Nigerian Online Journal of Educational Sciences and Technology (NOJEST)

Volume 1, Number 1, 2020

NIGERIAN ONLINE JOURNAL OF EDUCATIONAL SCIENCES AND TECHNOLOGY (NOJEST)

NIGERIAN ONLINE JOURNAL OF EDUCATIONAL SCIENCES AND TECHNOLOGY

http://ujh.unilag.edu.ng

EXAMINING THE RELATIONSHIP BETWEEN THE RASCH AND 2-PL MODELS OF IRT IN SELECTING ITEMS IN A CONSTRUCTED TEST FOR CURRICULUM DEVELOPMENT

Aliyu R. T. & Akinoso S. O.
Department of Mathematics Education,
Faculty of Education, *Lead City University, Ibadan*
[2]Department of Science and Technology Education,
Faculty of Education, University of Lagos, Nigeria
sakinoso@unilag.edu.ng

To cite this article:
Aliyu R. T. & Akinoso S. O. (2020). Examining the relationship between the Rasch and 2-pl models of IRT in selecting items in a constructed test for curriculum development. *Nigerian Online Journal of Educational Sciences and Technology (NOJEST),* 1 (1), 77-98

EXAMINING THE RELATIONSHIP BETWEEN THE RASCH AND 2-PL MODELS OF IRT IN SELECTING ITEMS IN A CONSTRUCTED TEST FOR CURRICULUM DEVELOPMENT

Aliyu R. T.[1] & Akinoso S. O.[2]

Article Info

Article History

Received:
05 February 2020

Accepted:
20 April 2020

Keywords

Test-Developments, Item Response Theory (IRT), Rasch model, 2-PL model, Classical Test Theory (CTT)

Abstract

Test developments are critical in the measurement of learning and education attainments. Its precision and accuracy is a major task in an academic setting in Nigeria and beyond. The scales employed for achievement tests are often based on classical test theory (CTT) approach with a drawback of the variability of results in different samples of the same population or from the same pool of items in some Africa countries like Nigeria. This work explores the method of the Rasch model and the 2-PL model of Item Response Theory (IRT) to examine the relationship between the models used on a constructed Mathematics Aptitude Test (MAT) items. A 120 items instrument with a reliability value of 0.86 was developed by the researcher. Mean Square (MNSQ) and ZSTD of fitness of Winsteps and Two-Parameter Model (2-PL) of Bilog-Mg3 were used to investigate how well the Mathematics fit the Models. Eventually, thirty-three (33) items whose parameters are known scaled through the Rasch model and were confirmed to measure the same construct (uni-dimensionality) while 13 items are significant but do not fit into the 2-PL model at p< 0.05. Rasch shows that only 33 items fit into the model while the 2-PL model shows that 107 items fit into its model. This shows that a great disparity occurs between Rasch and 2-PL model. These items are banked for curriculum development purposes. This paper concludes that the Rasch model is preferred to the 2-PL model and therefore recommends its usage in curriculum development in Nigerian schools and Examination bodies in Africa and beyond.

Introduction

Introduction

Assessment is an essential component of learning and teaching, as it allows the quality of both teaching and learning to be judged and improved. It often determines the priorities of education, influences practices and affects learning in general. Changes in curricula and learning objectives are ineffective if assessment practices remain the same as learning and teaching tend to be modelled against the test. The aim of modern testing is not just to present a group of test-takers with a set of items but to administer items that are informative and challenging for each testee. The psychometric methods that allow the scores of test-takers attempting different sets of items to be compared directly are based either on the Rasch model (Odili, Osadebe, & Aliyu, 2015) or item response theory (IRT) models (Michaela eta al, 2013). The Rasch model postulates that the probability of a person giving a correct response to an item is governed only by the person's ability and the item's difficulty, both of which can be represented as locations on the same underlying measurement scale. A person's ability is estimated from that individual's response to a set of items with previously estimated difficulties.

One parameter model, also known as the Rasch Model, uses only a single parameter, namely item difficulty to estimate an unobservable trait of a particular examinee. The two-parameter and three-parameter models are also widely used, especially in large scale assessment (Downing, 2003 and Odili, Osadebe, & Aliyu, 2015). The two-parameter adds an item discrimination parameter to the item difficulty, whereas the three-parameter model adds a 'guessing' parameter to item difficulty and item discrimination.

According to Aliyu (2015), the choice of an appropriate model depends on the type of test questions and their scoring. Another important consideration is that, in practice, the choice of models depends on the amount of data available. The larger the number of the parameter is, the more data are needed for parameter estimation, thus requiring more complex calculation and interpretation. In this case, the Rasch Model has some special properties that make it attractive to users. Rasch Model involves fewest parameters; therefore, it is easier to work with (Aliyu, 2013). Wright (1990) gives a more influential explanation in favour of the Rasch Model compared to a three-parameter model. These two models are the opposite in philosophy and practice. The three-parameter model will adjust to adapt whatever type of data (includes invalid responses). The Rasch model, however, has tight standards in controlling the data. Unlike the three-parameter model, invalid responses such as guessing on an item will not be accepted. It is described as an unreliable person's reliability. Critics of the Rasch Model often regard the model as having strong assumptions that are difficult to meet. However, these are values that make the Rasch Model more appropriate in practice than the two and the three-parameter models.

In any mathematical model, it is important to assess the fit of data to the model. If item misfit with any model is diagnosed as due to poor item quality, for example confusing distractors in a multiple-choice test, then the items may be removed from that test form and rewritten or replaced in future test forms. If, however, a large number of misfitting items occur with no apparent reason for the misfit, the construct validity of the test will need to be reconsidered for curriculum development and the test specifications may need to be rewritten. Thus, misfit provides invaluable diagnostic tools for test developers, allowing the hypotheses upon which test specifications are based to be empirically tested against data. To this end, the researchers want to examine the relationship between the Rasch model and the 2PL model of IRT in selecting items in a constructed test for efficiency and effective assessment.

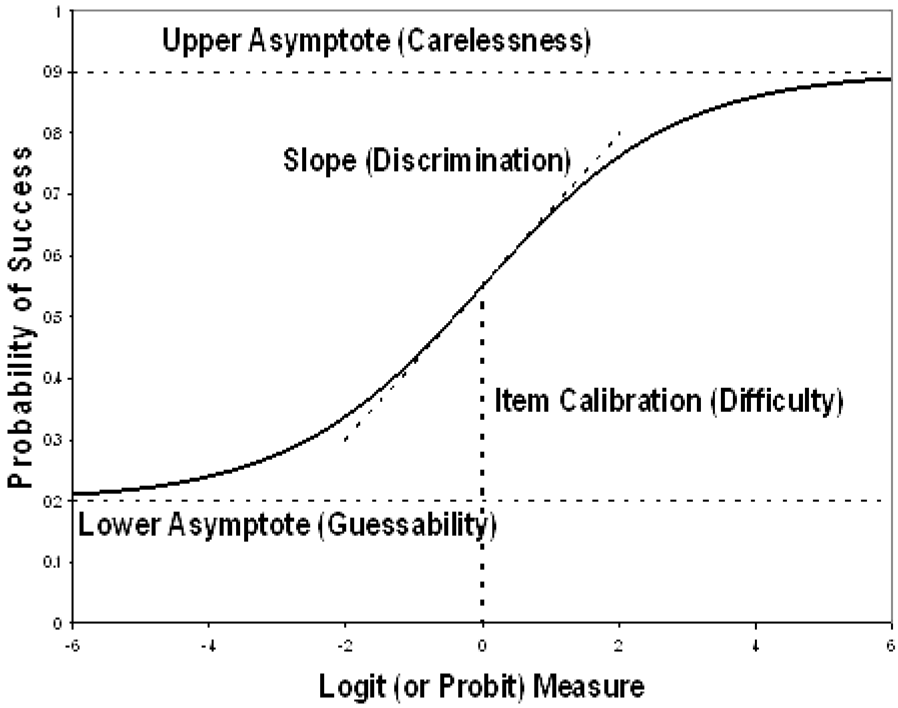
There are several methods of assessment for assessing fit for curriculum development purposes, such as a chi-square statistic, or a standardized version of it. Two and three-parameter IRT models adjust item discrimination, ensuring improved data-model fit, so fit statistics lack the confirmatory diagnostic value found in one-parameter models, where the idealized model is specified in advance. Data should not be removed based on fitting the model, but rather because a construct relevant reason for the misfit has been diagnosed, such as a non-native speaker of English taking a Mathematics test written in English. Such a candidate can be argued to not belong to the same population of persons depending on the dimensionality of the test, and, although one parameter IRT measures are argued to be sample-independent, they are not population independent, so misfit such as this is constructed relevant and does not invalidate the test or the model. Such an approach is an essential tool in instrument validation. In two and three-parameter models, where the psychometric model is adjusted to fit the data, future administrations of the test must be checked for fit to the same model used in the initial validation to confirm the hypothesis that scores from each administration generalize to other administrations. If a different model is specified for each administration to achieve a data-model fit, then a different latent trait is being measured and test scores cannot be argued to be comparable between administrations.

