IOSR Journal of Research & Method in Education (IOSR-JRME)
e- ISSN: 2320–7388,p-ISSN: 2320–737X Volume 8, Issue 2 Ver. VI (Mar. – Apr. 2018), PP 01-07
www.iosrjournals.org

Estimating Standard Errors of Irtparameters of Mathematics Achievement Test Using Three Parameter Model

Ogbonna, Joseph U[1], Opara, Ijeoma M[2],
[1,2]Department of Educational Psychology, Guidance and counseling, Faculty of Education, University of Port Harcourt, Rivers State, Nigeria.
Corresponding Author: Ogbonna, Joseph U

Abstract: *This study focused on the estimation of standard error of estimation using 3plm of item response theory of test items of achievement test in junior secondary school mathematics in Rivers state of Nigeria. The design for the study was instrumentation research design. The instrument for data collection was 50 itemed mathematics achievement test designed by the researcher applying irt procedures . A pilot study instrument of 100 MAT items was administered to 200 examinees using 2plm to obtain SEEs and item parameters, the discrimination and difficulty indices , these were used to select the items that constitute the instrument for the study. A total sample of 2000 examinees was used for the study. Marginal maximum likelihood estimation of Ex-calibre 4.2 was used to estimate the item's parameters and standard errors . The finding shows that the standard errors of the discrimination parameters (SEE_a) ranges from 0.08 of item 37 to 0.115 of item 1, the difficulty index (SEE_b) ranges from 0.030 of items 37 and 39 to 0.103 of item and the guessing parameter (SEE_c) ranges from 0.039 of item 25 and 32 to 0.076 of item 1. It is recommended that standard error of estimation should be estimated and used to validate the quality of items to be retained in a test.*

Keywords: *Standard errors, IRT, item parameters, and mathematics.*

Date of Submission: 09-04-2018 Date of acceptance: 23-04-2018

I. Introduction

Mathematics as a subject occupies a crucial role in the economic, technological development of any nation. The importance of it had made it compulsory subject in both junior and senior secondary schools and as a key in studying other sciences like physics, chemistry engineering etc, for it provides wide range of skills in problem solving and logical reasoning.Githua and Mwangi (2003) opined that life without mathematics is an almost impossibility and that it would be difficult to live a normal life in very many parts of the world without it.

Achievement in mathematics becomes very crucial in determining the ability of a student choosing specific career or vocation. However, mathematics is important in our daily and practically every human activity is seasoned with the application of mathematical skills. The need for high quality professional development programme in mathematics and science have become increasingly important in the current climate of educational reform (Blank, Alas & Smith, 2007)Achievement in mathematics by secondary school students are usually affected by several factors including learning styles, personality type, lack of confidence in the subject among others, but in most cases, factors inherent in the measurement instrument or the test items may affect the performance of the examinee in a given examination. Therefore,to an instrument developer,constructing or developing a mathematics achievement instrument demands a meticulous approach to produce a valid and reliable instrument. A quality testing instrument is function of careful item writing and validation to control errors in the test psychometric properties.

A quality testing is a panacea to the issues of measurement as it provides enough evidence to accurately make decision and improve educational practice and efficiency. Testing procedures and tools should be reliable within the framework of test theory or model of application. Testing procedures determine the credibility of the scores provided by the items and validly aids in estimating the abilities of the test takers.

Test developers or designer in their art of measurement tool development should be careful to ensure that items so developed are relevant to the construct under consideration, thus proper management of errors inherent in items becomes necessary.

Test are developed under two main framework, classical test theory (CTT) and item response theory (IRT) also known as latent trait theory (LTT). Maximizing the benefit of each theory demands that test items be developed in line with the associated models and procedures. Test models provide a comprehensive framework linking observable variables such as true scores and ability scores.

CTT is the oldest and widely used, it assumes equal contribution of items to the overall score of an examiner, the major aim is to understand and improve reliability. Reliability is obtained by correlating scores from two parallel tests, when error variance is small, reliability measure increases, nevertheless, the item statistics considered in CTT are difficulty and discrimination.

The major criticism leveled against CTT is that its reliability is group specific, scores are test specific and standard error of measurement for a fixed set of items is constant and therefore it assumes that error is the same for everybody and hence lacks predictability.

