International Journal of Research
International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 4, April 2015
Available at http://internationaljournalofresearch.org

Introduction of Item Response Theory (IRT) Models in the Development and Validation of College Mathematic in Attaining Quality Education for National Values

ALIYU, R. TAIWO.
(08068561852)
taiwo_aliyu@yahoo.com
Department of Guidance and Counselling (Measurement and Evaluation Units)
Faculty of Education, Delta State University, Abraka

ABSTRACT

Test developments that measure the facets of learning with precision and accuracy in education have been a major task in academic setting in Nigeria. The scales used to measure performance rely on classical test theory (CTT) approach. This has been faulted since it yields different results when different samples of subjects from a population or different samples of items from the same pool are selected. Due to this shortcoming, the researcher employed the modern method of measurement of IRT to construct and validate college Mathematics. This will lead to attaining quality education in Nigeria. This paper is an empirical study. It is an instrumental research which adopted a multistage sampling technique in selecting the samples from the two states involved. The instrument consists of 50 items with a reliability value of 0.83. Chi-square goodness of fit test of MULTILOG was used to investigate how well the college Mathematics fit the 3-parameter logistic model of IRT while factor analysis was used to establish the unidimensionality of the items. Thirty-five (35) items scaled through the 3-parameter logistic model of IRT and were confirmed to measure the same construct (unidimensionality) while Chi-square established the fitness of the items. Recommendations are made in this paper why to employ IRT principles of test construction in Nigeria.

INTRODUCTION

Naturally man is interested in knowing the extent his set objectives have been achieved in a particular content. If the aim of education is to effect changes in the individual then there must be a kind of assessment. In education, individuals come to school to learn, it is therefore necessary to “ask questions” in relation to the extent to which the learner has been able to attain the objectives. This brings into focus the concept test, its construction and validation.

In the teaching-learning process, testing and evaluation of the students progress is a common place event. In fact, evaluation of the students’ progress is a major part of the teachers’ task. The question or task is how do we evaluate the students learning progress? One of such way of going about this is the use of specially prepared test, which measures all aspect of the identified instructional objectives and content domain.

Test Developers are basically concerned about the quality of test items and how examinees respond to it when constructing tests. Nenty (2004) said in educational practice, one of the principal tasks is the development of tests that measure the facets of learning with the greatest precision and accuracy. Psychometricians generally determine the validity and reliability of such tests. Experts have suggested that for measurement to be meaningful, the object to be measured, the numerals that will be assigned and the rules of assigning of numerals must be well defined. Since measurement is the assignment of numerals to objects, or event according to a stipulated rules (Adedoyin,2010). Therefore, in order to measure the performance of students in an academic setting, measurement instruments called tests are administered. Yoloye (2004) defined a test as a systematic procedure in which individuals are presented with a set of constructed stimuli to which they respond. The responses enable the tester to assign the testees numerals or sets of numerals from which inferences can be made about the testees performance.

There are different types of tests such as oral test, achievement test, performance test, psychological test, teacher-made test, standardize test, formative and summative tests (Aliyu, 2012). However, the seminar paper focuses on achievement test as part of test construction process analysis of items.

Psychometric theory offers two approaches or methods in analyzing test data: classical test theory (CTT) and item response theory (IRT). Both theories enable to predict outcomes of psychological tests by identifying parameters of item difficult and ability of test takers. Both are concerned to improve measures of validity and reliability. There are some identified issues in the CTT that concerns with calibration of item difficulty, sample dependence of coefficient measure and estimates of measurement error which in turn is addressed by the item response theory. This purpose of this paper is to construct and validate mathematics achievement test with IRT model.

STATEMENT OF PROBLEM

The poor performance of students in mathematics in our colleges has been of great concern in the education industry. WAEC (2009) and NECO (2010) have responded to the declining trend in student's performance by advocating for the new approach of analyzing their test data. The commonest method of assessment of students' performance has always been the classical test theory (CTT) which lack objectivity in the cognitive and psychomotor traits of the candidates. Therefore, the statement of problem if put in a question form is: How suitable is the development and validation of a CMAT (that would bring about firmness in the measurement of the achievement of examinees) able to determine students achievement in College Mathematics Achievement Test (CMAT) with item response theory model?

RESEARCH QUESTIONS

This study therefore, attempts to answer the following questions.

1. How does the first assumption of IRT of CMAT establish?
2. What are the estimates of CMAT item parameters using the three-parameter logistic model of IRT?

HYPOTHESIS

In the light of the above stated research questions one null hypothesis is formulated.

1. There is no significant fit of CMAT items to the three parameter logistic model of IRT.

PURPOSE OF THE STUDY

The main purpose of the study is to:

1. Develop and validate college Mathematics Achievement Test with item response theory model
2 Estimate the 3-parameter logistic models of CMAT items
3. Investigate if there is a significant fit of CMAT items to the 3 – parameter logistic model of IRT.