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

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

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

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



4-PL IRT Item Characteristic Curve

Upper and lower asymptotes are notoriously difficult to estimate, so it appears that Lord abandoned his 4-PL model, and the value of ci in the 3-PL model is, on occasion, imputed from the number of options in a multiple-choice item, instead of being estimated directly from the data. Even the estimation of item discrimination usually requires constraints, such as "a_i cannot be negative or too big." The dichotomous Rasch model, however, provides an opportunity to estimate the first approximation to these parameters. These estimates can be useful in diagnosing whether the behaviour they reflect could be distorting the Rasch measures. In the dichotomous Rasch model, $c_i = 0$, $d_i = 1$ and $a_i = 1$. We can, however, treat the Rasch values as starting values in a Newton-Raphson iterative processed intended to find the maximumlikelihood values of each of these parameters, in a context in which all other parameter values are known.

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

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

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

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

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

$$\hat{c}_i = \left[\frac{\sum_{n,x=1} e^{-(\theta_n - \delta_i)} - \sum_{n,x=0} 1}{\sum_{n,x=1} e^{-2(\theta_n - \delta_i)} + \sum_{n,x=0} 1}\right] \qquad \hat{a}_i = 1 + \left[\frac{\sum_n \left(M_{niX_{ni}} - \sum_{k=1}^{m} P_{nik} M_{nik}\right)}{\sum_n \left(\sum_{k=1}^{m} M_{nik}^2 P_{nik} - \left(\sum_{k=1}^{m} M_{nik} P_{nik}\right)^2\right)}\right]$$

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

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

Returning to the dichotomous model, the lower asymptote (guessability) is: where 0 <= c_i <= 1. The Rasch expectation of c_i is 0. The upper asymptote (carelessness) is: where 0 <= d_i <= 1. The Rasch expectation of d_i is 1. In practice, it is convenient to use only observations in the lower tail for estimating the lower asymptote, in the centre for estimating discrimination, and in the upper tail for estimating the upper asymptote (Adapted from Aliyu, 2015 work).

Data Requirements for Design and Analysis with the Rasch Model

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

Objective of the Study

The study examined the relationship between the Rasch and 2-PL model of IRT using the MAT items. The study

i. Found out the difficulty index of each item in the constructed Mathematics Aptitude Test (MAT) using the Rasch model
ii. Found the total number of items that fit into the Rasch model
iii. Determined the difficulty index of each item in the constructed Mathematics Aptitude Test (MAT) using the 2-PL model
iv. Find the total number of items that fit into the 2-PL model

Research Questions

The following research questions were used for this study.

Research Question 1: What are the difficulty index of each item in the constructed Mathematics Aptitude Test (MAT) items using the Rasch model?

Research Question 2: What is the total number of MAT items that fit into the Rasch model?

Research Question 3: What are the difficulty index of each item in the constructed Mathematics Aptitude Test (MAT) items using the 2-PL model of IRT?

Research Question 4: What is the total number of MAT items that fit into the 2-PL model of IRT?

Methodology

This study focuses on the relationship between the Rasch and 2-PL in a developed multiple-choice Mathematics Aptitude Test for curriculum development. The instrumentation research design was adopted. The population for this study consists of all senior secondary school two students (SSII) in Oyo State. The simple random sampling techniques of balloting were used for the selection of the ten (10) senior secondary schools. The sample size for the study was 600 respondents which were selected using a non-proportionate stratified random sampling technique from the selected schools at 60 testees each. The instrument used for this study is the Mathematics Aptitude Test (MAT) developed by the researcher contained 150 items. The test content consists of three components based on a well-designed Test Blue Print covering the six levels of the cognitive domain of learning. It consists of three components of aptitude test which include: Verbal Aptitude test with the highest number of fifty (50) items; Abstract Aptitude Test which contains fifteen (43) items and Numerical/Quantitative Aptitude Test with (57 items). This shows how the 150 test items in the MAT were distributed among the content

areas as well as the instructional objectives. This was done to address the content validity of the instrument. A total of 120 items that formed the MAT were drawn using the Classical Test Theory (CTT) procedure after the experimental try-out and revision of the test items. The difficulty and the discrimination indices found were used in selecting a total of one hundred and twenty test items. This was validated by two experts in the field for both content and face validities.
The reliability of the MAT was established with the use of Kuder-Richardson 20 (KR-20) on 50 testees who were not part of the sample used for the study. The calculated reliability coefficient was 0.86 which indicated that the test items could be administered to the targeted audience. The research questions were analyzed using Winsteps and BILOG-MG3 statistical software to determine the: difficult level of MAT using the Rasch and 2-PL models of IRT. In WINSTEPS, the measures are determined through iterative calibration of items using the MAT. Research questions 1 and 2 were answered using winsteps while research questions 3 and 4 were answered using Bilog-Mg3.

Results

The results obtained in this study are presented below. Winsteps 3.75 and Bilog-Mg3 were used to answer the research questions:

Research Question 1: What is the difficulty index of each item in the constructed Mathematics Aptitude Test (MAT) using the Rasch model?

Table 1: Difficulty indices of MAT using infit and outfit of MNSQ and ZSTD indices of Rasch

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
39	33	600	3.03	.18	1.02	.2	1.78	3.3	-.04	.12	94.5	94.5	I0039
33	42	600	2.77	.16	1.07	.6	2.02	4.7	-.16	.14	93.0	93.0	I0033
37	47	600	2.65	.15	1.08	.6	1.81	4.2	-.12	.15	92.2	92.1	I0037
31	54	600	2.49	.14	1.06	.6	1.80	4.6	-.10	.16	91.0	91.0	I0031
44	55	600	2.47	.14	1.10	.9	1.75	4.4	-.15	.16	90.8	90.8	I0044
43	56	600	2.45	.14	1.04	.4	1.63	3.8	-.04	.16	90.7	90.6	I0043
35	59	600	2.39	.14	1.07	.6	1.64	4.0	-.07	.16	90.2	90.1	I0035
42	90	600	1.89	.12	1.06	.8	1.33	3.1	.03	.20	85.0	85.0	I0042
38	91	600	1.88	.12	1.11	1.4	1.55	4.8	-.09	.20	84.8	84.8	I0038
47	92	600	1.87	.12	1.08	1.1	1.39	3.6	-.01	.20	84.7	84.6	I0047
34	100	600	1.76	.11	1.07	1.0	1.30	3.0	.03	.21	83.3	83.3	I0034
30	122	600	1.51	.10	1.13	2.2	1.38	4.4	-.05	.22	79.7	79.6	I0030
28	126	600	1.46	.10	1.17	2.9	1.40	4.7	-.10	.23	79.0	79.0	I0028
46	127	600	1.45	.10	1.19	3.2	1.53	6.1	-.17	.23	78.8	78.8	I0046
41	129	600	1.43	.10	1.18	3.2	1.52	6.1	-.15	.23	78.5	78.5	I0041
84	131	600	1.41	.10	1.21	3.7	1.57	6.7	-.20	.23	78.2	78.1	I0084
45	140	600	1.32	.10	1.19	3.6	1.46	5.9	-.14	.24	77.5	76.7	I0045
85	147	600	1.25	.10	1.23	4.5	1.52	7.0	-.20	.24	74.7	75.5	I0085
66	164	600	1.10	.09	1.14	3.1	1.25	4.1	.00	.25	71.2	72.8	I0066
89	174	600	1.01	.09	1.20	4.7	1.35	5.9	-.09	.26	69.3	71.2	I0089
99	177	600	.98	.09	1.22	5.3	1.38	6.6	-.13	.26	70.2	70.7	I0099
75	178	600	.97	.09	1.14	3.6	1.22	4.0	.01	.26	68.7	70.6	I0075
110	187	600	.90	.09	1.24	6.1	1.35	6.5	-.13	.26	65.2	69.3	I0110
10	188	600	.89	.09	1.24	6.2	1.35	6.6	-.14	.26	65.0	69.1	I0010
98	192	600	.86	.09	1.18	4.8	1.30	5.9	-.05	.26	68.0	68.5	I0098
36	199	600	.80	.09	1.12	3.5	1.19	4.0	.06	.27	69.8	67.6	I0036