Initem response theory framework which is modern and an improvement on CTT, IRT is able to estimate item parameters independent of test takers characteristics. IRT is concerned with accurate test scoring and development (Xinningand Yung, 2014). IRT specify relationship between observed response and underlying unobservable construct, provides a means to evaluate scores on the ability and hence assumes that performance of an examinee can be totally be predicted or explained from one or more abilities (Hambleton and Swaminaltan 1985).

The standard error considers the variability of all means from samples of same population and provides a way to measure the average distance between a sample mean and a population mean. In order words the standard error gives researchers an indication of how accurate their sample data represents their intended population (Agresti& Finlay, 1997)

CTT standard error of measurement (SEM) is used to produce confidence internal and it is how much an estimate of how much error there is in a test (Obinne, 2011). Test reliability provide a basis for the measurement of the validity of a test and consequently assess the standard error of measurement. In IRT concept of standard error of estimation is adopted similar to the reliability concept in CTT.

Standard errors of estimation(SEE) emphasis on how confident we can be at each ability level (McAlphine 2002). The concept (SEEs) is derived from the item information. Itemphasizes how much we can learn about the latent trait from an item and each item has item information curve. In three-parameter latent trait (3plm) each item is described by three parameters (difficulty, discrimination and guessing) and each item parameter has its own standard error of estimate. SEE is a measure of precision of an item parameter (Thissen&Wainer ,1982). A smaller SEE indicates greater precision hence more information provided by the item. Item parameter SEEs are very essential components in IRT framework used in the area of differential item functioning (DIF), testing item parameter drift and in the determination of retention and rejecton of test item. One criteria for rejecting an item is when an item's difficulty SEE is equal or greater than a predetermined value (Toland 2008). In order words, items are retained if their item difficulty SEE is less than one (1). There, the accuracy of parameter SEEs are essential however , it is affected by different testing conditions as test length and examinee sample size. Several IRT software have been in use to determine to the item parameters and its SEEs. This study adopted marginal maximum likelihood estimation procedure in ex- caliber 4.2 for dichotomously scored responses

A high information value predicts small standard error of measurement. Item information in IRT is based on fisher information matrix. Test information is the aggregate of item information values across all the items in the scale (Lord, 1980).

Item information is maximized when difficulty level approaches the theta (ability level) and when the discrimination index increases as the guessing parameter tends to zero for 3 parameter IRT model. In addition, the higher the discrimination of an item the more information provided. (Mcaphine 2002) while the higher the c_i the lack information provided by the item

The standard errors of estimation is calculated b

$$SE(\hat{\theta}) = \frac{1}{\sqrt{\sum_{i=i} I(\theta)}}$$

While SE (θ) = the standard error estimation at ability level θ

$\sum_{i=i}(\theta)$ = sum of the item information ability level. θ for all items in the test.

As test information increases, the standard error of estimation decreases.The errors in IRT are more complicated and are connected to the maximum likelihood estimation statistical procedures used in estimating item parameters and examinee's ability. Estimating item parameters in 2plm and 3plm of IRT, problem encountered could be resolved through pre-specifying the expected item and ability distribution. Ranking problem encountered in the attempt to rank items according to difficulty order are usually altered with the involvement of discrimination index and item difficulty are ability dependent

Sincerely introducing the guessing parameter changes the interpretation of other parameters. The threshold parameter b is the value of theta at which respondent have at (0.5- 0.50) x 100% chance of responding correctly to the item. The aim of this study is to estimate the standard error of IRTparameters of 50 itemed mathematics achievement test constructed by the researchers meant for junior school examinees.

The objectives of the study are:

1. To find items parameters of a 50 itemed mathematics achievement test.
2. To find the standard error of estimation of the items parameters- difficulty, discrimination and guessing indices.
3. To find good/ fairly good items based on their SEEs

The following research questions guided the study.