SIGNIFICANCE OF THE STUDY

The study intends to make college lecturers to see that the test that measure achievement in college mathematics is criterion- referenced so that test scores directly conveyed level of competence in defined mathematics domain. The study equally intends to make the presence or absence of item bias determine by test developers.

International Journal of Research
International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 4, April 2015
Available at http://internationaljournalofresearch.org

DELIMITATION AND SCOPE

This study is delimited to only 2 colleges of education in Oyo and Delta state, one federal and the other state. It also focuses on the development and validation of CMAT with item response theory model.

REVIEW OF RELATED LITERATURE

A mathematics achievement test was developed in an effort to overcome the dismail performance of teacher-education graduates in the mathematics of colleges of Education in Nigeria NCE. This assessment is in line with the objectives of mathematics curriculum of both the state and federal colleges of education in Nigeria.

Measurement is central to the construction of quality student assessment, even in the case of teacher-made or classroom designed or non standardized assessment. Measuring variables is one of the necessary steps in the research process (Osadebe & Aliyu, 2012). Romel (2009) said what follows are the statistical tool to analyze the data. Thus , Bond & Fox (2001) are of the opinion that the interpretation of data analyses can only be as good as the quality of measures. Although many test and measurement textbooks present classical test theory as the only way to determine the quality of an assessment. The IRT offers a sound alternative to the classical approach. Romel (2009) said that CTT is rooted in a process of dependability rather than measurement. It does not rely on item difficulty variable for precision and calibration or on total score for indicating the measure ability (Thissen, 2001).

The IRT is a general statistical theory about examinee item and test performance and how performance relates to the abilities that are measured by the items in the test. item responses can be discrete of continuous and can be dichotomously or polychotomously scored; item score categories can be ordered or unordered (Hambleton, & Jones, 2009). IRT is item centred in their fundamental estimation of person ability (Adedoyin, Nenty & Chilisa, 2008).

Latent trait models in test construction are utilized for purpose of constructing equivalent test forms, developing tests that discriminate between ability levels and improving customized test system. IRT can be used to investigate item bias (Romel, 2009). A set of items is considered unbiased if all subpopulations are equally affected by the same sources of variance, thus producing similar item characteristic curves (ICCS) for both groups (Okpong, 2006). If a test item has different connotative meanings for different groups, then examinees performance on that item may be subject to sources of variation that are unrelated to ability level. This refers to differential item function (DIF) and can cause item bias (Romel, 2009). Also, a set of items is considered unbiased if a source of irrelevant variance does not give an unfair advantage to one group over another (Odili, 2012). DIF detection procedures can investigate the effects achievement tests have on different subpopulation (Romel, 2009).

Item response theory (IRT) is a collection of statistical models and methods used for two broad purposes in the measurement of achievement outcomes: item analysis and scale scoring. The family of IRT models (Dichotomous and Polytomous models) describes, in probabilistic terms, the relationship between a person's response to a question (a scale) and his or her standing on the construct (e.g., achievement) being measured by the scale. Specifically, IRT models predict the probability of choosing each response category as a function of an underlying, unobserved trait and item parameters (Hambleton, Swaminathan, & Rogers, 1991). For item analysis, the IRT model characterizes each scale item with a set of properties that describes its ability to discriminate among individuals at different levels along a trait continuum. For scale scoring, IRT uses the full information from a person's responses to each

e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 4, April 2015
Available at http://internationaljournalofresearch.org

item to estimate their standing on the measured construct. Scale scoring using IRT estimates a score along the continuum of the construct being measured for persons who provide a particular sequence of item responses. Usually a person's score estimates include a measure of central tendency and a description of variability that is reported as a standard error of measurement. The IRT scale score may be computed using only the item parameters and the responses of a single individual to any arbitrarily selected set of items, and this is the basis for computer adaptive testing. For items such as multiple choice items, the parameter c_1 is used in attempt to account for the effects of guessing on the probability of a correct response. It indicates the probability that very low ability testees will get this item correct by chance, mathematically represented as a lower asymptote. A four-option multiple choice item might have an IRF like the example item; there is a ¼ chance of an extremely low ability candidate guessing the correct answer, so the c_1 would be approximately 0.25. This approach assumes that all options are equally plausible, because if one option made no sense, even the lowest ability person would be able to discard it, so IRT parameter estimate methods take this into account and estimate a c_1 based on the observed data.

A reasonable assumption is that each examinee responding to a test item possesses some amount of the underlying ability. Thus, one can consider each examinee to have a numerical value, a score, which places him or her somewhere on the ability scale. This ability score will be denoted by the Greek letter theta, θ. At each ability level, there will be a certain probability that an examinee with that ability will give a correct answer to the item. This probability will be denoted by P (θ). In the case of a typical test item, this probability will be small for examinees of low ability and large for examinees of high ability. If one plotted P (θ) as a function of ability, the result would be a smooth S-shaped curve such as shown in Figure 1.1 The probability of correct response is near zero at the lowest levels of ability. It increases until at the highest levels of ability, the probability of correct response approaches 1. This S-shaped curve describes the relationship between the probability of correct response to an item and the ability scale. In item response theory, it is known as the item characteristic curve. Each item in a test will have its own item characteristic curve (Baker, 2001).