27	201	600	.78	.09	1.02	.6	1.03	.8	.23	.27	67.8	67.4	I0027
91	208	600	.73	.09	1.23	6.8	1.31	6.8	-.10	.27	60.7	66.5	I0091
54	209	600	.72	.09	1.14	4.4	1.17	3.8	.05	.27	58.3	66.4	I0054
60	217	600	.66	.09	1.04	1.4	1.04	1.0	.21	.27	59.7	65.5	I0060
40	223	600	.61	.09	1.07	2.3	1.09	2.4	.17	.28	60.7	64.9	I0040
50	224	600	.60	.09	1.06	2.2	1.09	2.4	.17	.28	63.2	64.8	I0050
97	226	600	.59	.09	1.11	3.6	1.16	4.1	.10	.28	62.8	64.6	I0097
88	234	600	.52	.09	.95	-1.7	.94	-1.7	.35	.28	69.7	63.9	I0088
62	237	600	.50	.09	1.01	.4	1.03	1.0	.26	.28	66.8	63.7	I0062
3	243	600	.46	.09	.97	-1.1	.98	-.5	.32	.28	65.3	63.3	I0003
96	243	600	.46	.09	1.12	4.4	1.14	4.2	.09	.28	55.0	63.3	I0096
103	243	600	.46	.09	.97	-1.1	.98	-.5	.32	.28	65.3	63.3	I0103
119	247	600	.43	.09	.96	-1.5	.97	-1.0	.34	.28	67.8	63.0	I0119
19	248	600	.42	.09	.96	-1.4	.97	-.9	.34	.28	67.7	63.0	I0019
51	261	600	.32	.09	1.10	4.0	1.12	3.9	.13	.29	55.5	62.5	I0051
69	262	600	.32	.09	.95	-2.2	.94	-2.0	.37	.29	70.2	62.5	I0069
25	278	600	.20	.09	1.15	6.2	1.17	5.9	.05	.29	51.2	62.4	I0025
95	281	600	.18	.09	.93	-2.8	.93	-2.6	.39	.29	66.0	62.4	I0095
67	288	600	.12	.09	1.12	4.8	1.11	4.1	.12	.29	50.5	62.5	I0067
58	299	600	.04	.09	1.02	.7	1.02	.7	.27	.29	61.0	62.7	I0058
94	299	600	.04	.09	.91	-4.1	.90	-3.9	.44	.29	70.3	62.7	I0094
82	302	600	.02	.09	.91	-3.7	.91	-3.7	.43	.29	67.7	62.8	I0082
87	304	600	.01	.09	.92	-3.6	.91	-3.3	.42	.29	69.3	62.9	I0087
17	306	600	-.01	.09	.96	-1.8	.95	-1.8	.36	.29	64.3	62.9	I0017
117	306	600	-.01	.09	.96	-1.8	.95	-1.8	.36	.29	64.3	62.9	I0117
76	309	600	-.03	.09	1.25	9.5	1.27	9.5	-.09	.29	46.8	63.0	I0076
80	310	600	-.04	.09	.91	-3.9	.90	-3.7	.43	.29	70.7	63.0	I0080
13	321	600	-.12	.09	.87	-5.5	.86	-5.5	.49	.29	70.7	63.4	I0013
113	321	600	-.12	.09	.87	-5.5	.86	-5.5	.49	.29	70.7	63.4	I0113
53	323	600	-.13	.09	.93	-3.0	.92	-3.1	.41	.29	68.3	63.4	I0053
78	323	600	-.13	.09	.90	-4.0	.89	-4.2	.45	.29	70.3	63.4	I0078
73	328	600	-.17	.09	.97	-1.2	.97	-1.1	.34	.29	65.5	63.6	I0073
63	330	600	-.18	.09	.87	-5.5	.85	-5.6	.50	.29	72.7	63.7	I0063
59	331	600	-.19	.09	.94	-2.5	.93	-2.4	.39	.29	69.2	63.7	I0059
8	342	600	-.27	.09	.89	-4.4	.87	-4.5	.47	.29	69.0	64.2	I0008
108	342	600	-.27	.09	.89	-4.4	.87	-4.5	.47	.29	69.0	64.2	I0108
23	347	600	-.31	.09	.86	-5.6	.83	-5.9	.52	.29	71.3	64.4	I0023
2	349	600	-.32	.09	1.08	2.7	1.08	2.6	.18	.29	59.0	64.6	I0002
102	349	600	-.32	.09	1.08	2.7	1.08	2.6	.18	.29	59.0	64.6	I0102
74	350	600	-.33	.09	.83	-6.5	.81	-6.6	.55	.29	74.8	64.6	I0074
61	352	600	-.35	.09	.90	-3.6	.89	-3.6	.44	.29	70.8	64.7	I0061
71	355	600	-.37	.09	.97	-1.0	.96	-1.1	.34	.29	67.7	64.9	I0071
18	357	600	-.38	.09	.92	-2.9	.92	-2.6	.41	.29	70.5	65.0	I0018
118	357	600	-.38	.09	.92	-2.9	.92	-2.6	.41	.29	70.5	65.0	I0118
114	361	600	-.41	.09	.96	-1.4	.97	-1.0	.35	.29	67.8	65.3	I0114
14	362	600	-.42	.09	.96	-1.3	.97	-.9	.35	.29	67.7	65.3	I0014
11	371	600	-.49	.09	1.00	.1	1.03	1.0	.27	.29	67.7	65.9	I0011
77	371	600	-.49	.09	.90	-3.5	.87	-3.9	.46	.29	69.3	65.9	I0077
111	371	600	-.49	.09	1.00	.1	1.03	1.0	.27	.29	67.7	65.9	I0111
5	373	600	-.51	.09	.94	-2.1	.92	-2.2	.39	.29	69.7	66.1	I0005
105	373	600	-.51	.09	.94	-2.1	.92	-2.2	.39	.29	69.7	66.1	I0105
65	374	600	-.51	.09	.97	-1.1	.94	-1.7	.35	.29	65.2	66.2	I0065
6	375	600	-.52	.09	1.02	.8	1.06	1.6	.24	.29	68.3	66.2	I0006
26	375	600	-.52	.09	1.03	.9	1.06	1.6	.24	.29	67.7	66.2	I0026
106	375	600	-.52	.09	1.02	.8	1.06	1.6	.24	.29	68.3	66.2	I0106
83	379	600	-.55	.09	.87	-4.2	.85	-4.4	.49	.29	72.3	66.5	I0083
52	383	600	-.58	.09	1.09	2.8	1.12	3.0	.14	.29	63.7	66.9	I0052
92	389	600	-.63	.09	.98	-.6	.96	-1.1	.33	.29	65.2	67.4	I0092
48	393	600	-.66	.09	1.08	2.4	1.12	2.7	.15	.29	66.8	67.8	I0048
68	398	600	-.70	.09	.81	-5.8	.75	-6.4	.59	.29	74.5	68.3	I0068

90	401	600	-.73	.09	.97	-.8	.97	-.7	.33	.28	71.0	68.7	I0090
115	401	600	-.73	.09	.85	-4.3	.81	-4.7	.52	.28	71.0	68.7	I0115
15	402	600	-.74	.09	.86	-4.2	.81	-4.6	.52	.28	70.8	68.8	I0015
22	406	600	-.77	.09	.83	-4.8	.77	-5.4	.55	.28	71.5	69.3	I0022
81	413	600	-.83	.09	.90	-2.7	.87	-2.8	.44	.28	74.5	70.1	I0081
70	421	600	-.90	.09	.95	-1.2	.96	-.8	.35	.28	73.8	71.1	I0070
109	421	600	-.90	.09	.94	-1.4	.94	-1.1	.36	.28	73.5	71.1	I0109
9	422	600	-.90	.09	.95	-1.3	.95	-1.1	.36	.28	73.7	71.2	I0009
64	422	600	-.90	.09	.92	-1.9	.88	-2.4	.41	.28	72.7	71.2	I0064
57	424	600	-.92	.09	.87	-3.2	.82	-3.7	.48	.28	74.3	71.5	I0057
55	428	600	-.96	.09	.86	-3.5	.82	-3.6	.50	.28	75.7	72.0	I0055
72	435	600	-1.02	.10	.93	-1.5	.93	-1.2	.37	.27	76.3	73.0	I0072
16	439	600	-1.06	.10	.87	-3.0	.76	-4.5	.50	.27	73.3	73.6	I0016
116	439	600	-1.06	.10	.87	-3.0	.76	-4.5	.50	.27	73.3	73.6	I0116
24	443	600	-1.09	.10	.90	-2.2	.84	-2.8	.43	.27	75.8	74.2	I0024
79	443	600	-1.09	.10	.91	-2.1	.85	-2.6	.42	.27	76.8	74.2	I0079
56	444	600	-1.10	.10	.91	-1.9	.86	-2.5	.42	.27	76.3	74.3	I0056
93	445	600	-1.11	.10	.87	-2.8	.78	-3.9	.49	.27	74.8	74.5	I0093
86	448	600	-1.14	.10	.89	-2.4	.83	-3.0	.45	.27	76.8	74.9	I0086
7	451	600	-1.17	.10	.89	-2.3	.81	-3.2	.45	.27	76.0	75.4	I0007
107	451	600	-1.17	.10	.89	-2.3	.81	-3.2	.45	.27	76.0	75.4	I0107
112	458	600	-1.24	.10	.93	-1.4	.88	-1.9	.38	.26	75.7	76.4	I0112
12	459	600	-1.25	.10	.93	-1.3	.88	-1.9	.38	.26	75.8	76.6	I0012
21	459	600	-1.25	.10	.89	-2.2	.85	-2.3	.44	.26	76.8	76.6	I0021
32	464	600	-1.30	.10	.92	-1.6	.86	-2.1	.40	.26	77.2	77.4	I0032
100	464	600	-1.30	.10	.90	-2.0	.81	-3.0	.44	.26	77.8	77.4	I0100
49	478	600	-1.44	.10	.95	-.8	.96	-.5	.32	.25	80.2	79.7	I0049
120	481	600	-1.48	.11	.92	-1.4	.79	-2.9	.41	.25	80.5	80.2	I0120
1	482	600	-1.49	.11	1.01	.1	.99	-.1	.24	.25	80.7	80.3	I0001
20	482	600	-1.49	.11	.92	-1.3	.79	-2.8	.40	.25	80.7	80.3	I0020
101	483	600	-1.50	.11	1.01	.2	.99	-.1	.24	.25	80.8	80.5	I0101
4	503	600	-1.74	.11	.85	-2.1	.65	-4.3	.51	.23	83.8	83.8	I0004
104	503	600	-1.74	.11	.85	-2.1	.65	-4.3	.51	.23	83.8	83.8	I0104
29	504	600	-1.76	.11	.91	-1.3	.93	-.7	.36	.23	84.0	84.0	I0029
MEAN	309.4	600.0	.00	.10	.99	-.4	1.06	.0			71.9	70.9	
P.SD	125.3	.0	1.09	.02	.11	3.1	.27	3.8			8.7	8.2	