1. What are the items parameters of a 50 itemed mathematics achievement test?
2. What are the standard error of estimation (SEEs) of the item parameters (SEEa, SEEb ,SEEc)?
3. What are the SEEs of items selected as good/fairly good items

II. Methodology

The design for this investigation is instrumentation research design for it involves the development of a research instrument in mathematics using IRT procedures.A pilot study instrument of 100 itemed multiple-choicedmathematics achievement test based on item response theory framework or procedures was validated by experts in the field of mathematics. The test itemswere drawn from Junior Secondary School syllabus in Rivers State. This was administered to two hundred (200) examinees (100 boys and 100 girls) drawn from the target population of all male and female JSS3 Students in Rivers State who sat for the 2017 JSCE. Their responses were analysed using marginal maximum likelihood estimation of the ex-calibre 4.2 developed by Harwell, Baker and Zwarts (1997) for both item parameters and standard error estimations (SEEs) of the parameters using the 2pl trait model. In this pilot study guessing parameter (c= 0) was assumed same for 2 pl. Eighty-nine (89) items from the 100 items fitted the 2pl model, 50 items from the 89 items were selected based on the general rule for selecting good and fairly good items. This 50 items constituted the final version of the instrument.

The 50 itemed mathematics achievement test (final version) was administered to a sample of two thousand (2000) students (1000 boys and1000 girls) drawn from the 36,233 JSS 3 students population in Rivers State of Nigeria from 10 LGAs of the state through multi-stage sampling procedures. Direct administration was adopted in testing students and scripts collection.

The students response was collected, prepared and items calibrated using the marginal maximum livelihood estimation of ex-calibre 4.2 software developed by Harwell, Baker &. Zwarts(1995).The excalibre software is modern and easy to use for logistic models for binary response based on 3pl model.

III. Results

Research question 1: What are the item parameters of a 50 itemed mathematics achievement test?
Table 1 displays items parameters of the 50 itemed MAT administered to 2000 examinees using the 3 pL model. It presents the item parameters, and any flags for each calibrated item. The F flag indicates that the item fit statistic (z Resid for dichotomous / chi-square for polytomous) was significant, and the item did not fit the IRT model.

Table 1: *Item Parameters for All Calibrated mathematics Items*

Seq.	Item ID	P	R	a	B	C	Flag(s)
1	1	0.695	0.169	0.411	-0.251	0.352	
2	2	0.606	0.237	0.621	0.306	0.327	
3	3	0.367	0.314	1.416	1.144	0.277	
4	4	0.581	0.208	0.533	0.278	0.260	
5	5	0.323	0.371	1.219	1.046	0.190	
6	6	0.431	0.285	0.981	0.953	0.266	
7	7	0.302	0.398	1.401	1.037	0.179	
8	8	0.399	0.300	1.603	1.034	0.284	
9	9	0.442	0.314	1.044	0.838	0.262	
10	10	0.340	0.351	1.371	1.057	0.219	
11	11	0.394	0.309	1.522	0.995	0.270	
12	12	0.297	0.436	2.155	0.979	0.189	
13	13	0.268	0.442	1.953	1.042	0.168	
14	14	0.284	0.366	1.640	1.153	0.191	
15	15	0.368	0.278	1.300	1.168	0.259	
16	16	0.339	0.350	1.467	1.022	0.216	
17	17	0.272	0.415	1.649	1.080	0.169	
18	18	0.312	0.402	1.740	0.992	0.196	
19	19	0.368	0.349	1.212	0.954	0.220	
20	20	0.340	0.359	1.522	1.020	0.220	

21	21	0.292	0.297	1.681	1.263	0.215	
22	22	0.345	0.423	1.768	0.875	0.207	
23	23	0.370	0.343	1.298	0.961	0.228	
24	24	0.365	0.298	1.414	1.097	0.253	
25	25	0.250	0.381	1.911	1.172	0.170	
26	26	0.327	0.397	1.559	0.969	0.199	
27	27	0.391	0.347	1.420	0.919	0.249	
28	28	0.271	0.447	1.711	1.010	0.158	
29	29	0.439	0.387	1.290	0.662	0.238	
30	30	0.375	0.356	1.392	0.925	0.231	
31	31	0.355	0.438	1.633	0.820	0.201	
32	32	0.250	0.454	1.575	1.021	0.136	F
33	33	0.381	0.417	1.767	0.762	0.220	
34	34	0.328	0.338	1.343	1.070	0.208	
35	35	0.346	0.417	1.909	0.857	0.208	
36	36	0.330	0.414	1.642	0.894	0.191	
37	37	0.288	0.502	1.920	0.834	0.147	F
38	38	0.321	0.437	1.720	0.870	0.180	
39	39	0.271	0.483	2.095	0.928	0.154	F
40	40	0.369	0.392	1.466	0.853	0.216	
41	41	0.357	0.374	1.945	0.928	0.234	
42	42	0.416	0.351	2.130	0.849	0.284	F
43	43	0.380	0.348	1.290	0.905	0.227	
44	44	0.346	0.399	1.570	0.873	0.201	
45	45	0.348	0.437	1.768	0.799	0.194	
46	46	0.306	0.430	1.534	0.896	0.164	
47	47	0.333	0.458	1.723	0.787	0.174	
48	48	0.338	0.419	1.492	0.847	0.183	
49	49	0.359	0.377	1.301	0.856	0.196	
50	50	0.375	0.396	1.510	0.805	0.214	

