	59	0.003	2.209	0.021	2.27	50	72	50	73

Appendix VIII

Marginal fit (X^2) and Standardized LD X^2 Statistics for WAEC CRMAT (GPCM)

		Marginal										
Item	Label	X^2	1	2	3	4	5	6	7	8	9	10
1	Item1	0.2										
2	Item2	0.2	24.4									
3	Item3	0.0	16.0	20.7								
4	Item4	0.1	28.8	19.8	31.9							
5	Item5	0.2	11.6	24.2	14.6	24.4						
6	Item6	0.0	40.9	19.6	26.3	33.2	29.6					
7	Item7	0.0	8.5	12.4	37.1	20.4	13.8	9.4				
8	Item8	0.0	19.6	12.6	18.2	11.9	15.4	58.7	12.6			
9	Item9	0.0	10.4	24.6	9.8	17.9	10.2	21.7	8.9	24.1		
10	Item10	0.0	13.2	16.7	18.0	16.2	17.5	18.3	19.6	21.9	20.4	
11	Item11	0.0	12.1	19.1	18.6	20.8	11.4	19.7	8.0	13.9	14.7	33.8
12	Item12	0.0	8.2	17.6	7.8	14.6	16.1	16.8	7.1	33.2	9.7	17.0
13	Item13	0.0	18.7	5.1	8.9	17.3	9.2	11.6	7.5	13.8	10.0	3.7
14	Item14	0.0	18.6	10.9	18.6	18.5	18.7	31.9	4.3	9.3	11.4	15.3
15	Item15	0.0	10.7	10.9	21.7	11.8	19.7	23.5	20.2	21.4	12.2	12.6

		Marginal				
Item	Label	X^2	11	12	13	14
11	Item11	0.0				
12	Item12	0.0	30.9			
13	Item13	0.0	7.8	8.4		
14	Item14	0.0	18.9	6.8	24.2	
15	Item15	0.0	11.7	7.1	12.7	17.3

Appendix IX

Marginal fit (X^2) and Standardized LD X^2 Statistics for WAEC CRMAT (GRM)

		Marginal										
Item	Label	X^2	1	2	3	4	5	6	7	8	9	10
1	Item1	0.7										
2	Item2	2.1	22.1									
3	Item3	0.1	12.5	20.7								
4	Item4	0.7	26.6	19.5	32.3							
5	Item5	3.3	12.5	26.6	15.1	25.2						
6	Item6	0.1	42.4	20.6	29.2	32.5	28.2					
7	Item7	0.1	9.0	12.3	30.8	19.6	14.0	10.0				
8	Item8	0.0	18.3	13.9	19.9	11.8	16.7	57.3	12.8			
9	Item9	0.0	10.3	25.8	10.6	18.3	11.0	20.7	8.3	23.4		
10	Item10	0.0	13.3	18.0	18.8	16.6	18.3	18.4	18.8	21.9	19.6	
11	Item11	0.7	12.8	21.8	19.4	23.1	14.4	18.5	7.5	13.7	12.2	30.8
12	Item12	0.1	7.9	17.5	9.2	15.6	16.3	15.5	6.4	31.8	10.2	16.1
13	Item13	0.0	20.1	5.5	8.9	17.9	10.7	12.1	6.9	13.9	9.6	3.5
14	Item14	0.1	17.4	10.7	19.3	17.8	19.6	34.1	4.7	9.7	10.8	15.4
15	Item15	0.0	10.6	11.6	22.0	12.1	19.4	24.8	19.8	22.2	12.1	12.4

		Marginal				
Item	Label	X^2	11	12	13	14
11	Item11	0.7				
12	Item12	0.1	33.0			
13	Item13	0.0	7.8	9.8		
14	Item14	0.1	18.4	6.3	23.3	
15	Item15	0.0	11.4	6.9	12.6	16.7

Appendix X

Marginal fit (X^2) and Standardized LD X^2 Statistics for NECO CRMAT (GPCM)

		Marginal										
Item	Label	X^2	1	2	3	4	5	6	7	8	9	10
1	Item1	0.0										
2	Item2	0.2	5.5									
3	Item3	0.0	18.2	11.9								
4	Item4	0.0	21.6	11.0	10.3							
5	Item5	0.0	11.3	10.1	15.5	16.6						
6	Item6	0.0	12.9	8.8	11.6	10.7	15.6					
7	Item7	0.0	18.5	10.3	13.7	14.9	8.9	19.5				
8	Item8	0.0	19.5	17.1	24.5	17.4	11.2	14.2	24.0			
9	Item9	0.1	13.2	17.5	15.5	11.9	11.5	15.2	29.9	23.1		
10	Item10	0.0	19.9	18.4	15.5	20.6	13.8	22.4	16.3	11.7	24.4	
11	Item11	0.1	10.7	9.3	10.9	10.1	12.6	8.2	8.3	13.1	21.8	26.7
12	Item12	0.0	13.6	7.9	10.1	11.9	7.3	13.1	11.7	14.5	14.1	24.8
13	Item13	0.0	9.1	8.5	11.2	4.0	5.9	13.0	8.3	7.2	14.8	12.4
14	Item14	0.0	20.2	3.7	7.9	5.2	8.8	7.7	15.8	9.7	11.9	23.2
15	Item15	0.0	13.5	10.5	12.5	11.9	12.2	18.4	10.8	13.3	16.5	16.9

		Marginal				
Item	Label	X^2	11	12	13	14
11	Item11	0.1				
12	Item12	0.0	14.5			
13	Item13	0.0	13.6	35.5		
14	Item14	0.0	12.0	4.6	7.2	
15	Item15	0.0	15.1	9.3	17.3	31.2

Appendix XI

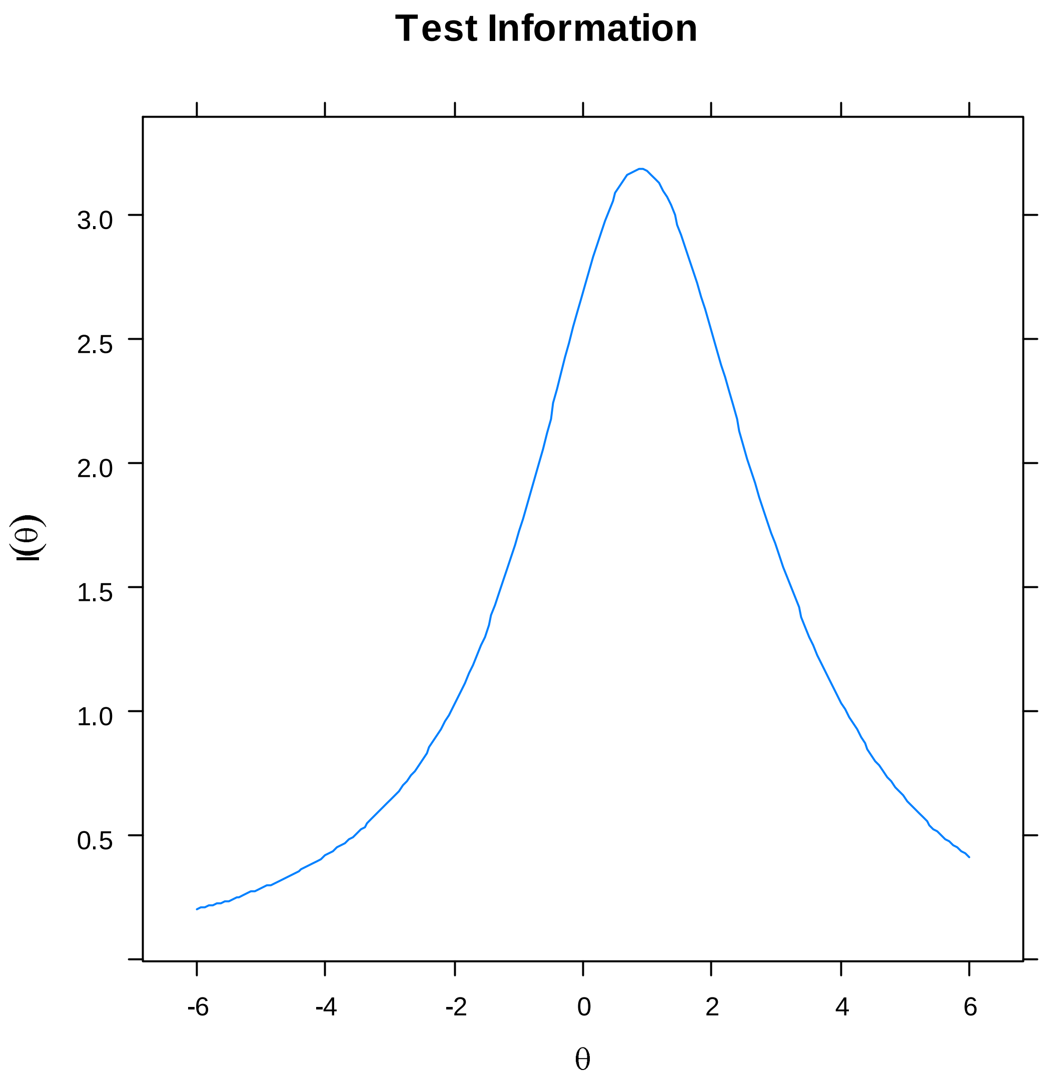
Marginal fit (X^2) and Standardized LD X^2 Statistics for NECO CRMAT (GRM)