In answering the **RQ 1**, the Winsteps software programme was used to calibrate the responses of the 600 testees to the 120 MAT items. Table 1 above shows the difficulty indices in the fourth column, item **39** is the most difficult item in the test. The difficulty of this item is estimated to be **3.03logits** with the standard error of **0.18** while item **29** is the easiest with **-1.76logits** and standard error of **0.11**.

Research Question 2: What are the total number of items in the constructed Mathematics Aptitude Test (MAT) that fit into the Rasch model?

In answering the research question 2, the **infit** and **outfit** columns for both **MNSQ** and **ZSTD** in table 1 above were equally used. The table indicates that 33 items fit into the Rasch model, the listed items are item **88, 62, 3, 103, 119, 19, 58, 17, 117, 73, 71, 114, 14, 11, 111, 65, 6, 26, 106, 52, 92, 90, 70, 109, 9, 72, 112, 12, 49, 1, 20, 101** and **29.** These are items that fell within the recommended value ranging between **0.6 - 1.4.** Also**,** some of the items showed a negative correlation which when removed improves the quality of the data; the reliability was improved. They should be kept for future use while the remaining highlighted **87** items are omitted, deleted or revised because of lack of fit to the model. These items are measuring something other than the intended content and construct. Therefore, 33 items met the model assumption which was an indication of their **unidimensionality**. The 33 items showed the construct validity of the MAT.

Research Question 3: What is the difficulty index of each item in the constructed Mathematics Aptitude Test (MAT) using the 2-PL model?

Table 2: Estimates of b and a parameter of MAT

ITEM	INTERCEPT S.E.	**SLOPE(a) S.E.**	**THRESHOLD(b) S.E.**	LOADING S.E.	ASYMPTOTE(c) S.E.	CHISQ **(PROB)**	DF
ITEM0001	0.646	**2.419**	**-0.558**	0.770	0.000	12.8	5.0
	0.083*	0.141*	0.118*	0.090*	0.000*	**(0.0252)**	
ITEM0002	0.679	**0.192**	**-3.527**	0.189	0.000	8.5	8.0
	0.052*	0.049*	0.966*	0.048*	0.000*	(0.3871)	
ITEM0003	0.499	**0.220**	**-2.272**	0.215	0.000	6.2	9.0
	0.049*	0.051*	0.600*	0.050*	0.000*	(0.7229)	
ITEM0004	0.380	**0.378**	**-1.007**	0.353	0.000	9.1	8.0
	0.049*	0.064*	0.228*	0.060*	0.000*	(0.3320)	
ITEM0005	0.249	**0.250**	**-0.996**	0.242	0.000	19.5	9.0
	0.047*	0.055*	0.303*	0.054*	0.000*	**(0.0209)**	
ITEM0006	0.182	**0.173**	**-1.048**	0.171	0.000	5.8	9.0
	0.046*	0.041*	0.366*	0.041*	0.000*	(0.7618)	
ITEM0007	0.365	**0.260**	**-1.405**	0.251	0.000	10.4	9.0
	0.048*	0.053*	0.358*	0.051*	0.000*	(0.3215)	
ITEM0008	0.310	**0.194**	**-1.595**	0.191	0.000	13.1	9.0
	0.047*	0.046*	0.455*	0.045*	0.000*	(0.1564)	
ITEM0009	0.174	**0.209**	**-0.833**	0.204	0.000	8.2	9.0
	0.046*	0.046*	0.289*	0.045*	0.000*	(0.5148)	
ITEM0010	0.421	**0.280**	**-1.504**	0.269	0.000	6.2	8.0
	0.049*	0.058*	0.382*	0.056*	0.000*	(0.6218)	
ITEM0011	0.456	**0.472**	**-0.966**	0.427	0.000	9.2	8.0
	0.052*	0.072*	0.203*	0.065*	0.000*	(0.3251)	
ITEM0012	0.193	**0.266**	**-0.725**	0.257	0.000	8.5	9.0
	0.046*	0.048*	0.217*	0.046*	0.000*	(0.4813)	
ITEM0013	0.388	**0.556**	**-0.697**	0.486	0.000	3.5	8.0
	0.052*	0.079*	0.153*	0.069*	0.000*	(0.8989)	
ITEM0014	0.524	**0.529**	**-0.990**	0.468	0.000	8.0	6.0
	0.055*	0.087*	0.221*	0.077*	0.000*	(0.2348)	
ITEM0015	0.182	**0.225**	**-0.809**	0.220	0.000	3.5	9.0
	0.046*	0.050*	0.276*	0.048*	0.000*	(0.9415)	

ITEM0016	0.377	**0.373**	**-1.011**	0.350	0.000	5.3	8.0
	0.050*	0.066*	0.244*	0.062*	0.000*	(0.7302)	
ITEM0017	0.374	**0.371**	**-1.007**	0.348	0.000	7.0	8.0
	0.049*	0.064*	0.235*	0.060*	0.000*	(0.5358)	
ITEM0018	0.340	**0.135**	**-2.522**	0.134	0.000	8.9	9.0
	0.047*	0.036*	0.750*	0.035*	0.000*	(0.4499)	
ITEM0019	0.273	**0.327**	**-0.836**	0.311	0.000	6.6	9.0
	0.048*	0.060*	0.227*	0.057*	0.000*	(0.6786)	
ITEM0020	0.220	**0.216**	**-1.015**	0.212	0.000	10.8	9.0
	0.046*	0.048*	0.317*	0.047*	0.000*	(0.2865)	
ITEM0021	0.159	**0.356**	**-0.446**	0.336	0.000	7.9	9.0
	0.047*	0.059*	0.155*	0.055*	0.000*	(0.5444)	
ITEM0022	0.244	**0.405**	**-0.602**	0.376	0.000	8.1	8.0
	0.048*	0.061*	0.156*	0.057*	0.000*	(0.4194)	
ITEM0023	0.215	**0.181**	**-1.188**	0.178	0.000	11.9	9.0
	0.046*	0.043*	0.378*	0.042*	0.000*	(0.2185)	
ITEM0024	0.296	**0.504**	**-0.588**	0.450	0.000	3.8	8.0
	0.050*	0.072*	0.137*	0.064*	0.000*	(0.8732)	
ITEM0025	0.380	**1.311**	**-0.290**	0.795	0.000	10.4	5.0
	0.067*	0.141*	0.068*	0.085*	0.000*	(0.0640)	
ITEM0026	0.478	**0.385**	**-1.242**	0.359	0.000	7.9	9.0
	0.051*	0.064*	0.264*	0.060*	0.000*	(0.5444)	
ITEM0027	0.563	**0.506**	**-1.111**	0.452	0.000	7.3	7.0
	0.054*	0.078*	0.224*	0.070*	0.000*	(0.3985)	
ITEM0028	0.239	**0.334**	**-0.717**	0.317	0.000	12.4	8.0
	0.048*	0.059*	0.202*	0.056*	0.000*	(0.1340)	
ITEM0029	0.397	**0.398**	**-0.998**	0.370	0.000	3.9	8.0
	0.050*	0.067*	0.234*	0.062*	0.000*	(0.8703)	
ITEM0030	0.255	**0.333**	**-0.764**	0.316	0.000	20.2	9.0
	0.047*	0.056*	0.196*	0.053*	0.000*	**(0.0167)**	
ITEM0031	0.363	**0.225**	**-1.610**	0.220	0.000	5.9	9.0
	0.048*	0.050*	0.434*	0.049*	0.000*	(0.7544)	
ITEM0032	0.213	**0.321**	**-0.663**	0.306	0.000	5.5	9.0
	0.047*	0.057*	0.196*	0.054*	0.000*	(0.7849)	
ITEM0033	0.271	**0.271**	**-0.999**	0.262	0.000	11.8	9.0
	0.047*	0.055*	0.278*	0.053*	0.000*	(0.2225)	
ITEM0034	0.312	**0.431**	**-0.724**	0.396	0.000	11.7	8.0
	0.049*	0.063*	0.168*	0.058*	0.000*	(0.1651)	
ITEM0035	0.321	**0.273**	**-1.175**	0.264	0.000	11.9	8.0
	0.047*	0.051*	0.293*	0.049*	0.000*	(0.1558)	

