International Journal of Interdisciplinary Research Methods

Vol.4, No.3, pp.11-22, October 2017

Published by European Centre for Research Training and Development UK (www.eajournals.org)

APPLICATION OF ONE PARAMETER LATENT TRAIT THEORY IN THE CONSTRUCT TEST ITEMS VALIDITY OF MATHEMATICS

Iweka, Fidelis Ph.D
Department Of Educational Psychology, Guidance and Counselling
University Of Port Harcourt, Nigeria

ABSTRACT: *Various criticisms have been leveled against psychological testing. A large proportion of the criticism pivots on the construct validity of test items. This article discusses the procedures to adopts in validating Mathematics test Items. The design of the study is instrumentation. A multi – stage sampling technique was used to acquire a sample size of 200 students for the study. The instrumentation for the study was a self- developed 150 objective mathematics test items. The content validity was examined based on some experts' judgment on the development of the items. The analysis of data was based on win steps analysis. The result showed that the 86 items not only met the one parameter-latent trait model assumption of measurement construct but also demonstrated good psychometric properties.*

KEYWORDS: one parameter latent trait theory, construct validity, unidimentionality and classical test theory.

INTRODUCTION

Mathematics is an efficient tool used in all sciences and for technological development of any nation. Mathematics is a science of all sciences; it is a universal part of human culture and significant in our daily life. Mathematics provides us with a broad range of skills in problem solving, logical reasoning and flexible thinking (Jayanthi, 2014). Poor performance in mathematics by students at all levels has persisted over a long time, not only in Nigeria but most Africa countries (Thissan,1991). Some researchers attributed the poor academic performance in mathematics to fault inherent in a test (Onunkwo, 2002). One of the ultimate purposes in educational measurement is to estimate testee's ability in a particular subject. This measurement always involve numerical numbers to certain traits or characteristics using a tool for physical traits, such as height, the process of assigning numbers can be done directly using a ruler.

However, psychological traits such as ability or proficiency are constructs. They are unobservable but can be measured directly using a tool called test. The design of tests to measure constructs, however, presents several problems. Since the measurement of psychological constructs is always done indirectly, there is always the possibility that researchers will select different types of behaviour to measure the same construct. As a consequence, different inferences will be concluded. Lack of well defined units in the measurement scale also poses problem, (Iweka, 2014). For example, an examinee who is unable to answer any test item does not mean that he or she has zero ability. Instead, all the items have difficulty index which is more than the examinees ability. The study of measurement problems and methods to overcome them is known as test theory. Test theories relate observable traits (such as test score) with unobservable traits (such as ability or

Print ISSN: ISSN 2398-712X, Online ISSN: ISSN 2398-7138

proficiency) for a measured construct using mathematics model. A theory is a set of interrelated concept, definition and propositions that explain or predicts events or situation by specifying relations among variables (Iweka, 2017). Test model provides a general framework linking observable variables such as true scores and ability scores and also are formulated within the framework of a test theory and do specify in considerable details the relationship among a set of test theoretical concepts and their relationship. A good test theory or model can provide a frame of reference for doing test design or solving other practical problems involving items and ability scores.

In creating quality tests for assessing the student's performance, many indices have been developed in order to construct valid and reliable instrument. These indices rely mostly on the two test theories; classical test theory (CTT) and latent trait theory also known as item response theory (IRT). These two framework have widely been used in test development, to ensure quality of measuring instrument. Classical test theory (CTT) has been the foundation for measurement theory for decades and it is a theory about test scores that introduces three concepts (Iweka, 2017): test score (often called the observed score), the theory suggests that any assessment will only reveal an individual's observed score, and that, this is not always reflective of their true score as there is something in the environment that impacts on individuals performance (error). The model is given as X = T + E where; X is the observed score, T equals the true score and E, the error variance of the obtained score. One of the major limitations of the CTT is that the item statistics (the difficulty index, P – value) and (the discrimination index, r – values) which are very essential in the application of CTT are sample dependent. These limitations are addressed and overcome in IRT.