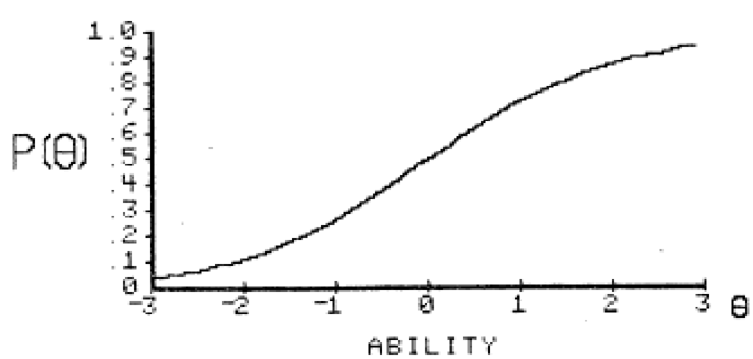


FIGURE 1-1. A typical item characteristic curve

Figure 1

The item characteristic curve is the basic building block of item response theory; all the other constructs of the theory depend upon this curve. There are two technical properties of an item characteristic curve that are used to describe it. The first is the difficulty of the item. Under item response theory, the difficulty of an item describes where the item functions along the ability scale. The second technical property is discrimination, which describes how well an item can differentiate between examinees having abilities below the item location and those having abilities above the item location. This property essentially reflects the steepness of the item characteristic curve in its middle section. The steeper the curve, the better the item can discriminate. The flatter the curve, the less the item is able to discriminate since the probability of correct response at low ability levels is nearly the same as it is at high ability levels. These two properties simply describe the form of the item characteristic curve (Baker, 2001).

IRT models come in many varieties, more than 100, (van der Linden and Hambleton, 1997) and can handle uni-dimensional as well as multidimensional data, binary (dichotomous) and polytomous response data, and ordered as well as unordered response data. The most commonly applied IRT models in achievement testing or measurement are the unidimensionality & multidimensional parametric families of dichotomous -response models, which are the Rasch Model - 1-parameter logistic (1-PL) function, The 2-parameter (2-PL) IRT model, and The 3-parameter (3-PL) model.

Models of Item Response Theory

There are three mathematical models commonly used in achievement measurement (one parameter or Rach Model; the two parameter model and the three parameter model) for the item characteristic curve. These models provide a mathematical equation for the relation of the probability of correct response to ability. Each model employs one or more parameters whose numerical values define a particular item characteristic curve. Such mathematical models are needed if one is to develop a measurement theory that can be rigorously defined and is amenable to further growth. In addition, these models and their parameters provide a vehicle for communicating information about an item's technical properties. The mathematical functions for the models are:

The Rasch, or One-Parameter, Logistic Model

This model was first published by the Danish mathematician Georg Rasch in the 1960s. The one-parameter logistic model, also known as the Rasch model (Aliyu 2015), assumes that there is no guessing parameter, i.e., $c_i = 0$ and that the discrimination parameter equals one, i.e., a = 1. The Rasch model specifies a 1-parameter logistic (1-PL) function. The equation for the Rasch model is given by the following:

$$P(\theta) = \frac{1}{1 + e^{-1.7(\theta - bi)}} \text{……… Rasch's Model}$$

Where: b is the difficulty parameter and θ is the ability level.

(Rasch, 1960, 1966).

The Logistic Function or Two Parameter Model

The two-parameter logistic model allows for different discrimination parameters per item and assumes that $c_i = 0$. The 2-parameter (2-PL) IRT model extends the 1-PL Rasch model by estimating an item discrimination parameter (α) and an item difficulty parameter. The discrimination parameter is similar to an item-total correlation and typically ranges from -0.5 to 2. An important feature of the 2-PL model is that the distance between an individual's trait level and item difficulty has a greater effect on the probability of endorsing highly discriminating items than on less discriminating items. Thus, more discriminating items provide greater information about a respondent than do less discriminating items. Unlike the Rasch model, discrimination needs to be incorporated, and the raw score is not sufficient for estimating trait scores. The equation for the two-parameter logistic model is given in the equation below.

$$P(\theta) = \frac{1}{1 + e^{-1.7a(\theta - bi)}} \ldots\ldots$$

Birnbaum's Model

where: e is the constant 2.718

b is the difficulty parameter

a is the discrimination parameter

L = a (θ - b) is the logistic deviate (logit) and

θ is an ability level.