ITEM0036	0.285	**0.228**	**-1.249**	0.223	0.000	11.1	9.0
	0.047*	0.046*	0.326*	0.045*	0.000*	(0.2708)	
ITEM0037	0.362	**0.252**	**-1.437**	0.245	0.000	10.1	9.0
	0.048*	0.053*	0.370*	0.052*	0.000*	(0.3463)	
ITEM0038	0.366	**0.252**	**-1.451**	0.245	0.000	17.8	9.0
	0.048*	0.054*	0.386*	0.053*	0.000*	**(0.0381)**	
ITEM0039	0.314	**0.243**	**-1.291**	0.236	0.000	7.0	9.0
	0.047*	0.052*	0.353*	0.051*	0.000*	(0.6321)	
ITEM0040	0.297	**0.184**	**-1.607**	0.181	0.000	9.1	9.0
	0.046*	0.045*	0.470*	0.044*	0.000*	(0.4252)	
ITEM0041	0.646	**1.208**	**-0.535**	0.770	0.000	12.8	5.0
	0.083*	0.141*	0.118*	0.090*	0.000*	**(0.0252)**	
ITEM0042	0.679	**0.192**	**-3.527**	0.189	0.000	8.5	8.0
	0.052*	0.049*	0.966*	0.048*	0.000*	(0.3871)	
ITEM0043	0.499	**0.220**	**-2.272**	0.215	0.000	6.2	9.0
	0.049*	0.051*	0.600*	0.050*	0.000*	(0.7229)	
ITEM0044	0.380	**0.378**	**-1.007**	0.353	0.000	9.1	8.0
	0.049*	0.064*	0.228*	0.060*	0.000*	(0.3320)	
ITEM0045	0.249	**0.250**	**-0.996**	0.242	0.000	19.5	9.0
	0.047*	0.055*	0.303*	0.054*	0.000*	**(0.0209)**	
ITEM0046	0.182	**0.173**	**-1.048**	0.171	0.000	5.8	9.0
	0.046*	0.041*	0.366*	0.041*	0.000*	(0.7618)	
ITEM0047	0.365	**0.260**	**-1.405**	0.251	0.000	10.4	9.0
	0.048*	0.053*	0.358*	0.051*	0.000*	(0.3215)	
ITEM0048	0.310	**0.194**	**-1.595**	0.191	0.000	13.1	9.0
	0.047*	0.046*	0.455*	0.045*	0.000*	(0.1564)	
ITEM0049	0.174	**0.209**	**-0.833**	0.204	0.000	8.2	9.0
	0.046*	0.046*	0.289*	0.045*	0.000*	(0.5148)	
ITEM0050	0.421	**0.280**	**-1.504**	0.269	0.000	6.2	8.0
	0.049*	0.058*	0.382*	0.056*	0.000*	(0.6218)	
ITEM0051	0.456	**0.472**	**-0.966**	0.427	0.000	9.2	8.0
	0.052*	0.072*	0.203*	0.065*	0.000*	(0.3251)	
ITEM0052	0.193	**0.266**	**-0.725**	0.257	0.000	8.5	9.0
	0.046*	0.048*	0.217*	0.046*	0.000*	(0.4813)	
ITEM0053	0.388	**0.556**	**-0.697**	0.486	0.000	3.5	8.0
	0.052*	0.079*	0.153*	0.069*	0.000*	(0.8989)	
ITEM0054	0.524	**0.529**	**-0.990**	0.468	0.000	8.0	6.0
	0.055*	0.087*	0.221*	0.077*	0.000*	(0.2348)	
ITEM0055	0.182	**0.225**	**-0.809**	0.220	0.000	3.5	9.0
	0.046*	0.050*	0.276*	0.048*	0.000*	(0.9415)	

ITEM0056	0.377	**0.373**	**-1.011**	0.350	0.000	5.3	8.0
	0.050*	0.066*	0.244*	0.062*	0.000*	(0.7302)	
ITEM0057	0.374	**0.371**	**-1.007**	0.348	0.000	7.0	8.0
	0.049*	0.064*	0.235*	0.060*	0.000*	(0.5358)	
ITEM0058	0.340	**0.135**	**-2.522**	0.134	0.000	8.9	9.0
	0.047*	0.036*	0.750*	0.035*	0.000*	(0.4499)	
ITEM0059	0.273	**0.327**	**-0.836**	0.311	0.000	6.6	9.0
	0.048*	0.060*	0.227*	0.057*	0.000*	(0.6786)	
ITEM0060	0.220	**0.216**	**-1.015**	0.212	0.000	10.8	9.0
	0.046*	0.048*	0.317*	0.047*	0.000*	(0.2865)	
ITEM0061	0.159	**0.356**	**-0.446**	0.336	0.000	7.9	9.0
	0.047*	0.059*	0.155*	0.055*	0.000*	(0.5444)	
ITEM0062	0.244	**0.405**	**-0.602**	0.376	0.000	8.1	8.0
	0.048*	0.061*	0.156*	0.057*	0.000*	(0.4194)	
ITEM0063	0.215	**0.181**	**-1.188**	0.178	0.000	11.9	9.0
	0.046*	0.043*	0.378*	0.042*	0.000*	(0.2185)	
ITEM0064	0.296	**0.504**	**-0.588**	0.450	0.000	3.8	8.0
	0.050*	0.072*	0.137*	0.064*	0.000*	(0.8732)	
ITEM0065	0.380	**2.662**	**-0.304**	0.795	0.000	10.4	5.0
	0.067*	0.141*	0.068*	0.085*	0.000*	**(0.0064)**	
ITEM0066	0.478	**0.385**	**-1.242**	0.359	0.000	7.9	9.0
	0.051*	0.064*	0.264*	0.060*	0.000*	(0.5444)	
ITEM0067	0.563	**0.506**	**-1.111**	0.452	0.000	7.3	7.0
	0.054*	0.078*	0.224*	0.070*	0.000*	(0.3985)	
ITEM0068	0.239	**0.334**	**-0.717**	0.317	0.000	12.4	8.0
	0.048*	0.059*	0.202*	0.056*	0.000*	(0.1340)	
ITEM0069	0.397	**0.398**	**-0.998**	0.370	0.000	3.9	8.0
	0.050*	0.067*	0.234*	0.062*	0.000*	(0.8703)	
ITEM0070	0.255	**0.333**	**-0.764**	0.316	0.000	20.2	9.0
	0.047*	0.056*	0.196*	0.053*	0.000*	**(0.0167)**	
ITEM0071	0.363	**0.225**	**-1.610**	0.220	0.000	5.9	9.0
	0.048*	0.050*	0.434*	0.049*	0.000*	(0.7544)	
ITEM0072	0.213	**0.321**	**-0.663**	0.306	0.000	5.5	9.0
	0.047*	0.057*	0.196*	0.054*	0.000*	(0.7849)	
ITEM0073	0.271	**0.271**	**-0.999**	0.262	0.000	11.8	9.0
	0.047*	0.055*	0.278*	0.053*	0.000*	(0.2225)	
ITEM0074	0.312	**0.431**	**-0.724**	0.396	0.000	11.7	8.0
	0.049*	0.063*	0.168*	0.058*	0.000*	(0.1651)	
ITEM0075	0.321	**0.273**	**-1.175**	0.264	0.000	11.9	8.0
	0.047*	0.051*	0.293*	0.049*	0.000*	(0.1558)	