		Marginal										
Item	Label	X^2	1	2	3	4	5	6	7	8	9	10
1	Item1	0.3										
2	Item2	0.4	5.2									
3	Item3	0.7	19.1	16.0								
4	Item4	1.1	24.0	12.1	11.8							
5	Item5	0.3	11.0	10.6	15.6	16.9						
6	Item6	0.0	13.4	9.8	12.5	9.8	15.5					
7	Item7	0.6	18.8	10.4	17.5	14.8	9.1	20.2				
8	Item8	0.4	19.5	16.5	24.1	17.7	10.6	14.3	23.5			
9	Item9	3.0	13.4	19.2	14.1	12.5	11.7	14.7	28.6	24.5		
10	Item10	0.6	20.5	19.0	16.1	20.3	13.1	19.3	16.7	12.2	24.5	
11	Item11	2.5	11.8	10.2	11.2	11.3	11.8	7.9	9.3	15.0	21.6	24.9
12	Item12	0.0	13.7	8.1	11.1	11.7	8.2	11.1	12.7	13.8	13.7	23.8
13	Item13	0.0	10.0	10.3	12.0	4.7	5.5	11.3	8.5	7.1	14.5	10.0
14	Item14	0.0	19.7	4.2	8.2	5.9	8.1	7.9	16.5	10.0	11.3	23.3
15	Item15	0.0	12.9	13.3	14.3	12.5	12.2	17.7	11.1	13.3	14.1	16.3

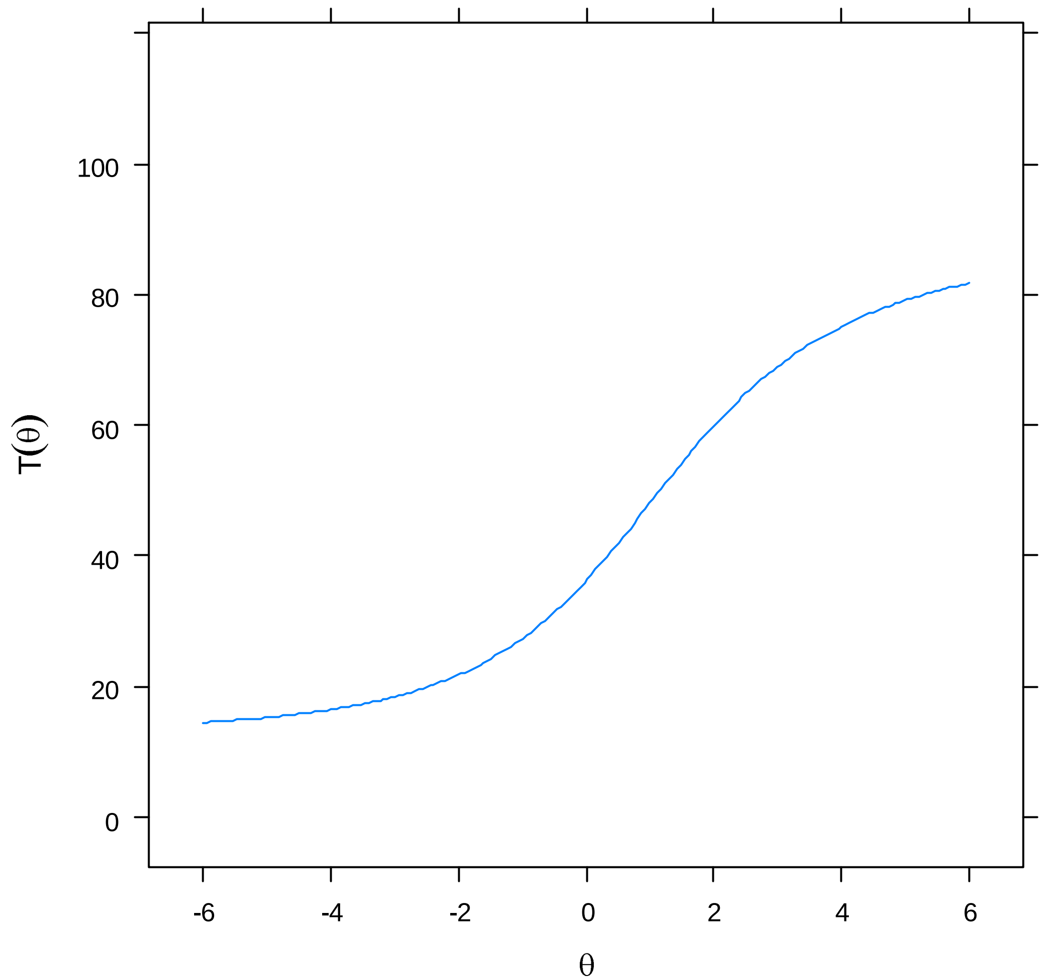
		Marginal				
Item	Label	X^2	11	12	13	14
11	Item11	2.5				
12	Item12	0.0	14.2			
13	Item13	0.0	13.7	35.2		
14	Item14	0.0	10.6	4.6	7.0	
15	Item15	0.0	14.7	9.5	18.2	32.1