The Three-Parameter Model

One of the facts of life in testing is that examinees will get items correct by guessing. Thus, the probability of correct response includes a small component that is due to guessing. Neither of the two previous item characteristic curve models took the guessing phenomenon into consideration. The 3-parameter (3-PL) model includes a pseudo-guessing parameter (c), as well as item discrimination and difficulty parameters. This additional parameter adjusts for the impact of chance on observed scores. The equation for the three-parameter model is:

$$P(\theta) = \frac{c_1 + [1 - c_1]}{1 + e^{-1.7ai(\theta - bi)}} \ldots\ldots..$$

Lord's Model

Where:

c is the guessing parameter and

θ is the ability level

IRT models are used as a basis for statistical estimation of parameters that represent the 'locations' of persons and items on a latent continuum or, more correctly, the magnitude of the latent trait attributable to the persons and items. The term latent is used to emphasize that discrete item responses are taken to be observable manifestations of the trait or attribute, the existence of which is hypothesized and must be inferred from the manifest responses. The other major body of psychometric theory of relevance to IRT is classical test theory. For tasks that can be accomplished using classical test theory, IRT generally brings greater flexibility and provides more sophisticated information. Some applications, such as computerized adaptive testing are enabled by IRT and cannot reasonably be performed using only classical test theory (Aliyu, 2015; Baker, 2001; Embretson and Reise, 2000)

RESEARCH METHOD

This study was designed to be an instrumental research, which is non-experimental since it involved the construction and validation of an instrument (CMAT). This was used in measuring students' performances.

POPULATION

The population of the study was made up of all State and Federal colleges of education in South-West and south-south geopolitical zone.

SAMPLE AND SAMPLING TECHNIQUES

Two colleges were used through random selection from South-West and south-south geopolitical zone. The colleges consist of one state and one federal. A total of 1000 students were used for the final administration of the instrument. This was done through multistage random sampling technique. At each state level, 2 out of the local government areas were randomly selected. At each local G.A 35% of the total number of schools were selected from each of the LGA while simple randomly sampling technique was used to select the students in order to arrive at a sample proportion of 1000 needed for the study.

VALIDITY OF THE INSTRUMENT

The instrument consists of two sections. Section A consists of the students' bio-data while section B consists of 75 items of CMAT which was developed by the researcher. The items were drawn from the NCE curricula for Mathematics by the Federal Ministry of Education (FME) using table of specification. It was equally verified by researcher's supervisor and two college Mathematics lecturers and two graduate Mathematics teachers who are members of Mathematics Association of Nigeria (MAN) at both Oyo and Delta state chapters. This was necessary to ensure both face and content validity. Some items were deleted while some were reconstructed which lead to the emergence of Fifty (50) items from the vetting exercise and were trial tested. They were administered to 60 students (20 boys and 40 girls) who were not part of the sample used.

RELIABILITY OF THE INSTRUMENT

A KR_{2O} reliability method was employed in testing the reliability coefficient of the instrument. The value obtained was 0.83. On the basis of the calculated reliability coefficient, the instrument was considered reliable for the study and administered to the 1000 samples.

DATA COLLECTION AND ANALYSIS

College lecturers assisted in the administration and collection of the data. The following statistical analyses were employed to prove the first assumption of the latent trait theory (unidimensionality), test the stated hypothesis and answered the research questions:

Factor Analysis using the principle component analysis (PCA) and rotated component matrix (RCM) was used to answer RQ1. This was done in order to establish the unidimensionality of CMAT i.e the extent to which CMAT measured a single trait. Chi-square goodness of fit of the MULTILOG was used to test the stated hypothesis. The 3-parameter logistic model of the item response theory (IRT) was used in analyzing the data collected for RQ2. The analysis used the regression technique of the latent trait theory (3-PL model) which involved the regression of each item on the latent ability. This brought about the estimation of the difficulty, discrimination and guessing indices of each of the items. The IRT estimation procedures using regression technique of the MULTILOG software was used in answering RQ2.

The following criteria were adopted in the selection of the items for inclusion in the final version of the CMAT: The items whose difficulty index ranges from -1 to 1.0, discrimination index ranges from 0.15 to 3.00 and the vulnerability to guessing did exceed 0.35 were all selected (Opasina, 2009).

DATA ANALYSIS AND PRESENTATION OF RESULTS

The purpose of the study was to develop and validate College Mathematics Achievement Test (CMAT) using the 3-parameter logistic model of the item response theory. The results obtained in the study are analyzed, presented and discussed in this stage.

Research Question RQ1: *How does the first assumption of IRT of CMAT establish?*

The first assumption of the item response theory of unidimensionality, which is the single trait for ability of CMAT was established. The process involved the use of the principle component analysis followed by factor rotation of the items.