ITEM0076	0.285	**0.228**	**-1.249**	0.223	0.000	11.1	9.0
	0.047*	0.046*	0.326*	0.045*	0.000*	(0.2708)	
ITEM0077	0.362	**0.252**	**-1.437**	0.245	0.000	10.1	9.0
	0.048*	0.053*	0.370*	0.052*	0.000*	(0.3463)	
ITEM0078	0.366	**0.252**	**-1.451**	0.245	0.000	17.8	9.0
	0.048*	0.054*	0.386*	0.053*	0.000*	**(0.0381)**	
ITEM0079	0.314	**0.243**	**-1.291**	0.236	0.000	7.0	9.0
	0.047*	0.052*	0.353*	0.051*	0.000*	(0.6321)	
ITEM0080	0.297	**0.184**	**-1.607**	0.181	0.000	9.1	9.0
	0.046*	0.045*	0.470*	0.044*	0.000*	(0.4252)	
ITEM0081	0.646	**1.208**	**-0.535**	0.770	0.000	12.8	5.0
	0.083*	0.141*	0.118*	0.090*	0.000*	**(0.0252)**	
ITEM0082	0.679	**0.192**	**-3.527**	0.189	0.000	8.5	8.0
	0.052*	0.049*	0.966*	0.048*	0.000*	(0.3871)	
ITEM0083	0.499	**0.220**	**-2.272**	0.215	0.000	6.2	9.0
	0.049*	0.051*	0.600*	0.050*	0.000*	(0.7229)	
ITEM0084	0.380	**0.378**	**-1.007**	0.353	0.000	9.1	8.0
	0.049*	0.064*	0.228*	0.060*	0.000*	(0.3320)	
ITEM0085	0.249	**0.250**	**-0.996**	0.242	0.000	19.5	9.0
	0.047*	0.055*	0.303*	0.054*	0.000*	**(0.0209)**	
ITEM0086	0.182	**0.173**	**-1.048**	0.171	0.000	5.8	9.0
	0.046*	0.041*	0.366*	0.041*	0.000*	(0.7618)	
ITEM0087	0.365	**0.260**	**-1.405**	0.251	0.000	10.4	9.0
	0.048*	0.053*	0.358*	0.051*	0.000*	(0.3215)	
ITEM0088	0.310	**0.194**	**-1.595**	0.191	0.000	13.1	9.0
	0.047*	0.046*	0.455*	0.045*	0.000*	(0.1564)	
ITEM0089	0.174	**0.209**	**-0.833**	0.204	0.000	8.2	9.0
	0.046*	0.046*	0.289*	0.045*	0.000*	(0.5148)	
ITEM0090	0.421	**0.280**	**-1.504**	0.269	0.000	6.2	8.0
	0.049*	0.058*	0.382*	0.056*	0.000*	(0.6218)	
ITEM0091	0.456	**0.472**	**-0.966**	0.427	0.000	9.2	8.0
	0.052*	0.072*	0.203*	0.065*	0.000*	(0.3251)	
ITEM0092	0.193	**0.266**	**-0.725**	0.257	0.000	8.5	9.0
	0.046*	0.048*	0.217*	0.046*	0.000*	(0.4813)	
ITEM0093	0.388	**0.556**	**-0.697**	0.486	0.000	3.5	8.0
	0.052*	0.079*	0.153*	0.069*	0.000*	(0.8989)	
ITEM0094	0.524	**0.529**	**-0.990**	0.468	0.000	8.0	6.0
	0.055*	0.087*	0.221*	0.077*	0.000*	(0.2348)	
ITEM0095	0.182	**0.225**	**-0.809**	0.220	0.000	3.5	9.0
	0.046*	0.050*	0.276*	0.048*	0.000*	(0.9415)	

ITEM0096	0.377	**0.373**	**-1.011**	0.350	0.000	5.3	8.0
	0.050*	0.066*	0.244*	0.062*	0.000*	(0.7302)	
ITEM0097	0.374	**0.371**	**-1.007**	0.348	0.000	7.0	8.0
	0.049*	0.064*	0.235*	0.060*	0.000*	(0.5358)	
ITEM0098	0.340	**0.135**	**-2.522**	0.134	0.000	8.9	9.0
	0.047*	0.036*	0.750*	0.035*	0.000*	(0.4499)	
ITEM0099	0.273	**0.327**	**-0.836**	0.311	0.000	6.6	9.0
	0.048*	0.060*	0.227*	0.057*	0.000*	(0.6786)	
ITEM0100	0.220	**0.216**	**-1.015**	0.212	0.000	10.8	9.0
	0.046*	0.048*	0.317*	0.047*	0.000*	(0.2865)	
ITEM0101	0.159	**0.356**	**-0.446**	0.336	0.000	7.9	9.0
	0.047*	0.059*	0.155*	0.055*	0.000*	(0.5444)	
ITEM0102	0.244	**0.405**	**-0.602**	0.376	0.000	8.1	8.0
	0.048*	0.061*	0.156*	0.057*	0.000*	(0.4194)	
ITEM0103	0.215	**0.181**	**-1.188**	0.178	0.000	11.9	9.0
	0.046*	0.043*	0.378*	0.042*	0.000*	(0.2185)	
ITEM0104	0.296	**0.504**	**-0.588**	0.450	0.000	3.8	8.0
	0.050*	0.072*	0.137*	0.064*	0.000*	(0.8732)	
ITEM0105	0.380	**1.311**	**-0.290**	0.795	0.000	10.4	5.0
	0.067*	0.141*	0.068*	0.085*	0.000*	**(0.0640)**	
ITEM0106	0.478	**0.385**	**-1.242**	0.359	0.000	7.9	9.0
	0.051*	0.064*	0.264*	0.060*	0.000*	(0.5444)	
ITEM0107	0.563	**0.506**	**-1.111**	0.452	0.000	7.3	7.0
	0.054*	0.078*	0.224*	0.070*	0.000*	(0.3985)	
ITEM0108	0.239	**0.334**	**-0.717**	0.317	0.000	12.4	8.0
	0.048*	0.059*	0.202*	0.056*	0.000*	(0.1340)	
ITEM0109	0.397	**0.398**	**-0.998**	0.370	0.000	3.9	8.0
	0.050*	0.067*	0.234*	0.062*	0.000*	(0.8703)	
ITEM0110	0.255	**0.333**	**-0.764**	0.316	0.000	20.2	9.0
	0.047*	0.056*	0.196*	0.053*	0.000*	**(0.0167)**	
ITEM0111	0.363	**0.225**	**-1.610**	0.220	0.000	5.9	9.0
	0.048*	0.050*	0.434*	0.049*	0.000*	(0.7544)	
ITEM0112	0.213	**0.321**	**-0.663**	0.306	0.000	5.5	9.0
	0.047*	0.057*	0.196*	0.054*	0.000*	(0.7849)	
ITEM0113	0.271	**0.271**	**-0.999**	0.262	0.000	11.8	9.0
	0.047*	0.055*	0.278*	0.053*	0.000*	(0.2225)	
ITEM0114	0.312	**0.431**	**-0.724**	0.396	0.000	11.7	8.0
	0.049*	0.063*	0.168*	0.058*	0.000*	(0.1651)	
ITEM0115	0.321	**0.273**	**-1.175**	0.264	0.000	11.9	8.0
	0.047*	0.051*	0.293*	0.049*	0.000*	(0.1558)	

ITEM0116	0.285	**0.228**	**-1.249**	0.223	0.000	11.1	9.0
	0.047*	0.046*	0.326*	0.045*	0.000*	(0.2708)	
ITEM0117	0.362	**0.252**	**-1.437**	0.245	0.000	10.1	9.0
	0.048*	0.053*	0.370*	0.052*	0.000*	(0.3463)	
ITEM0118	0.366	**0.252**	**-1.451**	0.245	0.000	17.8	9.0
	0.048*	0.054*	0.386*	0.053*	0.000*	**(0.0381)**	
ITEM0119	0.314	**0.243**	**-1.291**	0.236	0.000	7.0	9.0
	0.047*	0.052*	0.353*	0.051*	0.000*	(0.6321)	
ITEM0120	0.297	**0.184**	**-1.607**	0.181	0.000	9.1	9.0
	0.046*	0.045*	0.470*	0.044*	0.000*	(0.4252)	

* STANDARD ERROR

LARGEST CHANGE = 0.114914 1100.81002.0
(0.0156)

PARAMETER	MEAN	STD DEV
SLOPE	0.356	0.234
LOG (SLOPE)	-1.164	0.467
THRESHOLD	-1.136	0.588

To answer this research question, BILOG MG-3 software programme was used to calibrate the responses of 600 testees to the 120-items of Mathematics Aptitude Test. Table 2 above shows the item parameter estimates obtained using the two-parameter model (2-PL model); Difficulty indices are in column 4, which is the b, threshold.

RQ 4: What are the total number of items in the constructed Mathematics Aptitude Test (MAT) that fit into the 2-PL model?

To answer this research question, BILOG MG-3 software programme was used to calibrate the responses of 600 testees to the 120 items of MAT. Chi-square probability table of the Bilog MG was used in determining the fitness of the item at 0.05 level of significance. The difficulty index (b) of the MAT items is in the fourth column on the estimates of b parameters of the MAT table above with threshold (b) highlighted. Difficulty index (b) ranged from -3.527 to -.290. This shows that generally, the items are too simple for the respondents. By implication, **thirteen (13) items** were scientifically and statistically significant and do not fit into the 2-PL model of IRT, such items are **1, 5, 30, 38, 41, 45, 65, 70, 78, 81, 85, 110** and **118.** Therefore, by interpretation **107 items** fit into the 2-PL model. All item fit/misfit were determined at a 0.05 level of significance.