Appendix XII

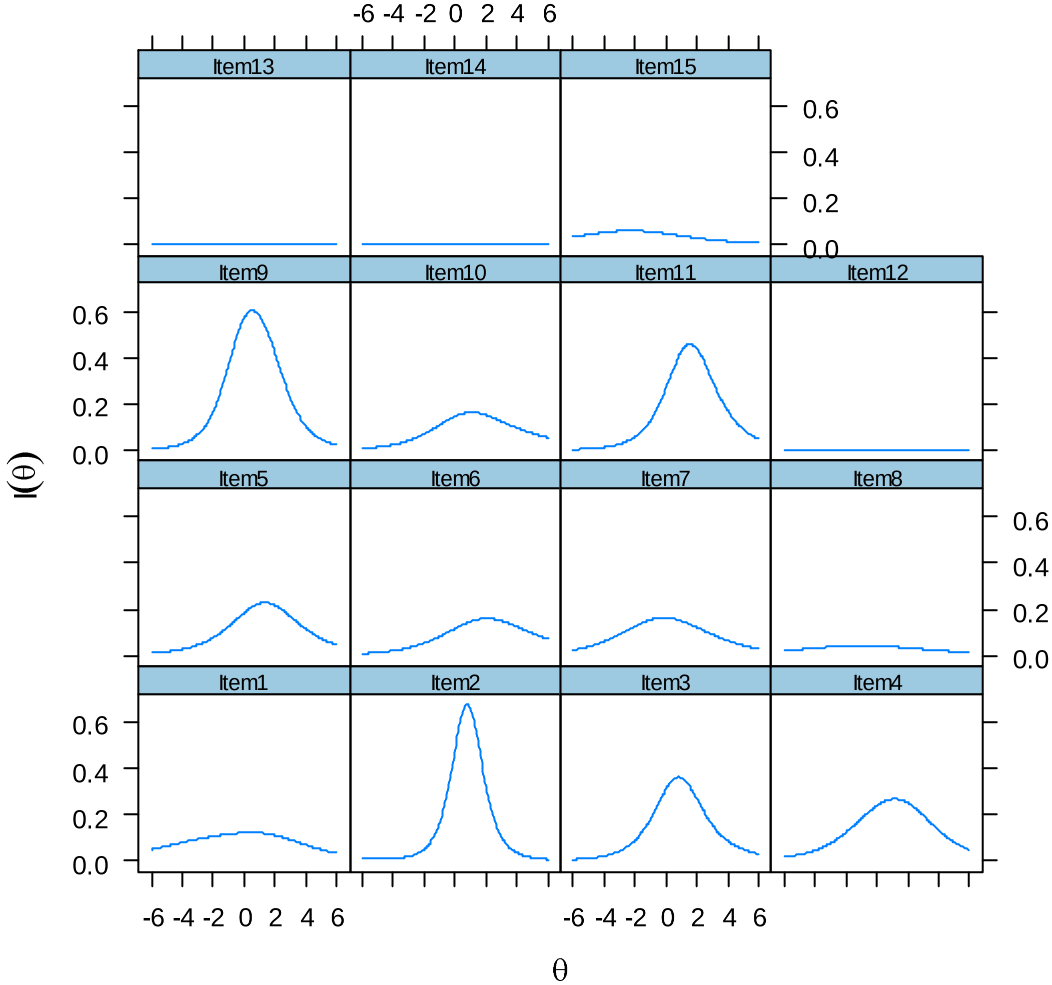
NECO Items graphs

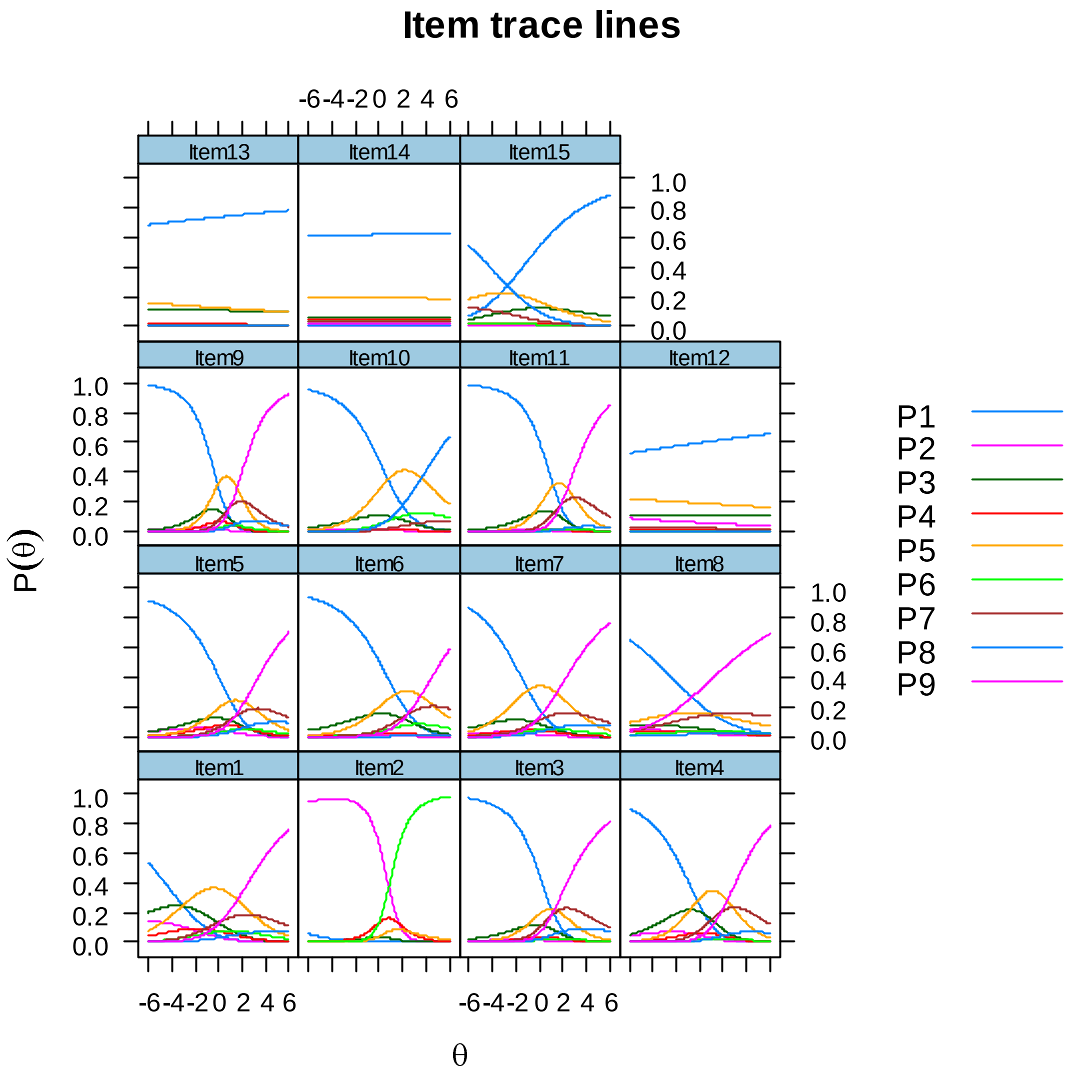


Expected Total Score



Item information trace lines

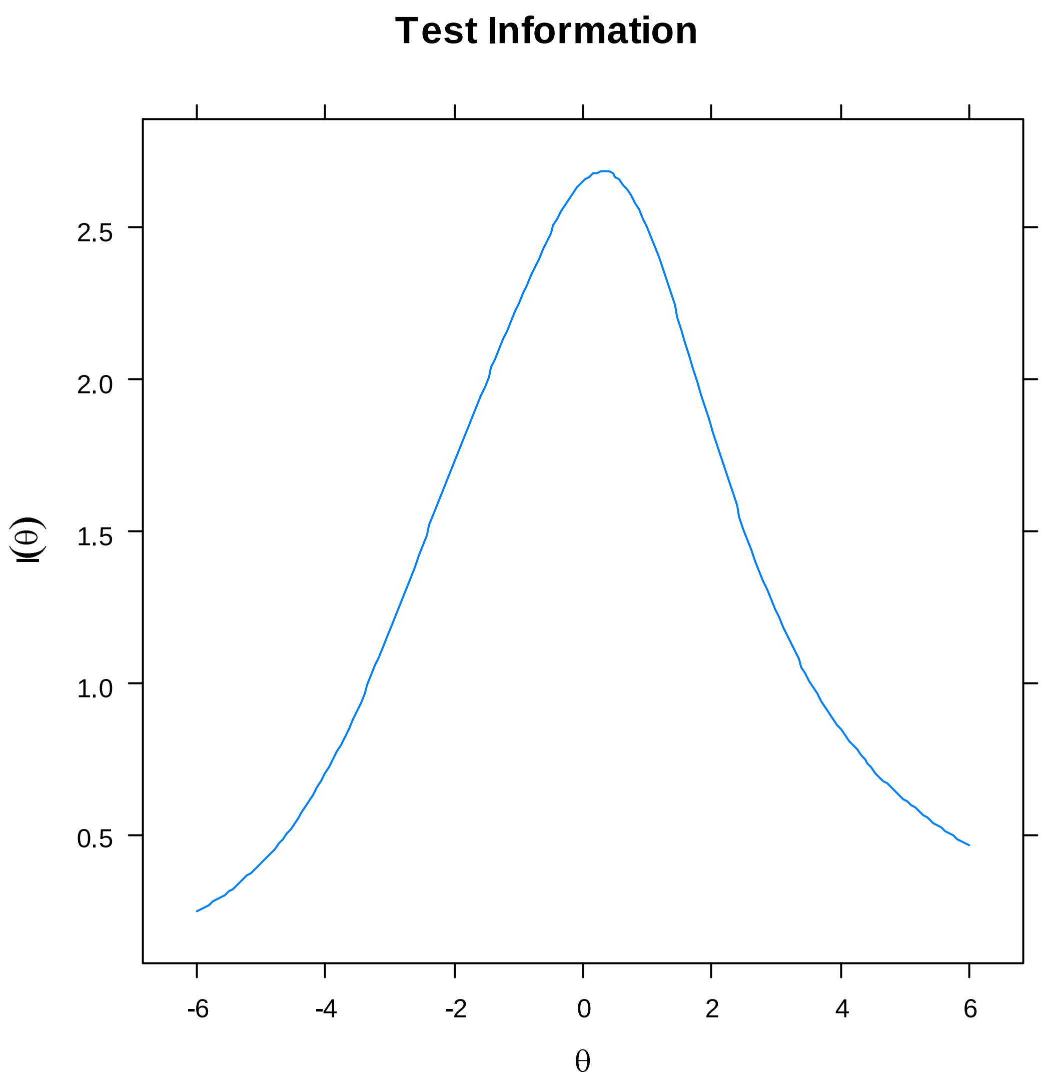


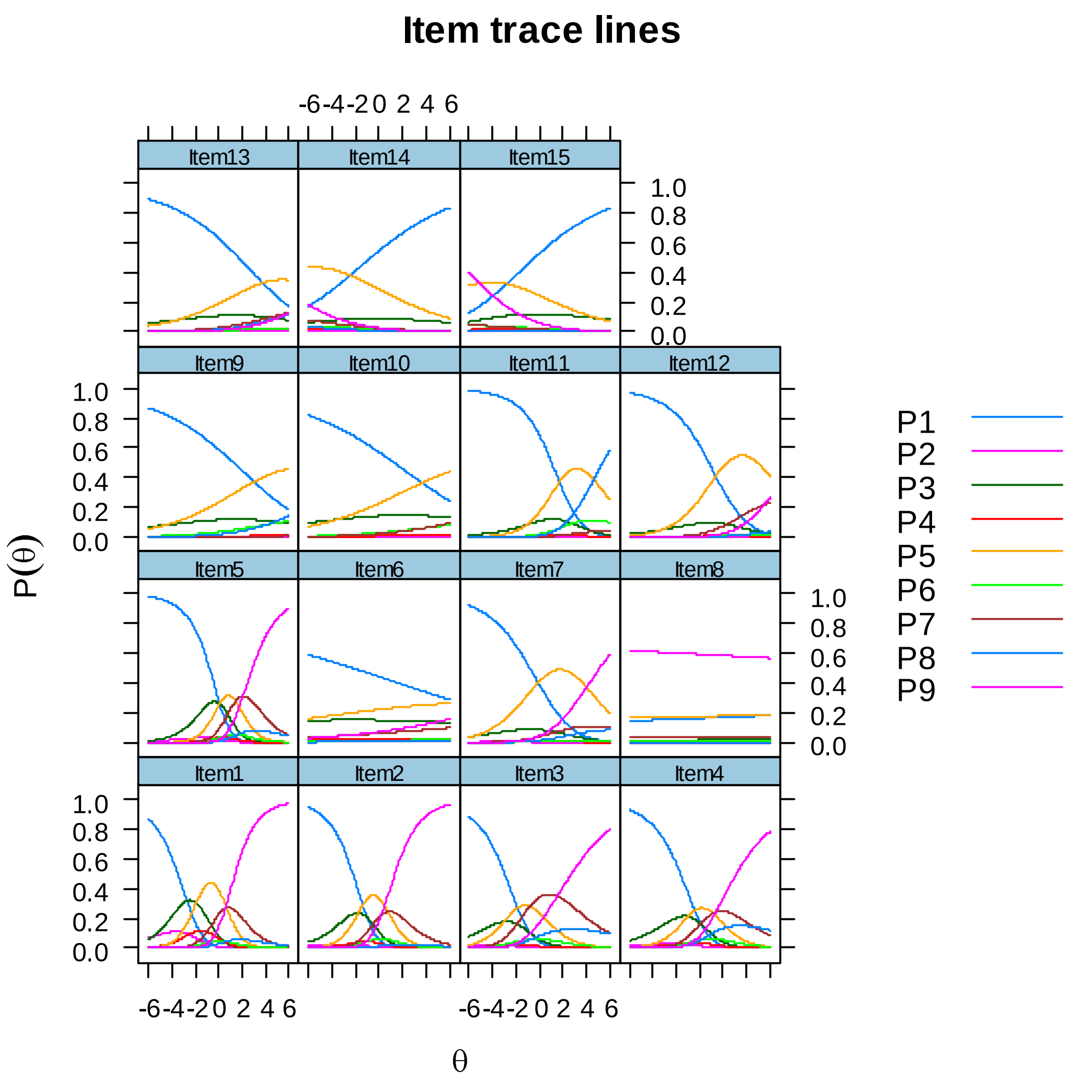