In table 1 above,items 32, 37, 39 and 42 were flagged which indicates that the item fit statistic (z Resid for dichotomous) was significant, and the item did not fit the IRT model. Table 1 also showed the item parameters, the difficulty index b ranges from -0.251 of item 1 to 1.26 of item 21, the Discrimination index a ranges from 0.411 of item1 to 2.130 of item 42. It could be observed that the items of the test discriminated well between the high ability and the low ability examinees. Forty-six out of fifty items have discrimination index greater than one (a > 1) while the guessing index c lowest is 0.136 of item 32 and the highest is 0.352 of item 1.Research question 2: What are the standard errors of estimation of the parameters (SEEa, SEEb, SEEc)?

Table 2: Standard errors of estimation of the item parameter of the 50- items MAT using the 3 pL model.

Item	*A*	*B*	*c*	*a SE*	*b SE*	*c SE*	*Chi-sq*	*df*	*p*	*z Resid*	*P*
1	0.411	-0.251	0.352	0.115	0.103	0.076	251.554	12	0	1.319	0.187
2	0.621	0.306	0.327	0.111	0.072	0.067	127.987	12	0	0.753	0.452
3	1.416	1.144	0.277	0.103	0.051	0.048	16.517	12	0.169	0.748	0.454
4	0.533	0.278	0.26	0.109	0.076	0.066	168.678	12	0	1.487	0.137
5	1.219	1.046	0.19	0.084	0.046	0.046	27.195	12	0.007	0.98	0.327
6	0.981	0.953	0.266	0.093	0.056	0.053	41.553	12	0	0.772	0.44
7	1.401	1.037	0.179	0.084	0.042	0.044	17.911	12	0.118	1.015	0.31
8	1.603	1.034	0.284	0.104	0.045	0.049	20.389	12	0.06	0.84	0.401
9	1.044	0.838	0.262	0.093	0.051	0.053	32.257	12	0.001	1.022	0.307
10	1.371	1.057	0.219	0.09	0.046	0.046	22.188	12	0.035	0.605	0.545
11	1.522	0.995	0.27	0.099	0.044	0.049	10.296	12	0.59	0.574	0.566
12	2.155	0.979	0.189	0.095	0.032	0.042	24.182	12	0.019	0.762	0.446
13	1.953	1.042	0.168	0.091	0.035	0.04	10.397	12	0.581	1.489	0.136
14	1.64	1.153	0.191	0.094	0.043	0.042	10.958	12	0.533	0.892	0.372
15	1.3	1.168	0.259	0.098	0.053	0.048	14.758	12	0.255	0.481	0.63
16	1.467	1.022	0.216	0.091	0.043	0.046	25.124	12	0.014	0.605	0.545
17	1.649	1.08	0.169	0.088	0.039	0.041	28.392	12	0.005	1.343	0.179
18	1.74	0.992	0.196	0.091	0.037	0.044	20.736	12	0.054	1.094	0.274
19	1.212	0.954	0.22	0.087	0.046	0.048	18.994	12	0.089	0.485	0.627
20	1.522	1.02	0.22	0.092	0.042	0.046	18.14	12	0.111	0.778	0.436
21	1.681	1.263	0.215	0.103	0.048	0.043	27.144	12	0.007	0.609	0.543