Discussion

Difficulty indices of the MAT items using the 2-PL model of IRT

The difficulty index (b) ranged from -3.527 to -.290. This shows that generally, the items are too simple for the respondents. By implication, **thirteen (13) items** were statistically significant and do not fit into the 2-PL model of IRT and by interpretation 107 items fit into the 2-PL model. All item fit/misfit were determined at a 0.05 level of significance. Among the items that fit into the 2-PL model were observed not to fit into the Rasch model.

Generally, an important aspect of the IRT approach is the selection of an IRT model to represent the data", the data were analyzed using Rasch and 2-PL models. The researcher's conclusion "is that for this assessment the Rasch model is preferred over the 2PL models because the model offers a significant improvement in the fit of the data to the model over the alternative models. In other words, the additional parameters estimated in the Rasch model are justified because they help provide a better fit to the data." This could be the result of the objectivity of the Rasch in item selection of fitness. Only items, 1, 65 and 70 were all recognized by both models. They are therefore suggested to be removed from the test instrument.

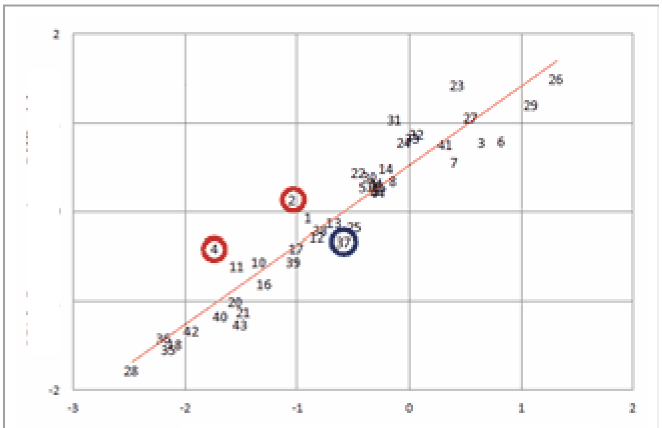
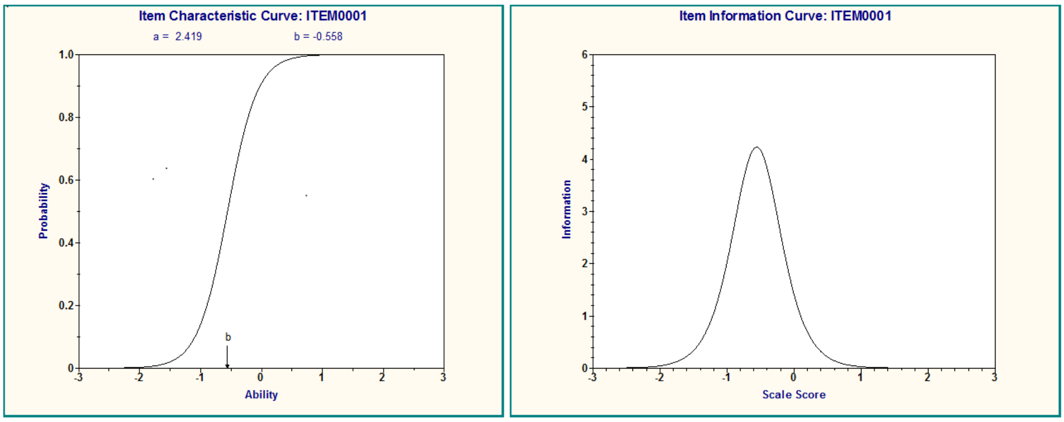


Figure 1. The plot of item difficulties from, Table 2. Person thetas are N(0,1).

Let's look more closely at these analyses. The researcher helpfully reports the item difficulties, b, according to Rasch and 2PL in Table 2. These are plotted in Fig.1. The person ability theta distribution is stated to be constrained to N(0,1) in both Rasch and 2PL analyses. In the Figure, items 1 and 65 have the highest 2PL discrimination and item 58 and 98 have the lowest discrimination. The Researcher attributes the average 0.5 z-score (unit-normal deviate) difference between the Rasch and 2PL estimates to the 2PL discrimination. He identifies Item 1 and 65 as more accurately estimated by Rasch than by 2PL because they met all the required prerequisites for item selection under the Rasch Objectivity standard. They equally have a positive PT measure correlation.



2-Parameter Model, Normal Metric **Item: 1**

The parameter a is the item discriminating power, the reciprocal (1/a) is the item dispersion, and the parameter b is an item location parameter.

To verify that the 1-PL analysis does correspond to a standard Rasch analysis, I simulated data using Bilog's 2-PL parameter estimates and an N(0,1) theta distribution. Rasch b-parameters for these data were estimated with Winsteps (chosen because its weighting capabilities allow an exact match in the data to the 4PL ogives and theta distribution). The plot of item difficulties is shown in Fig.2. The noticeable outliers are items 1 and 65 (which have high 2PL - discrimination values). Overall, this simulation confirms that the reported 1PL analysis reasonably matches a Rasch dichotomous analysis slightly.

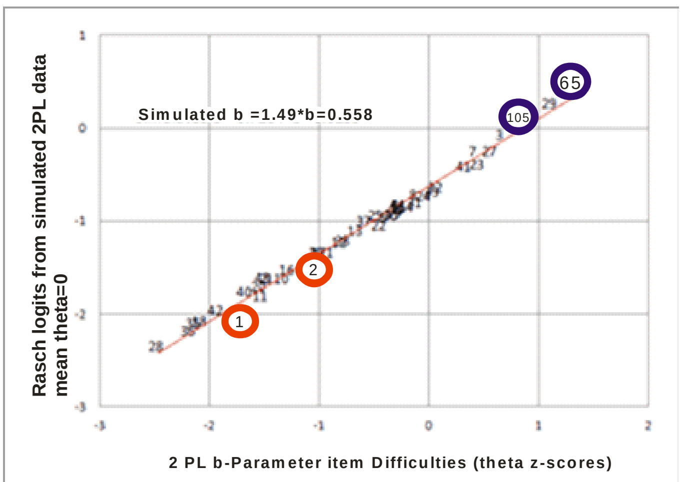
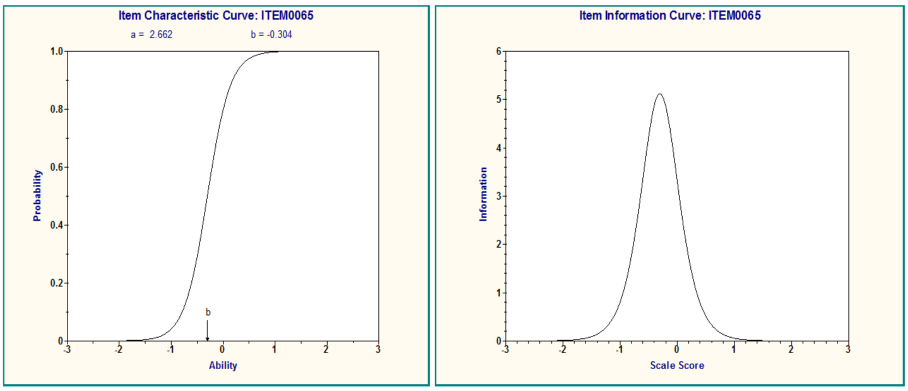


Figure 2. The plot of Rasch item difficulties estimated from data simulated with 2PL's estimates



2-Parameter Model, Normal Metric **Item: 65**

The parameter a is the item discriminating power, the reciprocal (1/a) is the item dispersion, and the parameter b is an item location parameter.

More interesting are the fit statistics for the simulated items from the Rasch analysis. All the items have acceptable fit statistics! The most under-fitting item is item 39 (highest difficulty value) with an outfit mean-square 1.78. The most over-fitting item is item 29 (with the highest 2PL discrimination) with an outfit mean-square of 0.93. The infit mean-squares are within the range of the outfit mean-squares. Surprisingly, item 1 (high 2PL discriminating value) only slightly under-fits with an outfit mean-square of 1.09, and item 65 (high 2PL discrimination) slightly over-fits due to its high 2PL discrimination. Though many simulated responses are flagged by Rasch as potential guesses, they are overwhelmed in the simulation by well-behaved data and so have little influence on the Rasch fit statistics. Surprisingly, if the original data did accord with the estimated 2PL parameters, then those data would also accord with the Rasch dichotomous parameters. Therefore, generally, the most appropriate model (i.e. the model involving the least number of estimated parameters) is preferred to represent the data" and this would motivate the selection of Rasch over 2PL!