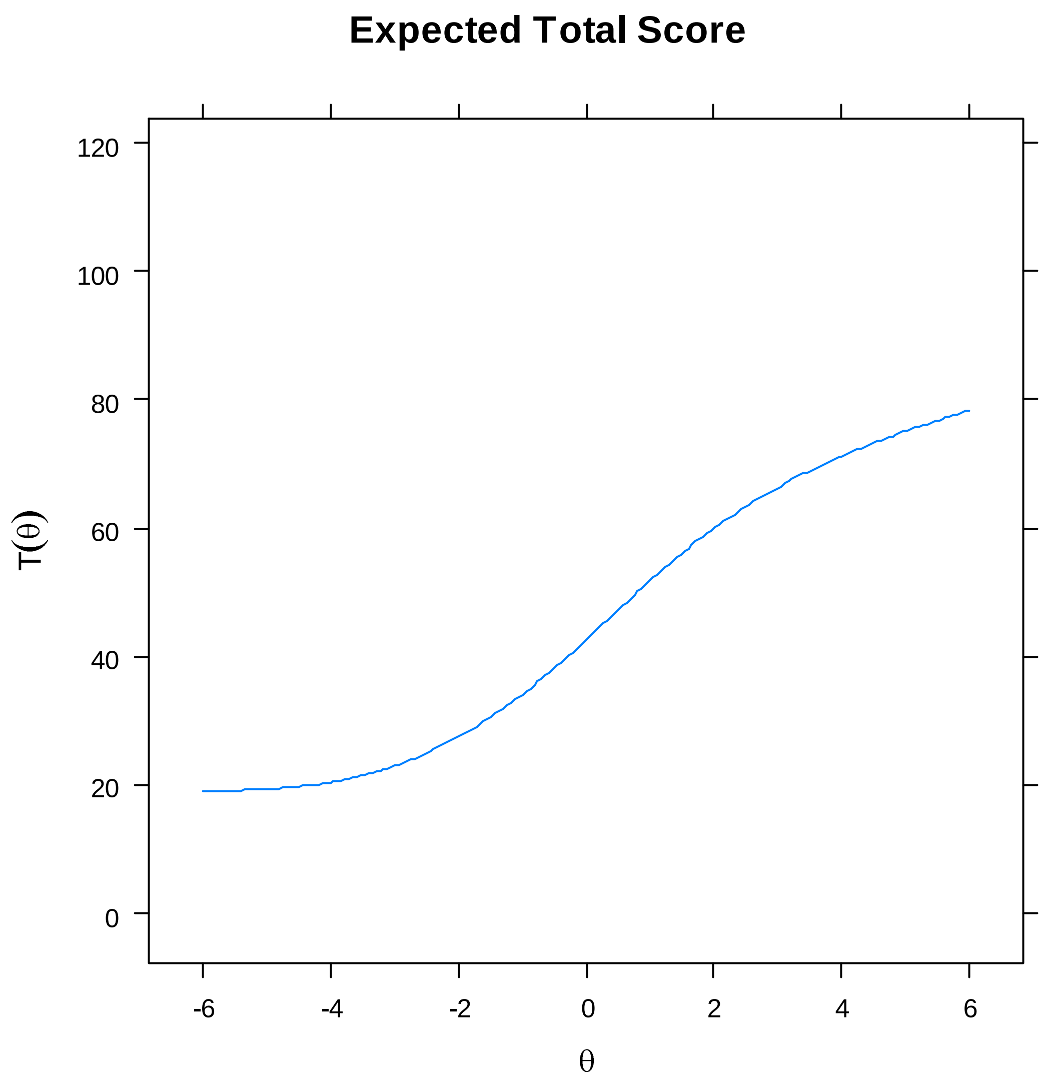
Item trace lines
Item13
Item14
Item15
Item9
Item10
Item11
Item12
Item5
Item6
Item7
Item8
Item1
Item2
Item3
Item4
P(θ)
θ
-6 -4 -2 0 2 4 6
0.0
0.2
0.4
0.6
0.8
1.0
P1
P2
P3
P4
P5
P6
P7
P8
P9