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	36.690	64.368	64.368	36.690	64.368	64.368	35.631	62.511	62.511
2	8.626	15.134	79.502	8.626	15.134	79.502	8.041	14.107	76.618
3	3.080	5.404	84.906	3.080	5.404	84.906	3.142	5.512	82.131
4	2.788	4.891	89.797	2.788	4.891	89.797	2.373	4.163	86.294
5	1.448	2.541	92.338	1.448	2.541	92.338	2.056	3.606	89.900
6	1.408	2.470	94.808	1.408	2.470	94.808	1.992	3.495	93.395
7	1.067	1.872	96.681	1.067	1.872	96.681	1.872	3.285	96.681
8	.632	1.110	97.790						
9	.504	.884	98.675						
10	.368	.645	99.320						
11	.183	.320	99.640						
12	.115	.201	99.841						
13	.057	.101	99.942						
14	.033	.058	100.000						
15	9.723E-16	1.706E-15	100.000						
16	3.644E-16	6.393E-16	100.000						
17	8.649E-17	1.517E-16	100.000						
18	7.632E-17	1.339E-16	100.000						
19	5.695E-17	9.991E-17	100.000						
20	4.693E-17	8.234E-17	100.000						
21	3.568E-17	6.260E-17	100.000						
22	2.149E-17	3.771E-17	100.000						
23	1.841E-17	3.230E-17	100.000						
24	1.153E-17	2.023E-17	100.000						
25	7.989E-18	1.402E-17	100.000						
26	5.592E-18	9.810E-18	100.000						
27	3.195E-18	5.606E-18	100.000						
28	1.743E-18	3.057E-18	100.000						
29	8.149E-19	1.430E-18	100.000						
30	1.552E-20	2.722E-20	100.000						
31	1.328E-32	2.329E-32	100.000						
32	1.508E-33	2.646E-33	100.000						
33	2.218E-34	3.892E-34	100.000						
34	3.444E-35	6.042E-35	100.000						
35	3.332E-36	5.846E-36	100.000						
36	-1.327E-51	-2.327E-51	100.000						
37	-2.701E-34	-4.739E-34	100.000						
38	-7.934E-34	-1.392E-33	100.000						
39	-1.192E-33	-2.092E-33	100.000						
40	-3.236E-33	-5.677E-33	100.000						

41	-4.226E-33	-7.414E-33	100.000						
42	-2.760E-20	-4.843E-20	100.000						
43	-3.064E-19	-5.375E-19	100.000						
44	-4.984E-19	-8.744E-19	100.000						
45	-1.574E-18	-2.762E-18	100.000						
46	-3.653E-18	-6.409E-18	100.000						
47	-5.005E-18	-8.781E-18	100.000						
48	-9.103E-18	-1.597E-17	100.000						
49	-1.139E-17	-1.998E-17	100.000						
50	-1.595E-17	-2.798E-17	100.000						

DISCUSSION

As shown in the table above, there are seven 7 components or factor structure obtained from PCA. The Eigen Value of above 1 was used to select the factors or component into the CMAT instrument. The table revealed the initial eigenvalue of each component and the extraction sums of squared loadings. The 7 components were further subjected to rotated component matrix for the selection of CMAT items which show construct validity of the instrument. 40 items emerged under the 7structured rotated factor matrix, the rotation equally converged in 5 iterations using Varimax with Kaiser Normalization rotation method.

Research Question 2: *What are the estimates of CMAT items parameters using the 3-parameter logistics model of IRT?*

Items No:	Normal Deviate	Difficulty index (b)	Discrimination index (a)	Guessing index (c)
1	0.467	-0.27	0.53	0.01
2	0.486	-0.24	0.83	0.03
3	0.499	0.00	0.42	0.05
4	0.490	0.28	2.07	0.32
5	0.486	3.35	0.19	0.07
6	0.486	0.34	0,79	0.06
7	0.490	0.51	0.97	0.30
8	0.470	0.37	1.08	0.19
9	0.486	0.86	1.89	0.20
10	0.490	1.00	0.80	0.23
11	0.494	0.38	1.33	0.00
12	0.499	-0.54	0.65	0.01
13	0.490	0.19	0.21	0.00
14	0.494	0.84	0.09	0.00
15	0.503	-0.98	0.99	0.23
16	0.490	0.85	1.14	0.27

17	0.497	1.34	0.29	0.20
18	0.486	-0.08	0.52	0.00
19	0.486	0.43	0.19	0.00
20	0.494	-0.03	1.35	0.00
21	0.502	0.03	2.99	0.00
22	0.490	1.00	0.18	0.00
23	0.481	0.19	0.46	0.01
24	0.481	0.04	1.27	0.00
25	0.490	-0.18	0.12	0.30
26	0.490	0.40	0.99	0.31
27	0.497	-0.06	1.11	0.07
28	0.499	0.45	2.33	0.09
29	0.497	-0.97	0.15	0.01
30	0.494	-0.31	0.62	0.02
31	0.497	0.04	0.51	0.00
32	0.486	0.03	0.51	0.05
33	0.486	-0.24	0.96	0.00
34	0.499	0.33	0.47	0.08
35	0.490	-0.72	0.35	0.25

DISCUSSION

In order to carry out this analysis, MULTILOG statistical software tool was used. The items were analysed using the IRT 3-logistic parameter of MULTILOG. The table above shows the estimated item parameters; the difficulty, discrimination and the guessing indices of the 35-CMAT items that measure up to the required standard. The difficulty index, b, of the CMAT items ranges from -0.98 to 1.00. In this case, the more difficult items have positive values, the easier items have negative values. Thus, it can be inferred that item 10 and 22 (difficulty index of 1.00) are the most difficult while item 15 (difficulty index of -0.98) is the least difficult. According to IRT from scientific software international - SSI (2003), items with difficulty range of -1.0 to +1.0 could be considered valid items as they are neither too difficulty nor too easy.