Latent trait theory otherwise known as the item response theory (IRT), strong true score theory or modern mental test theory evolved due to the weakness of CTT by providing a reporting scale on which examinees ability (the construct measured by the test) is independent of the particular choice of test items that are administered. The term latent is used to emphasize that discrete item responses are taken to be observable manifestation of hypothesized traits construct or attributes not directly observed, but which must be inferred from the manifest responses, hence it is called item response theory (IRT) due to its focus on the item as opposed to the test level of focus of classical test theory.

Latent trait model is seen as an improvement over CTT, it is more sophisticated and allows a researcher to improve the reliability of an assessment and has emphasized on three notions; a unidirectional trait denoted by θ, local independence of items and that the response of a person to an item can be modeled by a mathematical item, response function (IRF) or item characteristic curve (ICC). Latent trait theory is able to estimate the parameters of an item, independent of the characteristics of both test takers to which it is exposed and other items that constitute the test.
In IRT there exist different parameter models adjusting for different item properties leading to different ability estimation.

One-parameter model (also known as the Rash model) which adjust for item difficulty level as the trait level required for correctly answering a question. Two–parameter model (2p L) accounts for item difficulty and discrimination parameters. While three–parameter model (3PL) takes into

Print ISSN: ISSN 2398-712X, Online ISSN: ISSN 2398-7138

account the effect of item, guessing in addition to the difficulty and discrimination level of the item. This model assumes that the three parameters, difficulty, discrimination and guessing are combined for an estimate of a relationship between the probability of a correct response of an item and the trait level (ability) of an examinee.

However, one parameter model is the focus of the present study: Rash model has some special properties that make it attractive to users. It involves fewer parameter; therefore, it is easier to work with (Downing 2003). Critics of the one parameter model often regard the model as having strong assumptions that are difficult to meet. However, these are values that make the one parameter model more appropriate in practice. One major problem in measurement lies in the interaction between the person being measured and the instrument involved.

Performance of a person is known to be dependent on which instrument is used to measure his or her trait. However, this problem is circumvented by procedure of conjoint measurement in one parameter model. Iweka (2014) explained that in conjoint measurement, the unit of measurement is not the examinee or the item, but rather the performance of an examinee relative to a particular item. If Sn is an index for ability for examinee n on the trait being measured, and if θi is an index for the difficulty of the item i which relates to the trait being measured, then the unit of measurement is neither Sn nor θi but rather (Sn – θi) which is the difference between the ability of the examinee and the difficulty of the item. If the ability exceeds the item difficulty, then it is expected that the examinee will answer the item correctly. Conversely, if the difficulty exceeds ability, then it is expected that the examinee will answer incorrectly. In education, response on a particular item is always in uncertainties. Therefore, probabilistic approach has to be employed when explaining what happens when an examinee takes an item. Probabilities of correct response are between 0 and 1 and it does not permit proportion of correct answer to be expressed in internal scale. To overcome these problems, logistic transformation, which involves taking the natural logarithm is used.

As a final product, it can be shown that the probability of a person has correct response to item 1 as given in the formula:

P $(Yni = 1)\,\frac{e\ sn-\theta i}{1+e^{sn}-\theta 1}$ (Ahmed, 2012).

One parameter model offers procedure to transform test score into interval scale score in log-odd or logits unit. In order for the one parameter model measurement to have the examinee-free item difficulty and item-free examinee ability measurement two important assumptions must be met. Firstly, the data must meet the unidimensionality assumption, that is, they represent, a simple construct and secondly, one parameter model requires that the data must fit the model (Aridrich, 1999). It is also imperative to provide evidence on the psychometric properties of the test used from the framework of Rash model analysis.

According to Harwell, Baker and Zwarts (2003) two major threats to construct validity that are under investigation are construct-irrelevant variances that contaminate measurement of the main construct while in the later, the measurement fails to include important sub-dimensions of the construct. This implies that in construct validity, nothing should be left out while nothing irrelevant

Print ISSN: ISSN 2398-712X, Online ISSN: ISSN 2398-7138

or unimportant should be added. Jayanthi (2014) suggest that within the framework – irrelevant variance can be assessed by examining both dimensionality and fit of the measurement while significant gaps between the subsequent items provide indication of construct under-presentation.