This leads us to the scientific investigation of the items: quality control, efficiency, and effective assessment development. A major flaw in 2PL analysis is its lack of quality control of the data. What about items 1 and 65 with its high discriminating values? The researcher admits that there can be bad items but does not describe any attempt to discover if items 1 and 65 or any other of the 12 items are bad. However, "the (Rasch) model is then used as a yardstick that the item-response data must fit, or the item is discarded." The assumption is that item 1 and 65 fits the Rasch model and so is a good item (but did not fit the 2PL model statistics). The assumption is also that item 2 does not fit the 2PL model and so it would be discarded. The simulated evidence suggests that Rasch would keep items 1 and 65, but, based on the 2PL evidence, items 2 and 65 might be discarded.

The researchers reported the 2PL parameter estimates in Table 2. As we might expect, there is no correlation between 2PL item discrimination, a, and difficulty level, b, for item 1 and 65, they are with the highest discrimination value of 2.419 and 2.662, SE of 0.141 each respectively from the ICC. The difficulty values are -0.558 and -0.304 which negates the assumption that when items become more difficult, they discriminate more strongly between high and lower performers but was not so in this case in 2PL. The two items seem very simple but with high discrimination value! We might hypothesize that the 2PL analysis did not give us the true picture of these items while Rasch did. In this estimation, the maximum item discrimination appears to have been constrained to 2.0, so both items 1 (a=2.419) and item 65 (a=2.662) have the highest discrimination and 2PL has blindly accepted this pattern of item discrimination. Rasch analysis would flag the items with higher discriminations as over-fitting and perhaps locally dependent if unable to meet other conditions of the model fit. Items like 41 (a=1.208), 81 (a=1.208), 105 (a=1.311) were all discarded because they could not meet Rasch model standard fit. Therefore, if we are interested in measuring students' abilities, as opposed to describing this dataset, then we should seriously consider rejecting items recommended by the Rasch.

Discussion

From the data analyzed and described in the study, the 120 Items constructed showed that only a few of the items scaled through the Rasch model while a large number scaled through the 2-PL model. It, therefore, means that those items can be banked for future reference and use. Also, it was noted that few of the 33 items that fit into the Rasch model were not recognized by the 2-PL model whereas the majority of the 107 items of the 2-PL model did not fit into the Rasch model. This implies that the Rasch and the 2-PL models have functioned differently on some of the constructed MAT items. This shows the disparity between the two models. According to Bergan (2010), "In the Rasch approach, data that do not fit the theory expressed in the mathematical model are ignored or discarded. In the scientific (IRT) approach, the theory is discarded or modified if it is not supported by data." This view of "science" allows problematic data to control our thinking. Rasch takes a pro-active view of science. Every observation is an experiment that requires scrutiny. Was the experiment a success or a failure? Problematic data certainly should not be ignored, and if found to be fatally flawed must be discarded. Otherwise, we risk making false inferences that could have severe repercussions throughout the academic careers of these students.

Bergan (2010) reiterates that "it is expensive and risky to ignore objective data", but that is exactly what has happened in the 2PL analysis. The negative correlations and other potential aberrations in the objective data observed in Rasch have been ignored because the 2PL model has made no demands upon the quality of the data. Bergan admits that "Adherence to a scientific [IRT] approach does not imply that there are no bad items. Indeed, measurement conducted by the scientific approach facilitates effective item evaluation and selection." However, here it seems that 2PL does not accord with the scientific approach. It fails to examine the data. It hides problems in the data, and so acts against an effective evaluation. 2PL fails as a tool of science and curriculum development, but Rasch succeeds.

Conclusions

It was concluded that:

i. The difficulty indices range from 3.03logit to -1.76logit for the Rasch model.

ii. The difficulty indices range from -3.527logit to -.290logit for the 2PL model.

iii. 33 items fit into the Rasch model with the demonstration of good qualities because they were functioning in the intended ways while 107 items fit into the 2-PL model with their discrimination values ranging between .135 and 1.311 with the use of Bilog-MG3.

Recommendations

The study, therefore, recommends that the Rasch model should be adopted in test construction over the 2PL model since items fit to show the uni-dimensionality of the test. Also, item measure order in Rasch reduces any bias of any form according to literature. This will robust the curriculum, effectiveness of assessment in this era. Also, the researcher has observed that Aptitude Test items are not commonly used to determine students' placement at the next level. Therefore, recommends that the Aptitude Test such as MAT should be adopted for the placement of students in the schools' system. Rasch model of test development principle should be adopted, since, it does not discriminate between samples and also, shows high content and construct validity.

References

Ahmad, Z. K. & Nordin, A. (2012). Advance in Educational Measurement: A Rasch Model Analysis of Mathematics Proficiency Test. *International Journal of Social Science and Humanity, 2(3).*

Aliyu, R. T. & Akinoso, S. O. (2017). Development and validation of Mathematics Aptitude Test using the Rasch and 2-PL Models of IRT. *A paper published in Ibadan Journal of Educational Studies (IJES) 16(2), 1-15.*

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*

Aliyu, R. T. (2015). Development and validation of the Mathematics Achievement Test using the Rasch Model. *An Unpublished PhD Thesis at Delta State University, Abraka.*

Aliyu, R. T. & Ocheli, O.E. (2013). Development and Validation of College Mathematics with Item Response Theory (IRT) Models in Attaining Quality Education in Nigeria. *Delta Journal of Educational Research and Development (DJERD), 12(1), 130-140*

Andrich, D. (1992). "The application of an unfolding model of the PIRT type to the measurement of attitude". *Applied Psychological Measurement, 12, 33-35.*

Baghaei, P. & Amrahi, V. (2011). "Rasch Model as a construct validation tool" in *Rasch Measurement Transaction*, 22(1), 1145-1146

Bergan J.R. (2010) Assessing the Relative Fit of Alternative Item Response Theory Models to the Data. Tucson AZ: Assessment Technology Inc. http://ati-online.com/pdfs/researchK12/AlternativeIRTModels.pdf

Bond, T. G. & Fox, C. M. (2001). Applying *the Rasch model: Fundamental Measurement in Human Sciences, 1st ed,* Mahwah, NJ: Lawrence Erlbaum

Chen, S.Y., Ankenmann, R.D. & Chang, H.H. (2000). A comparison of item selection Rules at the early stage of computerized adaptive testing. Apply Psychological Measurement, 24, 241-255

Crocker, L. & Algina, J. (1986). Introduction to Classical and Modern Test Theory. *Orlando, FL: Holt, Rinehart and Winston Inc.*

Downing, S. M. (2013) "Item response theory: Applications of Modern Test Theory ", *Medical Education*, 37, 739-745.

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

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

Hambleton, R.K. & Swaminathan, H. (1985). Item Response Theory: Principles and Applications. *Boston: Kluwer. Nijhoff.*

Linn, R. L. (2000). Assessment and Accountability. Educational Researcher,29, (2) 4-6

Linacre, J. M. (2012). A user's guide to Winsteps

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

Nitko, A. J. (1996). Educational Assessment of Students. *The wright map. 2nd. ed. Englewood Cliffs, NJ: Merrill.*

Odili, J. N., Osadebe, P. U. & Aliyu, R. T. (2015). Assessment of Stability of Item The parameter in a Mathematics Achievement Test Under The Rasch Model. *A paper published in Journal of Association of Educational Researcher and Evaluators of Nigeria (ASSEREN), 1(1), 1-8*

Olaleye, O. O. & Aliyu, R. T. (2013). Development and Validation of Mathematics Achievement Test Items Using Item Response Theory (IRT) Models in Attaining Quality Education for National Development. *A paper presented and published in the Proceedings of Mathematics Association of Nigeria (MAN) at the 50th Anniversary of the Annual National conference of MAN, 82-95*

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

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

Reza, P., Baghaei, P & Ahmadi, H. S. (2011). Development and validation of an English Language Teacher competency test using Item Response Theory. *The International Journal of Education and psychological assessment, 8(2),54-68.*

Thissen, D., & Orlando, M. (2001) Test Scoring. Mahwah, NJ: Lawrence

ErlbaumAssociates. 3PL, Rasch, Quality-Control, and Science. J.M. Linacre. Rasch Measurement Transactions, 2014, 27(4) 1441- 1444

Wang, T & Vispoel, W. (1998). Properties of Abilities Estimation Method in Computer Adaptive Testing. *Journal of Educational Measurement*, 35, 109-135.

Wiberg M. (2004). Classical Test Theory Vs Item Test Theory. An Evaluation of the Theory Test in the Swedish driving-license Test. http:www.eedusci.umn.se/digital Assets159/5929-em-no-50p.d.f.cited 4/1/2016 1.43p.m.

Author Information

Aliyu R. Taiwo
Department of Mathematics Education,
Faculty of Education, *Lead City University*, *Ibadan*
aliyutaiwo2013@gmail.com

Sabainah Oyebola Akinoso
Department of Science and Technology Education,
Faculty of Education, University of Lagos, Nigeria
sakinoso@unilag.edu.ng