22	1.768	0.875	0.207	0.089	0.035	0.046	28.306	12	0.005	1.075	0.282
23	1.298	0.961	0.228	0.09	0.045	0.048	7.745	12	0.805	0.401	0.688
24	1.414	1.097	0.253	0.098	0.048	0.047	15.04	12	0.239	0.521	0.603
25	1.911	1.172	0.17	0.097	0.039	0.039	30.895	12	0.002	1.351	0.177
26	1.559	0.969	0.199	0.088	0.039	0.045	21.636	12	0.042	0.521	0.602
27	1.42	0.919	0.249	0.093	0.043	0.049	32.363	12	0.001	0.546	0.585
28	1.711	1.01	0.158	0.085	0.036	0.041	28.103	12	0.005	1.718	0.086
29	1.29	0.662	0.238	0.086	0.04	0.053	49.719	12	0	0.673	0.501
30	1.392	0.925	0.231	0.09	0.043	0.048	7.235	12	0.842	0.379	0.705
31	1.633	0.82	0.201	0.085	0.035	0.047	32.429	12	0.001	0.994	0.32
32	1.575	1.021	0.136	0.079	0.037	0.039	55.601	12	0	2.289	0.022
33	1.767	0.762	0.22	0.087	0.033	0.048	18.756	12	0.095	1.347	0.178
34	1.343	1.07	0.208	0.089	0.046	0.046	23.1	12	0.027	1.279	0.201
35	1.909	0.857	0.208	0.089	0.033	0.045	24.227	12	0.019	1.483	0.138
36	1.642	0.894	0.191	0.086	0.036	0.045	34.196	12	0.001	0.933	0.351
37	1.92	0.834	0.147	0.08	0.03	0.041	59.895	12	0	2.21	0.027
38	1.72	0.87	0.18	0.084	0.034	0.044	23.878	12	0.021	1.205	0.228
39	2.095	0.928	0.154	0.086	0.03	0.04	23.843	12	0.021	2.304	0.021
40	1.466	0.853	0.216	0.087	0.039	0.048	28.591	12	0.005	0.617	0.537
41	1.945	0.928	0.234	0.097	0.035	0.046	23.027	12	0.028	0.297	0.766
42	2.13	0.849	0.284	0.101	0.034	0.049	12.548	12	0.403	2.751	0.006
43	1.29	0.905	0.227	0.089	0.044	0.049	21.445	12	0.044	0.684	0.494
44	1.57	0.873	0.201	0.086	0.037	0.046	43.073	12	0	0.84	0.401
45	1.768	0.799	0.194	0.084	0.033	0.046	21.912	12	0.039	0.71	0.478
46	1.534	0.896	0.164	0.081	0.036	0.043	43.395	12	0	1.701	0.089
47	1.723	0.787	0.174	0.081	0.032	0.045	34.083	12	0.001	1.441	0.15
48	1.492	0.847	0.183	0.083	0.037	0.046	34.167	12	0.001	1.196	0.232
49	1.301	0.856	0.196	0.084	0.041	0.048	29.617	12	0.003	1.153	0.249
50	1.51	0.805	0.214	0.086	0.037	0.048	25.245	12	0.014	0.66	0.509

Table 3*: SEE of difficulty index (****b****) distribution of items of the MAT*

SEEb ≥ 1	$0.03 \leq SEE_b < 0.05$	$0.05 \leq SEE_b \leq 0.1$
1	5,7,8,10,11,12,13,14,15,17,16,18,19,20,21,22,23,24,25,26,27,28,29,30,31 ,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50	2,3,4,6,9
	44 items	5 items
Poor item	Retained items	Poor items / not retained

Table 3 above shows that 49 difficulty items SEEs are less than 1, only item 1 has SEE greater than 1.00. 44 items lie between 0.03 and 0.05

Table 4: SEE of the discrimination index (***a***) of the items of the MAT

0<SEEa<0.079	0.079 ≤SEEa< 0.1	SEEa ≥ 0.1
	5,6,7,9,10,11,12,13,14,15,16,17,18,19,20,22,23,24,25,26,27, 28,29,30,31,32,33,34,35,36,37,38,39,40,41,43,44,45,46,47,4 8,49,50	1,2,3,4,8,21, 42
0 items	43 items	7 items
	Good/ fairly good items	no good items

Table 4 indicates that 43 item discrimination SEEs are less than 1 while seven (7) items, item discrimination SEEs are greater than 1