Statement of problems

The most reliable means of assessing teaching and learning activities is by administering tests to the students: To maximize testing, one should aim to integrate all the major components of a course content, instruction, objectives, assessment and evaluation. Poor performance in mathematics by students at all levels has persisted over a long time and has been of great concern to the society at large. Therefore, there is an advocacy for new approach of analyzing test data. The statement of the problem therefore is to determine how suitable the development and construct validation of mathematics test items will be in determining students achievement in mathematics using one parameter model in order to solve the problem of lack of objectivity in students assessment which is inherent in the classical test theory method. Scope of the study: The study focused on using one parameter model to determine the construct validity of mathematics test items. The population of the study was made up of all S.S.2 students in Ogba-Egbema-Ndoni Local Government Area of Rivers State in Nigeria.

Purpose of the study: The aim of the study is to determine the construct validity of mathematics test using one parameter model. In specific terms, the study determined the psychometric characteristics of mathematics test items and examined the degree to which mathematics test items scores met the one parameter model expectations.

Research questions: The following research questions guided the study:

1. What are the estimates of the outfit and infit, indices of mathematics test items using one parameter model?
2. What are the reliability and validity coefficients of the mathematics test items using the one parameter model?

RESEARCH METHODOLOGY

The design of the study is instrumentation. Instrumentation research is a scientific investigation for meticulous development or construction of a test or measuring instrument that validity measures that concept or psychological construct, which it intends to measure with all accuracy (Kpolovie, 2010). The design is appropriate since the study involves construct validity of mathematics test items. A multi-stage sampling technique was used to acquire a sample size of 200 students for the study. Simple random sampling technique was used to select four schools from the local government area. Stratified random sampling was used to select 50 students each from each of the four schools to give the needed sample of 200 for the study.The instrument for this study was a self-developed mathematics test items. It is a 150 items test. The format is multiple-choice objectives with five (5) options lettered A-E. The item for the study is drawn from the senior secondary school (S.S.S.3) three syllabus. Effort was made to ensure that the topics for which the items were draw were those covered by the students in all schools selected. Test blue print was used to ensure content coverage.

Print ISSN: ISSN 2398-712X, Online ISSN: ISSN 2398-7138

Experts in the field of mathematics verified the instrument. These were done to ensure both content and face validity. Some items were deleted while some were reconstructed which led to the emergence of 150 items from the 200 items originally developed. They were administered to 50 student who were not part of the sample used for the trial testing of the instrument. Trial testing corrects, adjusts and revises the content of the test (Ojerinde, 2015), and test developers are majorly concerned about the quality of test items and how examinees respond to them. Item analysis was carried out on the responses of the trial testing to weed out the poor items using the xcalibre soft ware through marginal likelihood estimate for the difficulty and discrimination indices.

Items discrimination compares performance of upper group with high test scores and lower group (low test scores) on each item. The higher the value of D, the more adequately the item discriminates, (the highest value is 1.0) for exam with a normal distribution, a discrimination of 0.3 and above is good; 0.6 and above is very good. Attempts was made to select item with discrimination of 0.3 and above. Item difficulty is the percentage of test takers who respond correctly to a test item denoted by P (Ojerinde, 2013). An item with a P – value of 0.00 or 1.0 does not contribute to measuring individual differences and must be discarded.

Items with item difficulty of between 0.20 – 0.80 representing moderate difficulty were used for the study.

A Kuder Richardson 20 (KR_{20}) reliability technique was employed in testing the reliability coefficient of the instrument. The value obtained was 0.80. On the basis of the calculated reliability coefficient, the instrument was considered reliable for the study.

Method of Data Analysis

The students responses in the final test items from the sampled schools were prepared for the analyses using a Rasch model software, WINSTEPS version 3.75. In WINSTEPS, the measures are determined through iterative calibration of both person and item using the Mathematics Achievement test. In WINSTEPS, the outfit and the infit mean square provide indications of the discrepancies between the data and model's expectations. The range of acceptable fit for the study is between 0.7 and 1.3 for both fit indices (Bonds & Fax, 2001). The reliability and validity of the test items scores were determined by applying Rasch analysis for both item and examinees measures. A high reliability for both indices are required since they indicate a good replication if the comparable items/examinees are employed.