The discrimination index, a, ranges from 0.19 to 2.99. All the items have discrimination index in the positive direction which is an indication that all of them are satisfactory in the context of discrimination index. Therefore, it could be inferred that all the items are valid under item discrimination.

The third parameter of the 3-logistics model of IRT called the pseudo-guessing index, c, ranges from 0.00 to 0.32. Metra et aal (2003) cited in Opasina (2009) view valid items under guessing index to be those not above 0.35. This is an indication that the probability of vulnerability to guessing is not high which makes the items to be satisfactory within the context of a valid test. Therefore, for an item to be considered satisfactory under guessing index, the vulnerability to guessing must not be greater than 0.35. In this case, item 4 and 7 (c= 0.32 and 0.30) represented the highest probability due to chance for a correct response while items 11, 13, 14, 18, 19, 20, 21, 22, 24, 31 and 33 (c= 0.00) represented the least probability that a correct response occurred by chance.

Hypothesis 1: *There is no significant fit of CMAT items to the 3-parameter logistic model of IRT*

Items No:	OBSERVE PROPORTION OF TESTEES	EXPECTED PROPORTION OF TESTEES	CHI-SQUARE X^2	DEGREE OF FREEDOM Df

1	0.490	0.551	0.00677	1
2	0.4850	0.5618	0.01055	1
3	0.447	0.500	0.00562	1
4	0.345	0.326	0.00111	1
5	0.237	0.258	0.00171	1
6	0.331	0.415	0.01700	1
7	0.401	0.517	0.02603	1
8	0.431	0.513	0.01311	1
9	0.287	0.384	0.02450	1
10	0.370	0.429	0.00811	1
11	0.267	0.380	0.03360	1
12	0.557	0.619	0.00621	1
13	0.295	0.321	0.00211	1
14	0.304	0.316	0.00046	1
15	0.277	0.253	0.00228	1
16	0.309	0.424	0.03119	1
17	0.313	0.349	0.00371	1
18	0.454	0.515	0.00723	1
19	0.561	0.534	0.00137	1
20	0.410	0.509	0.01926	1
21	0.353	0.482	0.03452	1
22	0.396	0.421	0.00148	1
23	0.591	0.533	0.00631	1
24	0.386	0.486	0.02058	1
25	0.508	0.491	0.00059	1
26	0.293	0.390	0.02413	1
27	0.426	0.517	0.01602	1
28	0.240	0.394	0.06019	1
29	0.429	0.414	0.00054	1
30	0.501	0.567	0.00768	1
31	0.569	0.508	0.00733	1
32	0.432	0.494	0.00778	1
33	0.485	0.566	0.01159	1
34	0.672	0.717	0.00282	1
35	0.44	0.401	0.00483	1

DISCUSSION

The Chi-square goodness-of-fit was used to investigate how the CMAT items fit the 3-parameter logistic model of the latent trait theory. The Chi-square goodness of fit index was less than the critical value (x^2 at df 1= 3.840, $p < 0.05$). This means that the items characteristics curve to be specified by the values of the item parameter estimates could be said to fit the data. The degree of freedom (df) of 1 was obtained from the dichotomous nature of the examinee's responses that is correct or incorrect response. This agrees with Kerlinger eta al (2000) cited in Opasina (2009). The items therefore could be said to fit the 3-parameter logistic model of IRT with MULTILOG.

Conclusions and Recommendations

In this paper, an attempt was made to construct and validate college mathematics with IRT models. IRT provides several means of ensuring the generation of valid scores that would support valid quality-of-education-related decision which will enables the potential of every member of society to be identified, developed and utilized for the development of the individual and the nation at large.

Conclusion

The estimates of CMAT items parameters using the 3-logistics models of the IRT indicates that the threshold, b, the slope, a, and the pseudo-guessing, c, indices were all chosen within the standard range recommended by IRT from scientific software international (SSI 2003)

Chi-square goodness-of-fit shows that there is a significant fit of CMAT items to the 3-parameter logistic models of IRT of MULTILOG

Factor analysis using the PCA and Rotated Component Matrix (RCM) indicates the unidimensionalitry of the CMAT items which is the first assumption of IRT.

Recommendation

This paper recommends that IRT models should be adopted in test construction and validation over CTT, since CTT has varieties of shortcomings. Also, high priority should be given to research in Mathematics since it is the bedrock of all other subjects. This will enhance the performance of the individual and the development of the national values. Finally, the researcher has observed that the multiple choice items (objectives) in Mathematics are not commonly set in most of our colleges in Nigeria since this shows a very high content validity of the set items in mathematics for our students using the IRT models. Other statistical software tools such as BILOG-MG3, PARSCALE, WINSTEPS etc can be adopted in analyzing data.