Table 5: SEEs of the Guessing index (***c***) of the items of the MAT

0.0<SEEc< 0.039	0.04≤ SEEc ≤0.05	0.05≤ SEEc≤0.1
25,32,	3,5,7,8,10,11,12,13,14, 15,16,17,18,19,20,21,22,23,24,26,27,28,30,31, 33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49, 50	1,2,4,6,9,29
2 items	42 items	6 items

Table 5 shows that two (2) items lie in the interval 0.0<SEEc< 0.039, 42 items lie in the interval 0.04≤ SEEc ≤0.05 and 6 items in 0.05≤ SEEc≤1.00

Research question 3: What are the SEEs of items selected as good/fairly good items?

Table 6 shows items (**) in SEL as good/ fairly good items.

Item	*a*	*b*	*c*	*a SE*	*b SE*	*c SE*	*Chi-sq*	*df*	*p*	*z Resid*	*p*	*sel*
1	0.411	-0.251	0.352	0.115	0.103	0.076	251.554	12	0	1.319	0.187	*
2	0.621	0.306	0.327	0.111	0.072	0.067	127.987	12	0	0.753	0.452	*

3	1.416	1.144	0.277	0.103	0.051	0.048	16.517	12	0.169	0.748	0.454	*
4	0.533	0.278	0.26	0.109	0.076	0.066	168.678	12	0	1.487	0.137	*
5	1.219	1.046	0.19	0.084	0.046	0.046	27.195	12	0.007	0.98	0.327	**
6	0.981	0.953	0.266	0.093	0.056	0.053	41.553	12	0	0.772	0.44	*
7	1.401	1.037	0.179	0.084	0.042	0.044	17.911	12	0.118	1.015	0.31	**
8	1.603	1.034	0.284	0.104	0.045	0.049	20.389	12	0.06	0.84	0.401	*
9	1.044	0.838	0.262	0.093	0.051	0.053	32.257	12	0.001	1.022	0.307	*
10	1.371	1.057	0.219	0.09	0.046	0.046	22.188	12	0.035	0.605	0.545	**
11	1.522	0.995	0.27	0.099	0.044	0.049	10.296	12	0.59	0.574	0.566	**
12	2.155	0.979	0.189	0.095	0.032	0.042	24.182	12	0.019	0.762	0.446	**
13	1.953	1.042	0.168	0.091	0.035	0.04	10.397	12	0.581	1.489	0.136	**
14	1.64	1.153	0.191	0.094	0.043	0.042	10.958	12	0.533	0.892	0.372	**
15	1.3	1.168	0.259	0.098	0.053	0.048	14.758	12	0.255	0.481	0.63	**
16	1.467	1.022	0.216	0.091	0.043	0.046	25.124	12	0.014	0.605	0.545	**
17	1.649	1.08	0.169	0.088	0.039	0.041	28.392	12	0.005	1.343	0.179	**
18	1.74	0.992	0.196	0.091	0.037	0.044	20.736	12	0.054	1.094	0.274	**
19	1.212	0.954	0.22	0.087	0.046	0.048	18.994	12	0.089	0.485	0.627	**
20	1.522	1.02	0.22	0.092	0.042	0.046	18.14	12	0.111	0.778	0.436	**
21	1.681	1.263	0.215	0.103	0.048	0.043	27.144	12	0.007	0.609	0.543	***
22	1.768	0.875	0.207	0.089	0.035	0.046	28.306	12	0.005	1.075	0.282	**
23	1.298	0.961	0.228	0.09	0.045	0.048	7.745	12	0.805	0.401	0.688	**
24	1.414	1.097	0.253	0.098	0.048	0.047	15.04	12	0.239	0.521	0.603	**
25	1.911	1.172	0.17	0.097	0.039	0.039	30.895	12	0.002	1.351	0.177	***
26	1.559	0.969	0.199	0.088	0.039	0.045	21.636	12	0.042	0.521	0.602	**
27	1.42	0.919	0.249	0.093	0.043	0.049	32.363	12	0.001	0.546	0.585	**
28	1.711	1.01	0.158	0.085	0.036	0.041	28.103	12	0.005	1.718	0.086	**
29	1.29	0.662	0.238	0.086	0.04	0.053	49.719	12	0	0.673	0.501	**
30	1.392	0.925	0.231	0.09	0.043	0.048	7.235	12	0.842	0.379	0.705	**
31	1.633	0.82	0.201	0.085	0.035	0.047	32.429	12	0.001	0.994	0.32	**
32	1.575	1.021	0.136	0.079	0.037	0.039	55.601	12	0	2.289	0.022	***
33	1.767	0.762	0.22	0.087	0.033	0.048	18.756	12	0.095	1.347	0.178	**
34	1.343	1.07	0.208	0.089	0.046	0.046	23.1	12	0.027	1.279	0.201	**
35	1.909	0.857	0.208	0.089	0.033	0.045	24.227	12	0.019	1.483	0.138	**
36	1.642	0.894	0.191	0.086	0.036	0.045	34.196	12	0.001	0.933	0.351	**
37	1.92	0.834	0.147	0.08	0.03	0.041	59.895	12	0	2.21	0.027	**
38	1.72	0.87	0.18	0.084	0.034	0.044	23.878	12	0.021	1.205	0.228	**
39	2.095	0.928	0.154	0.086	0.03	0.04	23.843	12	0.021	2.304	0.021	**
40	1.466	0.853	0.216	0.087	0.039	0.048	28.591	12	0.005	0.617	0.537	**
41	1.945	0.928	0.234	0.097	0.035	0.046	23.027	12	0.028	0.297	0.766	**
42	2.13	0.849	0.284	0.101	0.034	0.049	12.548	12	0.403	2.751	0.006	*
43	1.29	0.905	0.227	0.089	0.044	0.049	21.445	12	0.044	0.684	0.494	**
44	1.57	0.873	0.201	0.086	0.037	0.046	43.073	12	0	0.84	0.401	**
45	1.768	0.799	0.194	0.084	0.033	0.046	21.912	12	0.039	0.71	0.478	**
46	1.534	0.896	0.164	0.081	0.036	0.043	43.395	12	0	1.701	0.089	**
47	1.723	0.787	0.174	0.081	0.032	0.045	34.083	12	0.001	1.441	0.15	**
48	1.492	0.847	0.183	0.083	0.037	0.046	34.167	12	0.001	1.196	0.232	**
49	1.301	0.856	0.196	0.084	0.041	0.048	29.617	12	0.003	1.153	0.249	**
50	1.51	0.805	0.214	0.086	0.037	0.048	25.245	12	0.014	0.66	0.509	**