Print ISSN: ISSN 2398-712X, Online ISSN: ISSN 2398-7138

Statistical Analysis of Data

Table 1: Item Correlation Order

Entry N0.	Total Score	Total Count	Measure	Measure S.E	Infit		Out fit		Pt measure		Exatmatch		Item
					MS	2 STD	MS	2 STD	Core	Exp	Obs %	Exp%	
6	71	200	1.09	.15	1.03	.6	1.03	.7	.03	.13	65.5	64.6	10006
5	87	200	.75	.14	.99	-.3	.99	-.3	.16	.14	57.5	58.0	10005
10	102	200	.44	.14	1.00	-.2	1.00	-.2	.15	.14	60.0	55.5	10010
39	106	200	.36	.14	.99	-.3	.99	-.2	.16	.14	61.5	55.7	10039
45	110	200	.28	.14	1.10	4.3	1.11	4.4	-.24	.13	48.0	56.6	10045
37	114	200	.20	.14	1.06	2.0	1.06	1.9	-.06	.13	50.0	57.8	10037
2	132	200	-.19	.15	1.03	.6	1.03	.6	.03	.13	66.5	66.0	10002
7	51	200	1.58	.16	1.05	.6	1.08	.9	-.07	.12	74.5	74.5	10007
99	147	200	-.55	.16	1.02	.3	1.04	.5	.02	.12	73.5	73.5	10099
89	155	200	-.77	.17	1.01	.2	1.03	.3	.06	.11	77.5	77.5	10089
43	96	200	.57	.14	1.02	.9	1.02	1.0	.07	.14	55.5	55.8	10043
82	148	199	-.60	.16	1.01	.2	1.02	.3	.07	.12	74.4	74.4	10082
48	128	200	-.10	.15	1.01	.4	1.02	.4	.07	.13	64.0	64.0	10048
22	72	200	1.07	.15	1.01	.2	1.01	.2	.10	.13	64.0	64.2	10022
88	137	199	-.32	.15	1.00	.0	1.01	.1	.11	.12	68.8	68.8	10088
51	178	200	-1.63	.23	.99	.0	.97	-.1	.12	.08	89.0	89.0	10051
17	78	200	.94	.15	1.05	1.3	1.05	1.4	-.04	.13	61.0	61.5	10017
50	96	200	.57	.14	1.04	2.0	1.04	2.0	-.01	.14	49.5	55.8	10050
26	65	200	1.23	.15	1.03	.6	1.04	.7	.01	.13	67.0	67.5	10026