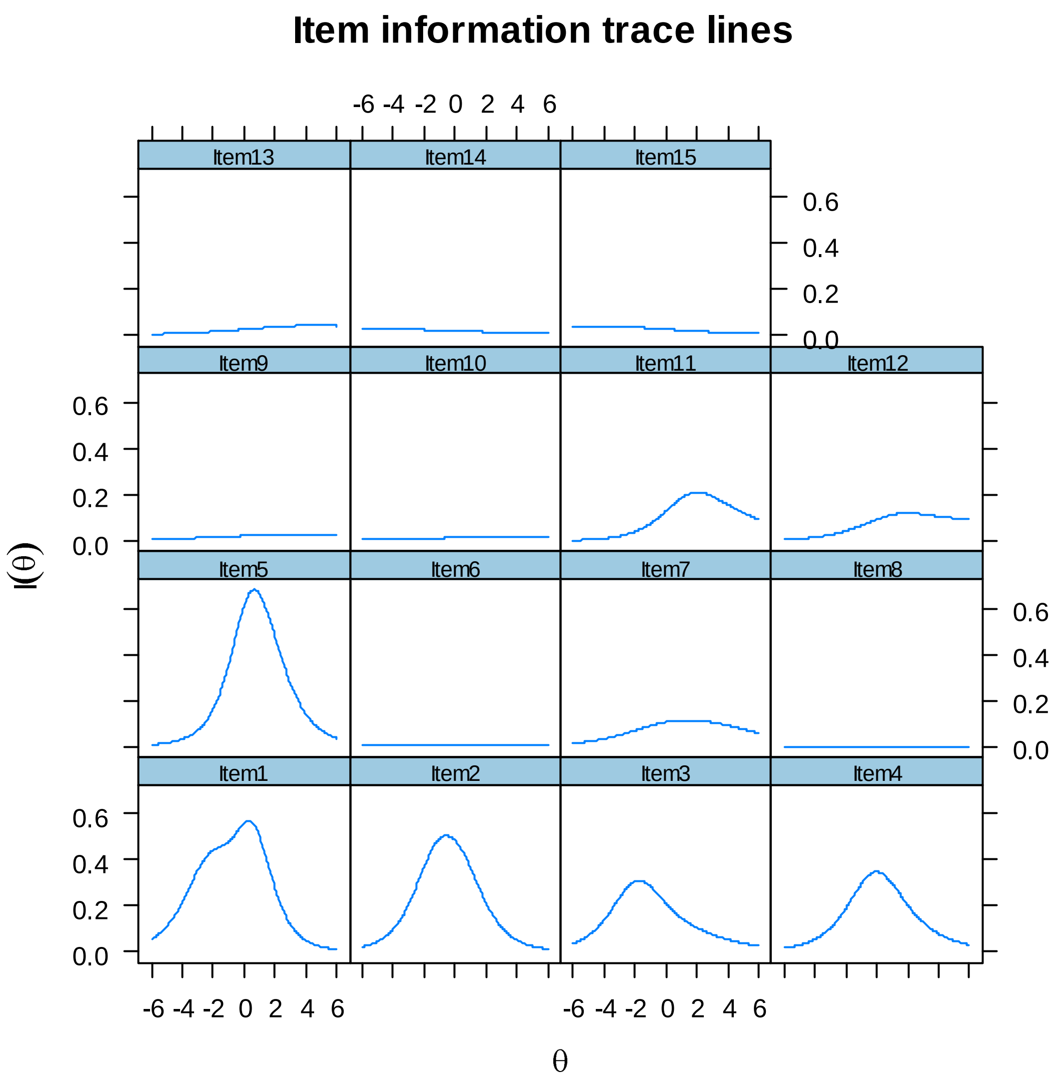
APPENDIXES XIII

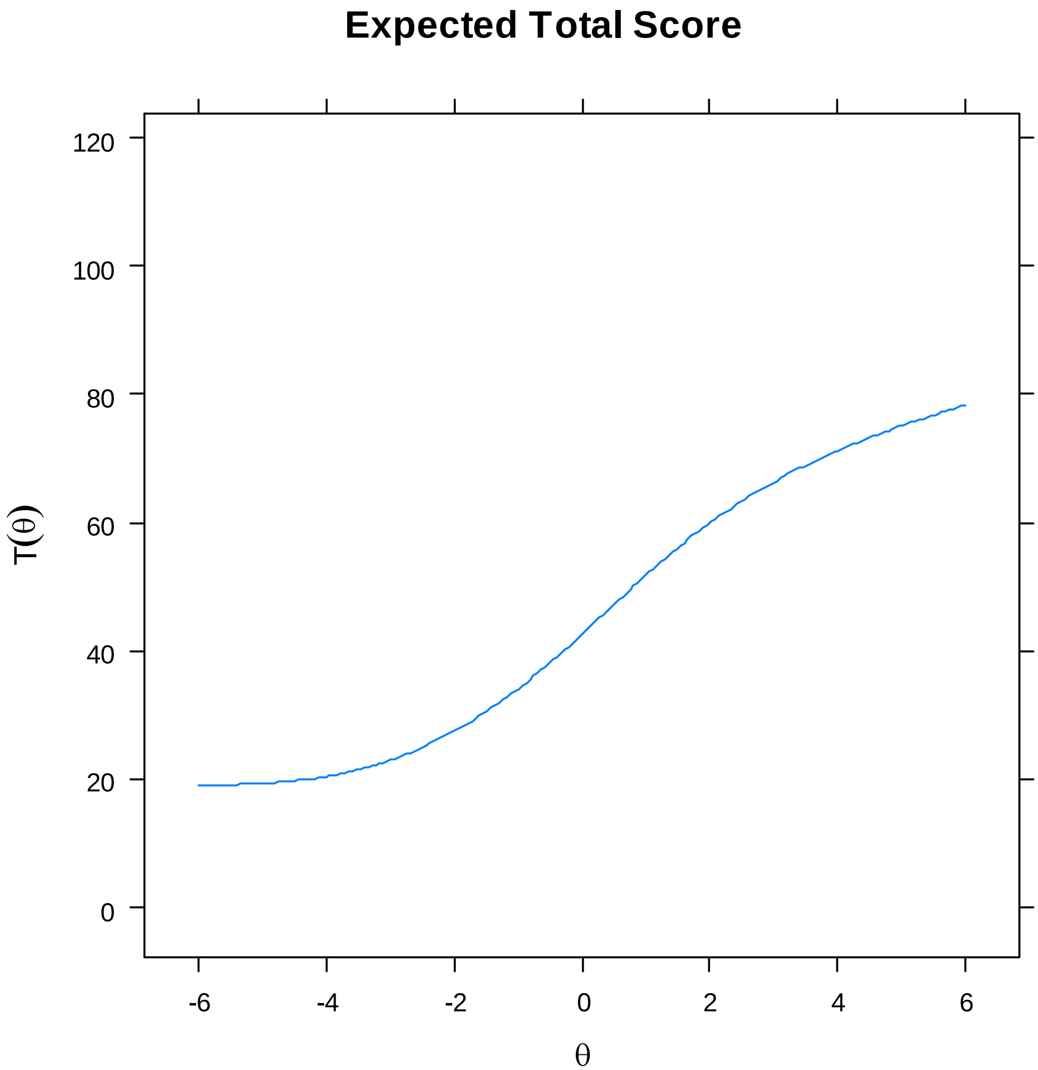
WAEC test Info



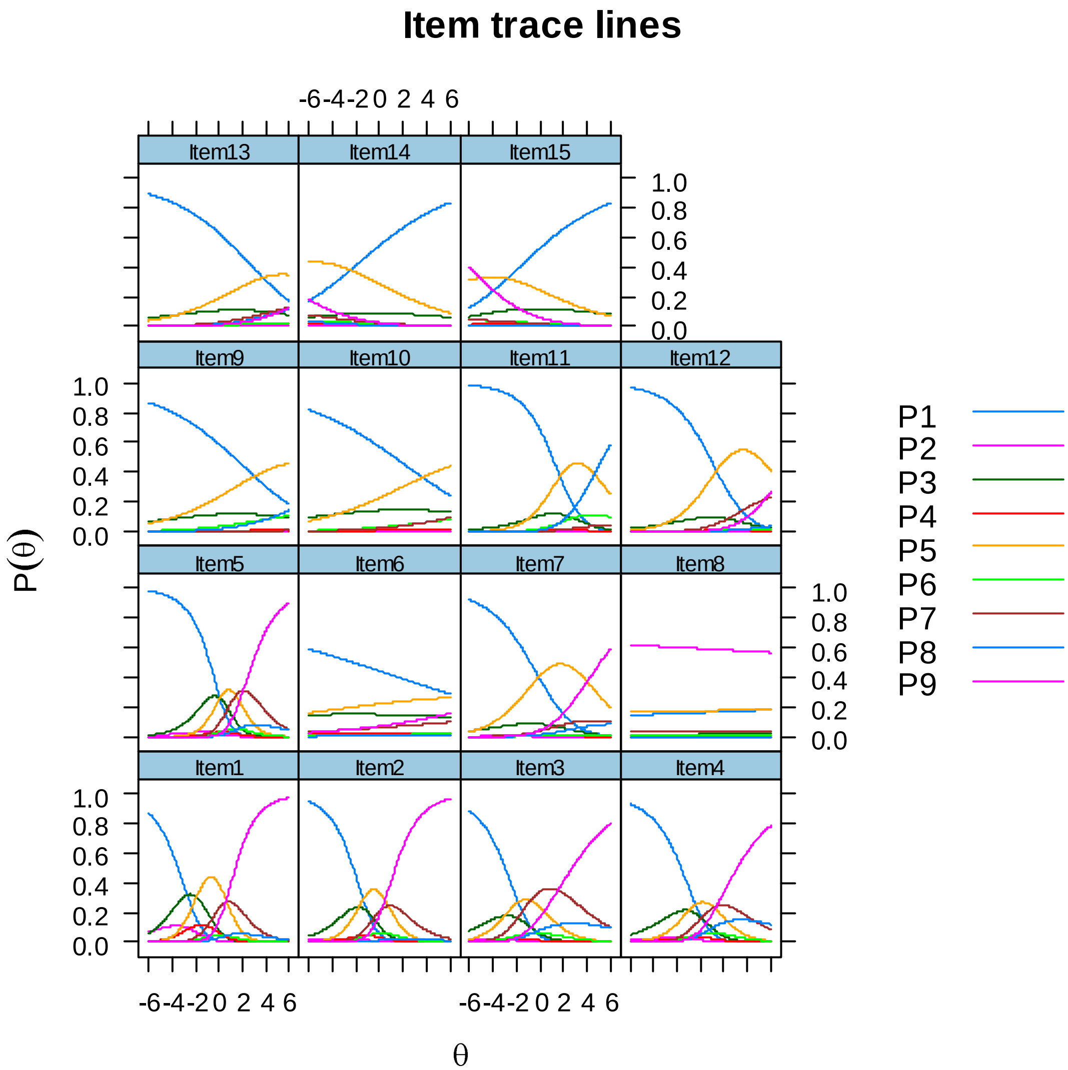

Item trace lines
Item13
Item14
Item15
Item9
Item10
Item11
Item12
Item5
Item6
Item7
Item8
Item1
Item2
Item3
Item4
-6-4-2 0 2 4 6
1.0
0.8
0.6
0.4
0.2
0.0
P(θ)
θ
P1
P2
P3
P4
P5
P6
P7
P8
P9

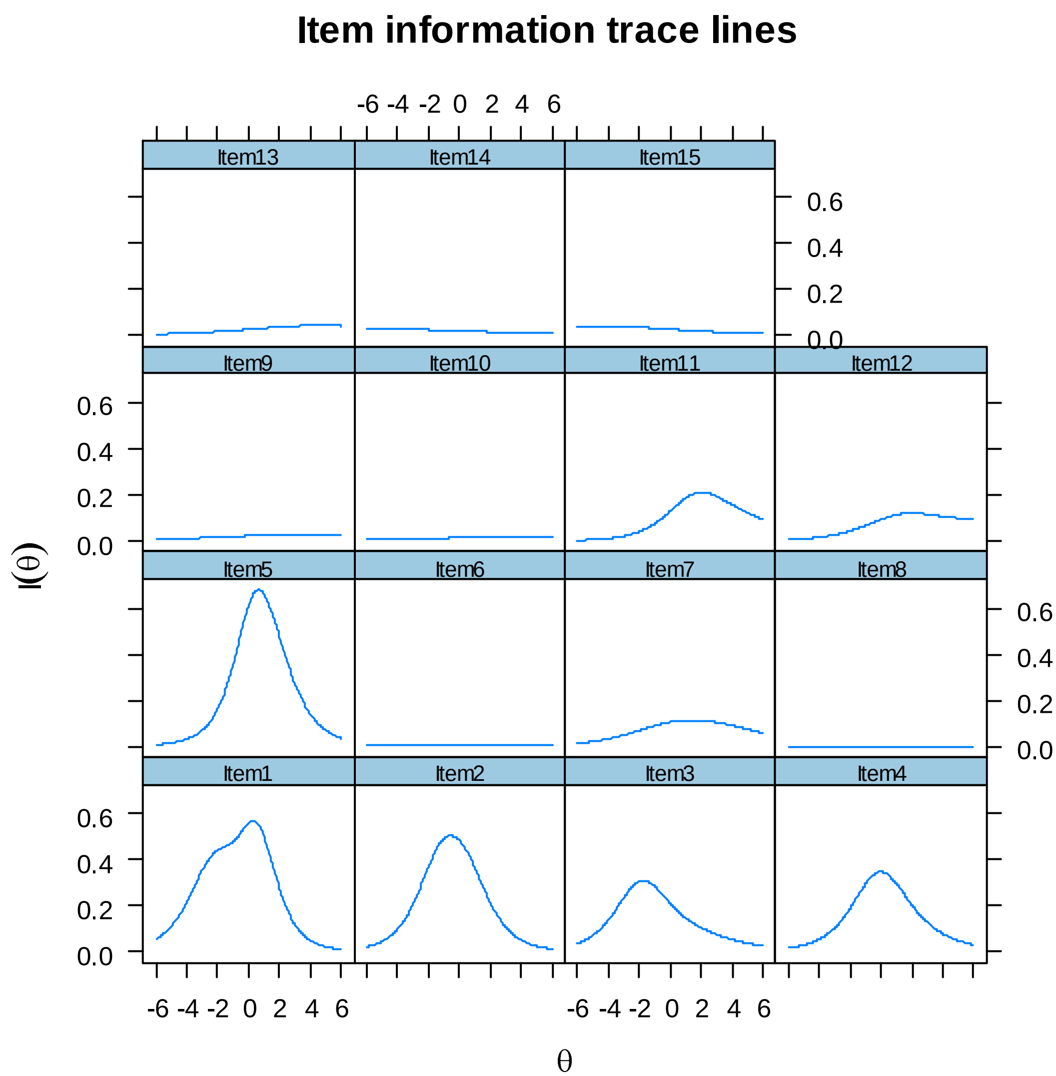

Expected Total Score
T(θ)
120
100
80
60
40
20
0
-6
-4
-2
0
2
4
6
θ


Item information trace lines
Item13
Item14
Item15
Item9
Item10
Item11
Item12
Item5
Item6
Item7
Item8
Item1
Item2
Item3
Item4
I(θ)
θ
0.0
0.2
0.4
0.6
-6 -4 -2 0 2 4 6



WAEC


Item trace lines
Item13
Item14
Item15
Item9
Item10
Item11
Item12
Item5
Item6
Item7
Item8
Item1
Item2
Item3
Item4
P(θ)
θ
-6-4-2 0 2 4 6
0.0
0.2
0.4
0.6
0.8
1.0
P1
P2
P3
P4
P5
P6
P7
P8
P9


Item information trace lines
Item13
Item14
Item15
Item9
Item10
Item11
Item12
Item5
Item6
Item7
Item8
Item1
Item2
Item3
Item4
I(θ)
θ
-6 -4 -2 0 2 4 6
0.0
0.2
0.4
0.6

APPENDIX XIV

Model-data-fit Assessment for WAEC Mathematics Constructed-response Achievement Test

Criterion	MODELS	
	GRM	GPCM
-2loglikelihood:	48411.17*	48450.68
Akaike Information Criterion (AIC):	48673.17	48712.68
Bayesian Information Criterion (BIC):	49334.51	49374.02

*Smaller value indicates a better fit

According to De Ayala (2009), the -2LL can be used to test the fit of competing IRT models. Thus, a smaller value of loglikelihood indicates a better fit. Therefore, the graded response model fits the data significantly better than the generalized partial credit model of the WAEC items. One reason for this was its capability to capture the variability in item discrimination. Also, marginal reliability obtained from this instrument was 0.72 and it reflected an average accuracy across the continuum. However, it is only when the total information function is somewhat uniformly distributed that this value accurately characterizes the precision of measurement across the continuum.