In table 6, 42 items double (**) are grouped as good/fairly good items based on the criteria that SEEa< 0.1, SEEb< 0.1 and SEEc<0.05 . These items are not to be rejected. The other 8 items (single *) are to be rejected.

IV. Discussion of findings

For 50 itemed MAT, the difficulty index ranges from -0.251 of item 1 to 1.26 of item 42. The discrimination index ranges from -0.411 of item 1 to 2.130 of item32 While the guessing index ranges from 0.136 of item32 to 0.352 of item1.

The standard errors of the discrimination parameters (SEE_a) shows that it ranges from 0.08 of item 37 to 0.115 of item 1, the difficulty index (SEE_b) ranges from 0.030 of items 37 and 39 to 0.103 of item and the guessing parameter (SEE_c ranges from 0.039 of item 25 and 32 to 0.076 of item 1. It could be observed that item 1 is no good item, the item discrimination parameter is low -0.411 which show that it did not discriminate between the high and low ability.Similarly, the item appears easy to the examinees with - 0.251 as the difficulty index and a see_b for the item as 0.013 which indicates a low information the guessing parameter is very high 0.352 and a see_c of 0.076 the highest compared to the SEEs of the rest of the items in the test. This item should be discarded. This agrees with Toland (2008) that the accuracy of standard error of estimate of b parameter under 3pl depends on the amplitude of the parameter being estimated.