Print ISSN: ISSN 2398-712X, Online ISSN: ISSN 2398-7138

100	154	200	-.74	.17	1.01	.1	1.02	.3	.08	.11	77.0	77.0	10100
97	145	199	-.52	.16	1.00	.1	1.01	.2	.10	.12	72.9	72.9	10097
66	150	200	-.63	.16	.99	-.1	.97	-.3-	.16	.12	75.0	75.0	10066
33	86	200	.77	.14	.93	-2.7	.92	-2.7	.40	.14	65.0	58.4	10033
3	79	200	.92	.15	.98	-.6	.98	-.6	.21	.13	64.5	61.1	10003
40	64	200	1.25	.15	.96	-.6	-96	.6	.26	.13	68.0	68.0	10040
11	124	200	-.01	.15	.96	-1.0	.96	-1.0	.27	.13	63.0	62.1	10011
79	132	200	-.19	.15	.98	-.3	.97	-.5	.20	.13	66.5	66.0	10079
27	64	200	1.25	.15	1.05	.9	1.06	1.0	-.05	.13	68.0	68.0	10027
94	142	200	-.43	.16	1.01	.2	1.00	.1	.09	.12	71.0	71.0	10094
60	146	199	-.55	.16	1.01	.1	1.01	.1	.09	.12	73.4	73.4	10060
85	140	200	-.38	.16	1.00	.0	1.00	.0	.13	.12	70.0	70.0	10085
90	148	200	-.58	.16	1.00	.0	1.00	.1	.11	.12	74.0	74.0	10090
53	174	200	-1.44	.21	.99	.0	.99	.0	.12	.09	87.0	87.0	10053
21	96	200	.57	.14	1.00	.0	1.00	.0	.14	.14	55.5	55.8	10021
8	74	200	1.03	.15	1.00	.0	1.00	.0	.13	.13	62.5	63.2	10008
55	166	200	-1.12	.19	1.00	.1	1.00	.0	.09	.10	83.0	83.0	10055
78	158	200	-.86	.17	1.01	.2	1.02	.2	.05	.11	79.0	79.0	10078
74	147	200	-.55	.16	.98	-.2	.97	-.4	.19	.12	73.5	73.5	10074
19	73	200	1.05	.15	.98	-.4	.99	-.3	.19	.13	65.0	63.7	10019
18	93	200	.63	.14	.99	-.5	.99	-.5	.18	.14	61.0	56.4	10018
71	158	200	-.86	.17	.98	-.2	.96	-.4	.20	.11	79.0	79.0	10071
34	72	200	1.07	.15	.97	-.7	.96	-.7	.25	.13	65.0	64.2	10034
38	82	200	.86	.15	.97	-1.1	.96	-1.1	.26	.14	61.0	59.9	10038
46	88	200	.73	.14	.95	-19	.95	-1.9	.31	.14	61.5	57.7	10046
41	88	200	.73	.14	.94	-2.5	.94	-2.5	.36	.14	65.5	57.7	10041
13	102	200	.44	.14	.99	-.6	.99	-.6	.18	.14	59.0	55.5	10013
1	95	200	.59	.14	.99	-.6	.99	-.7	.18	.14	55.0	55.9	10001
65	157	199	-.86	.17	.98	-.1	.97	-.3	.18	.11	78.9	78.9	10065
67	158	200	-.86	.17	.97	-.2	.94	-.5	.23	.11	79.0	79.0	10067
4	98	200	.53	.14	.97	-1.5	.97	-1.5	.24	.14	58.5	55.5	10004

Print ISSN: ISSN 2398-712X, Online ISSN: ISSN 2398-7138

76	140	199	-.40	.16	.96	-.6	.94	-.9	.28	.12	70.4	70.3	10076
70	146	200	-.53	.16	.97	-.4	.95	-.6	.24	.12	73.0	73.0	10070
84	144	200	-.48	.16	.99	-.2	.98	-.3	.17	.12	72.0	72.0	10084
80	138	200	-.33	.15	.99	-.2	.98	-.3	.17	.12	69.0	69.0	10080
30	91	200	.67	.14	.99	-.4	.99	-.4	.17	.14	60.0	56.8	10030
14	88	200	.73	.14	.99	-.3	.99	-.4	.17	.14	56.5	57.7	10014
72	148	200	-.58	.16	.97	-.3	.95	-.6	.23	.12	74.0	74.0	10072
75	148	200	-.58	.16	.97	-.5	.96	-.4	.21	.12	74.0	74.0	10075
83	125	200	-.03	.15	.98	-.5	.97	-.6	.21	.13	62.5	62.6	10083
77	148	200	-.58	.16	.99	-.1	.97	-.3	.17	.12	74.0	74.0	10077
57	152	200	-.69	.17	.99	-.1	.97	-.3	.17	.11	76.0	76.0	10057
20	127	200	-.08	.15	.97	-.6	.97	-.7	.23	.13	63.5	63.6	10020
44	99	200	.51	.14	.98	-1.2	.98	-1.2	.22	.14	59.0	55.5	10044
9	85	200	.79	.14	.98	-.6	.98	-.7	.20	.14	59.5	58.7	10009
56	155	199	-.80	.17	.97	-.2	.95	-.4	.22	.11	77.9	77.9	10056
63	147	200	-.55	.16	.99	-.2	.97	-.3	.18	.12	73.5	73.5	10063
12	92	200	.65	.14	.99	-.5	.99	-.5	.18	.14	60.5	56.6	10012
42	52	200	1.55	.16	.98	-.2	.98	-.2	.18	.12	74.0	74.0	10042
68	141	199	-.42	.16	.98	-.2	.97	-.4	.19	.12	70.9	70.8	10068
54	170	200	-1.27	.20	.98	-.1	.96	-.3	.19	.09	85.0	85.0	10054
47	100	200	.49	.14	.98	-1.1	.98	-1.1	.21	.14	61.0	55.5	10047
62	140	200	-.38	.16	.98	-.3	.97	-.4	.20	.12	70.0	70.0	10062
64	143	200	-.45	.16	.97	-.5	.95	-.7	.25	.12	71.5	71.5	10064
29	170	200	-1.27	.20	1.02	.2	1.07	.5	-.02	.09	85.0	85.0	10029
25	97	200	.55	.14	1.05	2.3	1.05	2.3	-.03	.14	50.0	55.6	10025
23	103	200	.42	.14	1.05	2.6	1.05	2.5	-.05	.14	48.5	55.5	10023
96	155	200	-.77	.17	1.04	.4	1.07	.7	-.05	.11	77.5	77.5	10096
31	118	200	.11	.15	1.03	.9	1.03	.9	.03	.13	57.5	59.3	10031
24	70	200	1.12	.15	1.02	.4	1.02	.5	.06	.13	64.0	65.1	10024
78	158	200	-.86	.17	1.01	.2	1.02	.2	.05	.11	79.0	79.0	10078
98	142	200	-.43	.16	1.00	.1	1.00	.1	.11	.12	71.0	71.0	10098
81	139	200	-.35	.15	1.00	.0	.99	-.1	.14	.12	69.5	69.5	10081
58	151	199	-.68	.17	.99	.0	.99	-.1	.14	.11	75.9	75.9	10058