References

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

[2.] Adedoyin,O.O. (2009). Investigating the invariance of person parameter estimates based on CTT and IRT. *International Journal Educational science Research and Review vol. 2(2), 117-113*

[3.] Aliyu, R.T. (2015). *Construct Validity of Mathematics Test Items using the Rasch Model.An International Journal of Social Science and Humanities Research. 3(2), 22-28*

[4.] Aliyu, R.T. (2012). *Development and validation of Mathematics Achievement test using the Rasch model. An unpublished ph.d proposal thesis in Delta state university, Abraka Nigeria.*

[5.] Aliyu, R.T. & Ocheli, E.C. (2012). *Development and validation of College Mathematics test (CMAT) with Item Response Theory (IRT) models in attaining quality education for national value. A paper presented in the international conference held at Delta state University (DELSU) between 13-17, November, 2012.*

[6.] *Anigbo, L.C. (2010). Demonstration of the multiple matrices simply technique in establishing the psychometric characteristics of large samples. Journal of education and practice vol 2 (3).*

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

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

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

[10.] Nenty, H. J. (2005). The application of item response theory in strengthening assessment role in the implementation of national education policy.

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

International Journal of Research
International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 4, April 2015
Available at http://internationaljournalofresearch.org

[12.] Vanderlinden, W. I. & Hambleton, R.T.(1997). A latent trait look at the pretest post test validation criterion referenced test items. *Review of Educational Research*

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

COUNSELLING PSYCHOLOGY DEPARTMENT
FACULTY OF EDUCATION
ABRAKA CAMPUS Time: 1 hr FINAL DRAFT
COLLEGE MATHEMATICS ACHIEVEMENT TEST-(CMAT)

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

SECTION A: Personal Information
i. Name of your School:
ii. Gender: Male ( ) Female ( )
iii. Age in years
iv. Place of residence - Rural ( ) Urban ( )
v. School type – Federal ( ) State ( )

SECTION B: - MAT

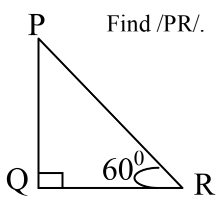
Attempt all the questions.

1. If (x - 3) is a factor of $2x^3 + 3x^2 - 17x - 30$, find the remaining factors A. (2x – 5)(x – 2) B. (2x – 5)(x + 4) C. (2x + 5)(x – 2) D. (2x + 5)(x + 2)
2. Simplify $\frac{^nP_4}{^nC_4}$ A. 24 B. 18 C. 12 D. 6
3. Find the coefficient of X^4 in the binomial expansion of $(1 - 2x)^6$ A. 320 B. 240 C. 250 D. 230
4. Evaluate $\log_{0.25} 8$ A. $\frac{3}{2}$ B. $\frac{2}{3}$ C. $-\frac{2}{3}$ D. $-\frac{3}{2}$
5. A particle is projected vertically upwards from a height 45metres above the ground with a velocity of 40m/s. How long does it take to hit the ground? [Take g = 10m/s] A. 1s B. 3s C. 7s D. 9s
6. A binary operation on the set of real numbers is defined by M * n = $\frac{Mn}{2}$ for all m, n ε R. If the identity element is 2, find the inverse of -5 A. $-\frac{4}{5}$ B. $-\frac{2}{5}$ C. 4 D. 5
7. Evaluate $\int_1^2 (x^2 - 4x)dx$ A. $\frac{11}{3}$ B. $\frac{3}{11}$ C. $-\frac{3}{11}$ D. $-\frac{11}{3}$
8. The locus of a point equidistant from the intersection of lines 3x – 7y + 7 = 0 and 4x – 6y + 1 = 0 is a A. line parallel to 7x + 13y + 8 = 0 B. circle C. semi circle D. bisector of the line 7x + 13y + 8 = 0
9. A stone is thrown vertically upwards and its height at any time t seconds is $h = 45t - 9t^2$. Find the maximum height reached A. 45.25m B. 45.50m C. 56m D. 56.25m
10. The initial velocity of an object is u = $\binom{-5}{3} m/s$. If the acceleration of the object is a = _ $\binom{3}{4} m/s^2$ and it moved for 3 seconds. Find the final velocity A. $\binom{-14}{15} m/s$ B. $\binom{-2}{1} m/s$ C. $\binom{4}{-9} m/s$ D. $\binom{-14}{-9} m/s$
11. The sum and product of the roots of a quadratic equation are $\frac{4}{7}$ and $\frac{5}{7}$ respectively. Find its equation A. $7x^2 - 4x - 5 = 0$ B. $7x^2 - 4x - 5 = 0$ C. $7x^2 + 4x - 5 = 0$ D. $7x^2 - 4x + 5 = 0$
12. Simplify $(\sqrt{6} + 2)^2 - (\sqrt{6} - 2)^2$ A. $2\sqrt{6}$ B. $4\sqrt{6}$ C. $8\sqrt{6}$ D. $16\sqrt{6}$
13. The nth term of a sequence is $n^2 - 6n - 4$. Find the sum of the 3rd and 4th terms A. 24 B. 23 C. -24 D. -25
14. Two bodies of mass 8kg and 5kg traveling in the same direction with speed x m/s and 2 m/s respectively collide. If after collision, they move together with a speed of 3.85m/s, find correct to the nearest whole number, the value of x A. 2 B. 5 C. 8 D. 13