A closer look at the distribution of difficulty parameters and associated standard error of estimates, it could be observed that items with difficulty index less than 1.00 has a higher standard error of estimates than items whose b parameter is greater or equal to 1.00 (b ≥), its also noticed that item 1 did not fit the 3pl model of IRT.Item 37 with the lowest standard error of estimate of 0.08 for ***a*** parameter of 1.92 and ***b*** item parameter of 0.834 with see_b of 0.03, guessing item parameter of 0.147 with SEE_C of 0.041 did not fit the 3pl model. Though the item discriminates well and moderate difficult level but guessing parameter of 0.03 which is virtually reduces the item to a 2 pl model and irrelevant to 3plm items 37 and 39 has 0.30 as their standard error of estimates and did not fit the 3plm nevertheless have moderate difficulty parameter of 0.834 and 0.928, they also have equivalent error of estimate of c-parameter of 0.41 and 0.4 respectively.

Items 25 and 32 have the lowest c-parameter standard error of estimates of 0.039 with c-item parameter of 0.17 and 0.136 respectively. The relationship between the c parameter and b and its associated standard errors could to be attributed to factors described by Toland (2008) as test length, underlying ability (θ) or other factors. In summary 43 items were retained based on the consideration of the parameters SEEs

V. Conclusion

The findings from the study show that 42 items from the 50 itemed MAT item parameters were considered as having their parameters SEEs which lie within the acceptable limit. It also been observed that the derivation of the standard errors of estimates depend on the item parameters, there is a strong relationship between the item defaulting and its standard error of estimate. The lower the standard error of an item parameter, the more item information the items provide especially for the difficulty and discrimination index of the items. The SEEc (standard error of estimation of guessing parameter) is determined by the number of item distracters and interpreted directly.

Retaining or rejecting an item is a function of the interaction of all the item parameter SEEs of the particular item, some of the items require review to constitute a fairly good items. Most items with extreme values in the item parameters compared to the rest of the item parameter do not fit the data in 3pl.

VI. Recommendations

Based on the findings, the following were recommended.

1. That test developers, should apply three-parameter logistic model in their effort to develop a good test.
2. That the standard error of estimates of item parameters be used to identify or differentiate between good and bad items in any test development process.
3. Item response theory procedures be made accessible to examination bodies for credible test item construction

References

[1] Agresti, A., & Finlay, B. (1997).*Statistical methods for the social sciences* (3rd edition).Upper Saddle River, NJ: Prentice Hall.

[2] Blank, R.K; Alas, N. & Smith, C. (2008). The power of professional development and self-efficacy in the teaching of mathematics. Washington DC: McGraw Hill Education.

[3] Githua, B. N. &Mwangi, J.G. (2003). Students' mathematics self-concept and motivation to learn mathematics; relationship and gender differences among Kenya's secondary school students in Nairobi and Rift Valley Provinces. International Journal of Educational Development; 23, 487-499.

[4] Hambleton,R.K,, &Swaminathan,H.(1985):*Itemresponse theory principlesand applications.*Boston,MA:KluwerAcademic Publishers.

[5] Harwell,M.R.,Baker,F.B.,&Zwarts,M.(1988).Item parameter estimationvia marginal maximum likelihood and anEM algorithm: A didactic. *Journal of EducationalStatistics, 13*,243-271.

[6] Lord,F.M.(1980).*Applicationsofitemresponsetheorytopractical testingproblems*.Hillsdale,Ltd.

[7] McAlpine, M. (2002) .Asummary of methods of item analysis.Robert Clark center for technological education. University ofGlasgow, blue paper no 2.

[8] Obinne,A.D.E.(2011).A psychometric analysis oftwo major examinations inNigeria:Standard rrorofMeasurement.*Internationa lJournal of Education Science* ,3(2):137-144

[9] Thissen, D. &Wainer, H. (1992). Somestandarderrors in item response theory.Psychometrika 47, 397-41

[10] Toland,M.D.(2008)Determining theaccuracyofitem transactions,19:3,1028-1029.Retrieved from http://www.rasch.org/rmt/ rmt193c.htm Trials Based on Item ResponseTheory," ControlledClinicalTrials,24,390–410.TrustPublishers,pp.66-83.

[11] Xinning, A. & Yung, Y. (2014). Item response theory: What it is and how you can use IRT procedure to applt it. *Paper SAS* 363-2014

Ogbonna, Joseph U "Estimating Standard Errors of Irtparameters of Mathematics Achievement Test Using Three Parameter Model." IOSR Journal of Research & Method in Education (IOSR-JRME) , vol. 8, no. 2, 2018, pp. 01-07.