Print ISSN: ISSN 2398-712X, Online ISSN: ISSN 2398-7138

69	160	200	-.92	.18	.99	.0	.99	-.1	.13	.11	80.0	80.0	10069
95	131	200	-.17	.15	1.00	.0	1.00	0.1	.13	.13	66.5	65.5	10095
73	146	199	-.55	.16	1.00	.0	1.00	.0	.12	.12	73.4	73.4	10073
87	129	200	-.12	.15	1.00	.0	1.0	1.1	.12	.13	64.5	64.5	10087
35	79	200	.92	.15	1.01	.3	1.01	.3	.10	.13	59.5	61.1	10035
97	145	199	-.52	.16	1.00	.1	1.01	.2	.10	.12	72.9	72.9	10097
59	144	200	-.48	.16	1.01	.2	1.00	.1	.09	.12	72.0	72.0	10059
92	148	200	-.58	.16	1.02	.3	1.02	.3	.04	.12	74.0	74.0	10092
86	144	200	-.48	.16	1.01	.2	1.02	.3	.07	.12	72.0	72.0	10086
93	150	200	-.63	.16	1.01	.1	1.03	.4	.07	.12	75.0	75.0	10093
16	98	200	.53	.14	1.04	1.8	1.04	1.8	.01	.14	51.5	55.5	10016
49	81	200	.88	.15	1.04	1.2	1.04	1.3	-.01	.13	57.5	60.3	10049
15	107	200	.34	.14	1.03	1.6	1.03	1.5	.01	.14	50.0	55.9	10015
28	72	200	1.07	.15	1.02	.6	1.03	.6	.04	.13	65.0	64.2	10028
61	155	200	-.77	.17	1.01	.1	1.01	.1	.09	.11	77.5	77.5	10061
91	148	200	-.58	.16	1.00	.0	1.01	.1	.11	.12	74.0	74.0	10091
52	176	200	-1.53	.22	1.00	.1	1.02	.2	.06	.09	88.0	88.0	10052
32	143	200	-.45	.16	1.05	.7	1.07	.9	-.06	.12	71.5	71.5	10032
Mean	120	.8	199	.9	.00	.16	1.00	.0	1.00	.0	67.7	67.7	
SD	33	.2		.3	.76	.02	.03	.9	.03	.9	9.3	8.9	