To further verify the model data fit analysis the following IRT statistics were obtained: ***b1, b2, b3, b4, b5, b6, b7, b8*** represents the category boundaries while ***a*** represents the discrimination indices were considered.

Item Statistics for WAEC Mathematics Constructed-response Achievement Test (GPCM)

S/No	Items	a	b1	b2	b3	b4	b5	b6	b7	b8
1	Item1	0.41	0.24	-5.66	0.53	-4.36	5.06	-3.72	5.08	-3.97
2	Item2	0.32	7.51	-9.52	3.85	-7.40	5.33	-3.84	8.69	-8.33
3	Item3	0.23	12.00	-13.44	6.94	-12.95	5.79	-7.43	6.28	-3.23
4	Item4	0.24	10.74	-10.46	6.57	-8.53	6.63	-4.39	4.65	-1.90
5	Item5	0.37	6.70	-6.53	5.66	-5.51	5.81	-2.96	6.20	-2.07
6	Item6	0.03	176.67	-133.79	69.73	-85.13	110.2	-59.21	92.3	-

							4		3	98.77
7	Item7	0.14	36.07	-25.19	20.01	-31.31	28.71	-11.78	7.22	-8.04
8	Item8	-0.00	-1609.57	905.70	-563.02	1290.77	1205.20	755.68	-838.93	1661.23
9	Item9	0.08	84.87	-63.68	38.38	-47.53	24.87	39.38	-31.12	NA
10	Item10	0.06	102.87	-80.86	46.04	-53.39	32.61	3.92	NA	NA
11	Item11	0.21	28.62	-19.68	13.26	-15.38	10.54	9.97	-6.59	NA
12	Item12	0.16	41.00	-29.04	16.09	-22.61	27.68	-12.01	17.63	-6.66
13	Item13	0.08	59.43	-37.37	26.06	-33.19	39.95	-15.46	6.90	5.21
14	Item14	-0.07	-98.47	70.71	-30.31	48.20	-43.80	5.16	-18.91	19.89
15	Item15	-0.07	-68.56	47.27	-30.60	40.81	-35.27	-4.35	-41.69	60.89

The adopted WAEC Mathematics constructed-response test items were scored over 8 with 9 categories including 0, 1, 2, 3, 4, 5, 6, 7 and 8. It was observed from Table that columns 3 and 4 upward gave the discrimination (a) index with values ranging from -0.00 to 0.41 and category boundaries (b1, b2…, b8) of polytomous item response theory model obtained from item calibration of r 3.6.2 software. It was observed that IRT frameworks did not give the estimates of all the item parameters of the WAEC Mathematics CR subjected to the item analysis process. The results suggested that within the IRT framework, all the items were analysed and the items do not fit the generalised partial model

APPENDIX XVI

Item Statistics for WAEC Mathematics Constructed-response Achievement Test (GRM)

S/No	Items	a	b1	b2	b3	b4	b5	b6	b7	b8
1	Item1	1.41	-3.16	-2.79	-1.64	-1.23	0.17	0.32	1.17	1.34
2	Item2	1.21	-2.27	-2.20	-1.37	-1.21	0.06	0.29	1.21	1.29
3	Item3	0.75	-3.81	-3.74	-2.69	-2.54	-0.79	-0.44	1.54	2.21
4	Item4	0.97	-1.60	-1.50	-0.62	-0.46	0.67	0.94	1.97	2.56
5	Item5	1.36	-0.81	-0.72	0.14	0.25	1.27	1.46	2.51	2.76
6	Item6	0.31	-0.76	1.61	1.36	1.74	5.76	6.13	8.26	8.54
7	Item7	0.65	-0.7	-0.71	-0.14	-0.11	3.35	3.47	4.29	4.73
8	Item8	-0.15	10.95	10.90	10.14	10.02	3.93	3.83	2.67	2.59
9	Item9	0.34	1.03	1.04	2.56	2.65	8.40	11.38	11.63	NA
10	Item10	0.26	1.08	1.10	3.56	3.71	11.38	14.69	NA	NA
11	Item11	0.80	0.90	0.91	1.61	1.67	4.17	5.19	5.44	NA
12	Item12	0.64	0.68	0.69	1.33	1.39	5.33	5.49	7.79	8.17
13	Item13	0.25	1.99	2.08	4.03	4.30	10.70	11.25	14.09	17.70
14	Item14	-0.36	-0.50	-0.51	-3.24	-7.18	-7.93	-9.46	-10.13	-10.13
15	Item15	-0.33	-0.28	-0.34	-1.88	-2.07	-6.89	-7.68	-8.40	-8.44

The adopted WAEC Mathematics constructed-response test items were scored over 8 with 9 categories including 0, 1, 2, 3, 4, 5, 6, 7 and 8. It was observed from the table that columns 3 and 4 upward gave the discrimination (a) index with values ranging from -0.36 to 1.41 and category boundaries (b1, b2…, b8) of polytomous item response theory model obtained from item calibration of r 3.6.2 software. It was observed that IRT frameworks did not give the estimates of all the item parameters of the WAEC Mathematics CR subjected to the item analysis process. The results suggested that within the IRT framework, all the items were analysed and the items fitted the graded response model. The model data fit estimates for NECO test items are presented

Model-data fit Assessment for NECO Mathematics Constructed-response Achievement Test

Criterion	**MODELS**	
	GR	GPC
-2loglikelihood:	51307.31	51260.94*
Akaike Information Criterion (AIC):	51565.31	51518.94
Bayesian Information Criterion (BIC):	52216.55	52170.18

*Smaller value indicates a better fit

According to De Ayala (2009), the −2LL can be used to test the fit of competing IRT models. Thus, a smaller value of loglikelihood indicates a better fit. Therefore, the generalized partial credit model fits the data significantly better than the graded response model for the NECO items. The single reason for this was its capability to capture the variability in the item discrimination. Also, marginal reliability obtained from this instrument was 0.70 and it reflected an average accuracy across the continuum. However, it is only when the total information function is somewhat uniformly distributed that this value accurately characterizes the precision of measurement across the continuum. To further verify the model data fit analysis the following IRT statistics were obtained: ***b1, b2, b3, b4, b5, b6, b7, b8*** represents the category boundaries while ***a*** represents the discrimination indices.

Item Statistics for NECO Mathematics Constructed-response Achievement Test (GPCM)

S/No	Items	a	b1	b2	b3	b4	b5	b6	b7	b8
1	Item1	0.17	2.13	-8.18	3.40	-9.62	9.73	-3.77	9.08	-8.706
2	Item2	0.47	-12.01	6.44	-2.98	2.57	2.67	NA	NA	NA
3	Item3	0.21	16.42	-9.96	6.00	-8.31	11.86	-8.78	7.35	-5.29
4	Item4	0.23	7.31	-6.47	5.17	-6.81	13.21	-7.55	8.98	-5.24
5	Item5	0.18	11.65	-5.29	2.81	-5.14	10.37	-4.37	7.63	-4.97
6	Item6	0.16	24.93	-17.44	13.33	-14.97	11.00	-1.60	21.95	-17.184
7	Item7	0.16	16.05	-11.21	4.43	-13.12	12.47	-5.48	7.40	-8.506
8	Item8	0.07	32.77	-12.14	7.73	-23.37	22.90	-21.05	29.19	-41.428
9	Item9	0.33	8.60	-6.39	2.24	-4.76	7.55	-3.24	5.44	-3.544
10	Item10	0.18	19.28	-10.75	13.50	-19.00	9.79	7.84	-6.78	NA
11	Item11	0.26	15.75	-9.87	9.62	-10.70	13.19	-7.20	10.80	-7.145
12	Item12	-0.01	-497.81	340.28	-171.08	224.30	-393.52	197.86	-298.15	387.426
13	Item13	-0.01	-379.32	232.13	-161.92	173.32	-257.61	35.99	-75.13	NA
14	Item14	-0.00	-3082.97	-1981.36	-134.94	685.66	-765.71	-110.16	-813.75	460.181
15	Item15	-0.09	-51.34	34.90	-18.05	22.36	-29.98	13.76	10.28	NA

The adopted NECO mathematics constructed-response test items were scored over 8 with 9 categories including 0, 1, 2, 3, 4,5, 6, 7 and 8 It was observed from the able that columns 2 upward gave the discrimination (a) index with values ranging from -0.00 to 0. 47 and category boundaries (b1, b2…, b8) of polytomous item response theory model obtained from item calibration of r 3.62software. It was observed that IRT frameworks did not give the estimates of all the item parameters of the NECO Mathematics CR subjected to the item analysis process. The results suggested that within the IRT framework, all the items were analysed and the items fitted the generalized partial credit model. Table 4.19 present the calibration of the test items using the graded response model.