International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 4, April 2015
Available at http://internationaljournalofresearch.org

15. If P = $\begin{pmatrix} 1 & -2 \\ 3 & 4 \end{pmatrix}$ and Q = $\begin{pmatrix} -2 & 3 \\ 1 & 0 \end{pmatrix}$, find PQ A. $\begin{pmatrix} 4 & 1 \\ -2 & 9 \end{pmatrix}$ B. $\begin{pmatrix} -4 & 1 \\ 2 & 9 \end{pmatrix}$ C. $\begin{pmatrix} -4 & 3 \\ -2 & 13 \end{pmatrix}$ D. $\begin{pmatrix} -4 & 3 \\ -2 & 9 \end{pmatrix}$
16. Evaluate $\lim_{x \mapsto 3} \left(\frac{x^2-2x-3}{x-3}\right)$ A. 4 B. 3 C. 2 D. 0
17. If $\begin{vmatrix} 4 & x \\ 5 & 3 \end{vmatrix}$ = 32, find the value of x A. 4 B. 2 C. -2 D. -4
18. Find the value of Cos ($60^0 + 45^0$) leaving your answer in surd form A. $\frac{\sqrt{6}+\sqrt{2}}{4}$ B. $\frac{\sqrt{3}+\sqrt{6}}{4}$ C. $\frac{\sqrt{2}-\sqrt{6}}{4}$ D. $\frac{\sqrt{3}-\sqrt{6}}{4}$
19. Find the equation of tangent to the curve $y = 4x^2 - 12x + 7$ at point (2, -1) A. y + 4x – 9 = 0 B. y – 4x – 9 = 0 C. y – 4x + 9 = 0 D. y + 4x + 9 = 0
20. The third term of geometric progression (G.P) is 10 and the sixth term is 80. Find the common ration A. 2 B. 3 C. 4 D. 8
21. Calculate, correct to one decimal place, the length of the line joining points x (3, 5) and y (5, 1) A. 4.0 B. 4.2 C. 4.5 D. 5.0
22. Evaluate ${}^{n+1}C_{n-2}$ if n = 15 A. 360 B. 3360 C. 1120 D. 560
23. In how many ways can the letters of the word TOTALITY be arranged? A. 6720 B. 6270 C. 6207 D. 6027
24. Evaluate $\int_0^{\frac{\pi}{4}} \sec^2 \theta d\theta$ A. 1 B. 2 C. 3 D. 4
25. Simplify $\frac{27^{\frac{1}{3}}}{16^{\frac{-1}{4}}}$ A. 6 B. 5 C. 4 D. 3
26. In PQR, <PQR is a right angle, /QR/ = 2cm and < PRQ = 60^0.

Find /PR/. A. $4\sqrt{3}cm$ B. $4cm$ C. $2\sqrt{3}cm$ D. $5\sqrt{3}cm$



27. For what values of x is the expression $\frac{x-5}{x^2-2x-3}$ not defined? A. 3, 1 B. -1, -3 C. -1, 3 D. 3, -2
28. The sides of a right –angles triangle in ascending order of magnitude are 8cm, (x – 2)cm and x cm. Find x A. 16 B. 17 C. 34 D. 90
29. If y = $\sqrt{ax-b}$ express x in terms of y, a and b A. x = $\frac{y^{2-b}}{a}$ B. x = $\frac{y+b}{a}$ C. x = $\frac{y-b}{a}$ D. x = $\frac{y^2+b}{a}$
30. If $k\sqrt{28} + \sqrt{63} - \sqrt{7} = 0$, find k A. -2 B. -1 C. 1 D. 2
31. Solve the inequality 2x + 3 < 5x A. x > 1 B. x < $\frac{3}{7}$ C. x > $\frac{3}{7}$ D. x <1
32. Solve the equation $3y^2 = 27y$ A. y = 0 or 3 B. y = 0 or 9 C. y = -3 or 3 D. y = 3 or 9
33. Given that $\log_4{}^x$ = -3, find the value of x A. $\frac{1}{81}$ B. $\frac{1}{64}$ C. 64 D. 81
34. The sum to infinity of a G.P is A. $\frac{a}{1-r}$ B. r^2 C. r - $\frac{1}{2}$ D. r – ½
35. Simplify $\frac{2x-1}{3} - \frac{x+3}{2}$ A. $\frac{x+7}{3}$ B. $\frac{x+8}{6}$ C. $\frac{x-11}{6}$ D. $\frac{x-4}{6}$