Table 1 is used to answer the research question 1. The outfit and infit columns for both MNSQ and ZSTD showed the indices. From the table, it will be discovered that item 7 with an estimated difficulty index of 1.58logits and standard error of 0.16 is the most difficult item in the test. Also item 51 with – 1.63 logits and standard error of 0,23 is the easiest item. It could be observed also

Print ISSN: ISSN 2398-712X, Online ISSN: ISSN 2398-7138

that out of the 100 items that fitted the IRT model, 14 items which are 45, 7, 37, 32, 27, 23, 96, 17, 25, 29, 49, 50, 41, and33 were classified as poor test items that should be omitted, deleted or revised because of lack of fit to the model. These items are measuring something other than the intended content and construct. They are construct irrelevant. Ten (10) items were classified as fairly good test items which could be revised or improved. There were 76 items that model assumption which is an indication of undimensionality of mathematics test items.
Therefore, these 76 items that fitted IPLM out of the 100 items were used to constitute the mathematics items as the construct qualities of the MTI.

Table 2: Person to Measure Correlation Summary

	INFIT		OUTFIT		TOTAL SCORE	MEASURE	COUNT	MODEL ERROR
	MNSQ	ZSTD	MNSQ	ZSTD				
MEAN	1.00	.0	1.00	.0	60.4	.49	99.9	.22
SD	.13	1.3	.13	1.3	5.7	.28	.7	.01
MAX	1.30	3.4	1.38	3.3	75.0	1.24	100.0	.24
MIN	.70	-4.0	.64	-3.9	45.0	-.23	90.0	.21

Real RMSE .22 True SD .16 SEPARATION .73 Person reliability .55
MODEL RMSE .22 TRUE SD .17 SEPARATION .77 Person reliability .57
S.E of PERSON MEAN = .02
Person raw score test reliability (cronback alpha'ssocre) = .55
The validity and reliability of the mathematics test items (MTI) using one parameter latent trait theory was estimated from the data in table 2.

The separation index of the persons is 0.73 which translates to a person strata index of 3.4. The strata index shows the number of distinct ability levels which can be identified by the test. The minimum person strata index is 2 which means that the test is able to distinguish between at least 2 strata of persons namely, high ability and low ability persons. A reliability index of at least 0.50 is required for a separation index of 1. The moderate reliability, separation and strata indices for this test are as a result of the low standard deviation of the person abilities (Iweka, 2014).
The crobach alpha (KR-20) person raw score test reliability of 0.55 was moderate, indicating that it was likely that the ordering of the examinees ability can be replicated since most of the variance was attributed to true variance of the mathematics test items.

Print ISSN: ISSN 2398-712X, Online ISSN: ISSN 2398-7138

Table 3: Summary of measured items for separation index.

	TOTAL SCORE	INFIT		OUTFIT		COUNT	MEASURE	MODEL ERROR
		MNSQ	ZSTD	MNSQ	ZSTD			
MEAN	120.8	1.00	.0	1.00	.0	199.9	.00	.16
SD	33.2	.03	.9	.03	.9	.3	.76	.02
MAX	178.0	1.10	4.3	1.11	4.4	200.0	1.58	.23
MIN	51.0	.93	-2.7	.92	-2.7	199.0	-1.63	.14

REAL RMSE .16 TRUE SD .74 SEPARATION 4.68 ITEM RELIABILITY .96
MODE RMSE .16 TRUE SD .74 SEPARATION 4.71 ITEM RELIABILITY .96
S.E of item mean 5.8

To investigate the representativeness of the test items, table 3 is used for checking the separation index. The separation shows the spread of the items along the variable void of gaps and targeted to person ability.The minimum index for item separation and item strata is 2. Therefore, the separation value for this test is 4.68. The item reliability 0.96 is a very good one which shows that the items are very reliable for administration. There is a very wide spread of difficulty in the items as the standard deviation of item difficulty estimates is 0.76 logits and the separation is 4.68. Consequently, we can rely on the representativeness of the test items.