APPENDIX XVII

Item Statistics for the NECO Mathematics Constructed-response Achievement Test (GRM)

S/No	Items	a	b1	b2	b3	b4	b5	b6	b7	b8
1	Item1	0.60	-4.95	-4.05	-2.28	-1.63	0.92	1.46	2.79	3.21
2	Item2	1.09	-5.84	0.51	0.64	1.26	1.53	NA	NA	NA
3	Item3	0.79	-0.33	-0.25	0.34	0.51	1.69	1.81	2.90	3.26
4	Item4	0.91	-1.35	-1.07	-0.08	0.21	1.91	2.05	3.21	3.50
5	Item5	0.82	-0.58	-0.32	0.34	0.76	2.10	2.40	3.37	3.79
6	Item6	0.79	0.02	0.07	0.96	1.07	2.77	3.32	4.67	4.74
7	Item7	0.55	-2.59	-2.42	-1.57	-1.20	1.38	1.77	2.92	3.40
8	Item8	0.33	-3.80	-3.36	-2.49	-2.01	0.02	0.45	2.23	2.50
9	Item9	1.45	-0.71	-0.65	-0.20	-0.01	1.22	1.37	2.04	2.23
10	Item10	0.86	-0.16	-0.08	0.47	0.52	2.60	3.43	3.76	NA
11	Item11	1.19	0.29	0.33	0.90	0.96	2.30	2.38	3.25	3.36
12	Item12	0.16	2.51	2.58	5.57	6.08	14.91	15.12	17.25	17.34
13	Item13	0.12	8.49	8.71	14.29	15.16	35.31	38.42	49.02	NA
14	Item14	0.11	4.49	4.52	6.93	8.89	21.93	27.59	36.54	39.54
15	Item15	-0.25	-0.54	-0.63	-2.69	-3.09	-6.95	-7.31	-8.84	NA

The adopted NECO Mathematics constructed-response test items were scored over 8 with 9 categories including 0, 1, 2, 3, 4, 5, 6, 7 and 8. It was observed from the table that columns 3 and 4 upward gave the discrimination (a) index with values ranging from -0.25 to 1.45 and category boundaries (b1, b2…, b8) of polytomous item response theory model obtained from item calibration of r 3.6.2 software. It was observed that IRT frameworks did not give the estimates of all the item parameters of the NECO Mathematics CR subjected to the item analysis process. The results suggested that within the IRT framework, all the items were analysed and the items fitted the generalized partial credit model. Similarly, the reliability coefficients of both WAEC and NECO were obtained using IRT and CTT. These values are presented in table 4.20 as follows:

Reliability Coefficients of WAEC and NECO Mathematics Constructed-response Items

Test Models	WAEC	NECO
CTT	0.47	0.59
GPC	0.70	0.70
GR	0.72	0.71

APPENDIX XVIII

Conditional Differences of WAEC Items: Intervals of size 7.8

Lower **Upper**	**12** **19.8**	**19.8** **27.6**	**27.6** **35.4**	**35.4** **43.2**	**43.2** **51**	**51** **58.8**	**58.8** **66.6**	**66.6** **74.4**	**74.4** **82.2**	**82.2** **90.1**
Item 1	-0.77	-0.77	0.43	0.21	0.58	-0.6	-0.41	0.22	0.5	0.67
Item 2	-0.41	-0.74	-0.48	-0.43	-0.04	-0.68	-0.46	-0.07	1.2	1.56
Item 3	-0.21	0.1	-0.25	0.06	-0.11	0.34	-0.36	1.28	0.1	-1.22
Item 4	0.2	-0.53	-0.94	-0.23	-0.49	0.19	0.42	0.44	2.1	1.56
Item 5	-0.07	0.16	0.56	0.48	0.22	-0.33	0.5	0.77	0.35	1.33
Item 6	-1.00	2.8	-0.16	0.85	0.2	-1.14	0.8	1.00	1.62	2.89
Item 7	-0.38	-0.87	0.55	-0.02	0.12	-0.13	0.27	-0.11	1.35	-1.11
Item 8	0.57	-2.83	0.75	0.02	-0.21	0.59	-0.49	0.11	0.55	-0.22
Item 9	0.00	-0.11	0.35	0.21	-0.31	-0.54	0.84	-0.44	-3.3	-1.56
Item 10	0.00	0.18	0.00	0.14	-0.04	0.07	1.36	-1.07	-3.8	0.89
Item 11	0.00	-0.21	0.27	-0.23	0.26	-1.45	0.77	1.66	0.9	1.56
Item 12	0.71	0.54	-0.09	-0.1	0.6	-0.59	-1.53	-2.22	0.9	-1.44
Item 13	1.14	-0.49	-0.13	-0.66	-0.17	1.17	-0.43	-1.71	-2.92	-4.78
Item 14	0.21	1.35	-0.3	0.26	-0.35	1.25	-1.34	-0.84	-3.2	-4.11
Item 15	0.18	1.78	-0.84	-0.63	-0.48	1.9	-0.97	0.47	0.4	1.78

Conditional Differences of NECO Items: Intervals of size 8.4

Lower	**9**	**17.4**	**25.8**	**34.2**	**42.6**	**51**	**59.4**	**67.8**	**76.2**	**84.6**
Upper	**17.4**	25.8	34.2	42.6	51	59.4	67.8	76.2	84.6	93.1
Item 1	0.81	0.57	0.02	0.11	-0.05	-0.75	1.75	-0.38	0.14	-0.50
Item 2	-0.49	0.01	-0.12	0.19	-0.21	0.24	0.09	0.33	0.41	0.20
Item 3	-0.55	0.44	-0.04	0.2	-0.16	-0.67	-1.41	-0.23	-0.73	0.60
Item 4	-0.06	-0.38	0.32	0.41	-0.37	-0.36	-0.56	-0.53	-1.33	1.80
Item 5	-0.09	-0.11	0.04	-0.05	0.73	1.54	0.96	-0.45	-0.38	-2.00
Item 6	0.04	-0.34	0.12	-0.09	0.17	-0.02	-1.16	-0.92	-0.02	0.60
Item 7	-0.89	0.98	-0.24	-0.61	0.1	0.16	-0.21	-0.36	0.27	-0.20
Item 8	0.00	-0.27	-0.23	-0.27	-0.33	0.15	-0.69	0.11	0.43	1.20
Item 9	0.06	-0.53	0.11	0.39	0.00	-0.36	-1.34	0.02	0.92	1.00
Item 10	-0.06	-0.20	0.24	-0.26	0.06	0.80	-1.01	-1.09	0.83	-0.40
Item 11	-0.08	-0.13	0.21	0.74	0.21	-0.90	0.94	0.2	-0.38	-0.60
Item 12	-0.01	0.03	0.11	-0.39	-0.39	-0.29	0.74	0.43	-0.16	-1.20
Item 13	0.48	-0.22	0.04	0.11	0.18	0.25	-0.18	0.74	-0.25	0.10
Item 14	-0.35	-0.51	-0.1	0.14	0.00	0.62	1.32	1.7	-0.44	-2.10
Item 15	0.01	-0.06	0.07	-0.26	0.05	-0.13	0.93	0.24	0.89	6.40