DISCUSSION OF FINDINGS AND RESULTS

The one parameter latent trait model analyses as presented in table 1 indicated that both means of infit MNSQ and outfit MNSQ values were close to the expected value of 1.00. The individual item values showed that infit MNSQ values ranged from 0.90 to 1.11 while outfit MNSQ values ranged from -1.00 to 1.20. The findings of the study proved that the scores demonstrated little variation from model expectation which is that there was evidence of consistency between the examinees' responses and items on the scale and the models expectations and the unidimensionality assumption of the construct validity was met. The reliability of item difficulty measures were 0.96, which was high, suggesting that the ordering of item difficulty was replicable with other comparable sample of examinees. Consequent upon the findings, threat regarding construct irrelevant – variance was minimum based on the dimensionality test as well as the within range fit indices. Summarily, there were 86 items that fit the one parameter latent trait model in the construct validation of mathematics test items with an indication of undimensionality.

RECOMMENDATIONS AND CONCLUSION

The findings from this study reveal that the Mathematics test items exhibited few negative point–measure correlations and has very few misfitting items. The test did exhibit a fairly low mean score, although, test-takers' abilities were nonetheless, reasonably well spread across items.
The limitation of the study, while attempting to provide validity evidence, did not include such analyses as, differential item or test functioning, unexpected response or item distracter analyses or person-item-map. These should most certainly be explored in more details to determine if there

Print ISSN: ISSN 2398-712X, Online ISSN: ISSN 2398-7138

are any items that are causing unexpected response patterns either across groups or across sections of the test. The use of one parameter latent trait theory (Rasch model) offers opportunity to deal with core measurement issues such as construct validity as well as providing richer interpretation regarding examinee performance. Theoretically, this study has added more evidence in favour of the one parameter latent trait theory as having the capacity to resolve some of the rudimentary issues in measurement. However, in order for construct validity to hold, the theory requires more evidence. Test developers would have to have a thorough understanding of the measured construct. This one parameter atent trait model analysis has provided useful information which not only can be used for future developments, modification and monitoring achievement assessments, but also for establishing a process of validating pedagogical assessment.

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, vol. 2 No. 3.

Aridrich, D. (1999). "Rasch Model as a coustruct validation tool," in Rasch Measurment Transaction, vol. 22 (1) pp. 1145-1146.

Bond, T.G., & Fox, C.M. (2001). Applying the Rasch Model: Fundamental Measurement in Human Science, 1st ed, Mahwah, N.J: Lawrence Eribaum.

Dowing, S.M. (2003) "Item response theory: Application of Modern test Theory" Medical Education, Vol 37, pp. 739-745.

Harwell, O.C; Baker, Z. & Zwarts P. (2003). A Rasch hierarchical measurement model. Journal of Educational and behavioural statistics, 26, 307-331.

Iweka Fidelis, (2017). Effects of Authentic and Jigsaw II, learning techniques on students academic achievement in Mathematics. Global Journal of Arts, Humanities and social sciences, vol 5, 2, pp18-24.

Iweka, F. (2014). Comprehensive Guide to test Construction and administration. Omoku: chifas.

Iweka, F.O.E. and Wokoma T.A. (2017) Attitudes of Teachers towards application of item response theory in technical colleges in Rivers State. British Journal of Education vol. 5 (6) pp 39-56.

Jayanthi, E.C. (2014). The application of an unfolding model of the PIRT type to measurement of attitude. Applied psychological measurement, vol.12 pp. 33-50.

Kpolovie, P.J. (2010). Advanced research methods. New owerri: Springfield publishers ltd.

Ojerinde, D. & Ifewulu, C.B. (2013). Item unidimensionality using 2010 UTME Mathematics, a paper presented at IAEA Conference in Astanakasakstan.

Ojerinde, D. (2015). Classical test Theory (CTT) vs item Response theory (IRT): An evaluation of the comparability of item analysis results: Lecture presentation, institute of Education; University of Ibadan.

Onunkwo, G.I.N. (2002). Fundamentals of educational measurement and evaluation. Owerri Onitsha: cape publishers international.

Thissan, S.E. (1991). Item response theory for psychologists, N.J: Lawrence-Eribaum.

Print ISSN: ISSN 2398-712X, Online ISSN: ISSN 2398-7138