STRONG TRUE TEST SCORE THEORY ANALYSIS OF ITEM LOCATION AND SLOPE INDICES OF OYO STATE JUNIOR SECONDARY SCHOOL CERTIFICATE MATHEMATICS TEST ITEMS

Asowo Ayobode Patricia

Obafemi Awolowo University, NIGERIA

Abstract. This study has been undertaken to investigate item location (difficulty) and slope (discrimination) indices of Oyo State Junior Secondary School Certificate Examination (JSSCE) mathematics multiple-choice test items using the strong true test score theory (STTST) method. The study adopted a descriptive survey design. A sample of 600 candidates was selected using multi-stage sampling techniques from a population of 95,419 students who sat for the 2016 JSSCE Mathematics Paper 1 in Oyo State, Nigeria. An adoption version of the 2016 Oyo State JSSCE Mathematics Paper 1 titled "Mathematics Test" (MT) instrument was adopted for information aggregation. Data collected were analysed using BILOG-MG 3.0 software bundle. The solutions for item location under the 1, 2, and 3PL models found easy items to be 50, 42 and 25, ideal items stood at 8, 13 and 23, while difficult items stood at 2, 4 and 9 items respectively. Results for item slope under the 2PL and 3PL models found items with high slopes to be 2 and 18, moderate slopes stood at 23 and 29, while poorly sloped items stood at 34 and 10 respectively. Lastly, the results for pseudo-guessing under the 3PLM found 30 items to have a high degree of probability. In closing, the Oyo State 2016 JSSC Mathematics Examination under the strong true test score theory analysis had a moderate psychometric quality. It showed that most examinees including low ability examinees, had at least a moderate probability of answering the items along the MT correctly considering the individual examinee item location on the test, which signified that, the MT could not discriminate between examinees that understood the open content and those who do not.

Keywords: assessment practices**,** Mathematics test (MT), strong true test score theory (STTST), item parameters

---

Introduction

The realness of a nation's entire educational system completely depends fundamentally on the quality of its assessment practices. It is apparently not the query on either how efficient and effective a teacher is in the employment of instructional materials, methods of teaching, or how intellectual the students are if the opportunity for assessment and feedback is not given a prospect, the objective of teaching may completely be defeated. With appraisal and feedback, teachers would be capable to get precise information about learners' cognitive gains, psychological change in behaviour, and other skills acquired by learners during the process of scholarship. Usually, assessment of what students' have learned in the school is carried out by carefully administering appropriate achievement scale for students for feedback and decision-making. This is to supply information on students' achievement level, growth, scope to which implemented educational policy is successful, and the set procedures necessary for improvements in the subsequent teaching-learning process.

At the onset of formal training in Nigeria, students were evaluated through a single examination administered at the conclusion of the school year (Oyedeji, 2016). Nevertheless, one final test at the remainder of the term to determine students' success or failed attempts would be unfair which would not present a truthful description of how individual child behaved on the exam. As prevalent factors of learners' possible future behaviour like sudden ill health, phobia, fatigue, anxiety, accident, administration errors, and other recurrent students' psychological trait could significantly affect this single decision-making assessment process. As an effect, the continuous assessment process was inaugurated into the school curriculum with a singular drive of realising the extent to which the effect of educational policy is successful which could be utilized to check curriculum quality through evaluation. This is performed by making valid interpretations about changes in examinee's behavioural characteristics that are taken through the process of educational activity and scholarship.

In Nigeria, Junior Secondary School Certificate Examination (JSSCE) is considered as a basic public examination that is held at the end of the course of study for the certification. The basic test is taken by each state of the federation through their respective Ministry of Education (MOE), for final year students of Upper Basic Education (UBE) program at the remainder of junior secondary schooling. Omole (2007) posited that this case of testing is conducted outside the control of the school, which commonly produces a summative evaluation of nominees at the conclusion of an instructional plan. The Junior Secondary School Certificate Examination (JSSCE) is the focus of this paper with reverence to school-based appraisal process (SBAP) in which Mathematics is among one of its core fields in Nigeria.

According to Opara (2012), Mathematics is learned at all grades of instruction in Nigeria, that is why at the upper basic level (i.e., JSS1 – JSS3), an individual examinee must pass the field before they are upgraded to a higher grade or must have a credit level pass before gaining admission in any higher institution. This demand has made mathematics an essential subject for students' progress from one stage of instruction to another, it is likewise understood that mathematics is one of those cases that have entered a non-encouraging students' performance at the credential level. Betiku (2001) observed that Mathematics is one of the subjects that are poorly taught in schools, widely hated by students, and students, particularly girls, run away from the subject, as they believed mathematics is meant for the strong and talented. All these factors have affected the proficiency and achievement of students in Mathematics, especially at Junior Secondary School Certificate Examination (JSSCE).

Although, precautionary measures and efforts have been set in position to assist students to have an alteration in attitude towards learning Mathematics, and in developing Mathematics items that are faultless or near complete. Despite the standards put in place, students who are not good enough in subjects like Mathematics still have difficulties in applying Mathematics concepts, precepts, and skills to courses in the domain of science teaching. Therefore, as part of the measures put in place, educational institutions, examination bodies, and other organizations have identified training through exams, either written or oral for candidates, as the best means of determining the competency level of a person after having been exposed to many test or training experiences. Obemeata (2000) described that test item used for this purpose, usually take some time to get up because the trial would have to possess a certain degree of trustworthiness and cogency as well as a high degree of usability for the purpose it was meant. Since the major intent of all test developers is to construct a test of desired quality by taking the appropriate items that would oppose the designated use of the trial, no matter what type of tool or procedure is utilised in addressing the measurement problems.

In assessing the character of an assessment tool, it is imperative to talk over the fundamental operations needed in the procedure such as appropriate measurement models and item characteristic functions required in the form. It is explained the Strong True Test Score Theory (STTST) also experienced as the Item Response Theory (IRT) as a modern educational and psychological measurement approach that positioned a notion about cognition and some many sophisticated statistics used to appraise cognitive processes. It is apparent in tests and measurement that STTST is based on the assumptions that, an intelligent individual should receive a bigger chance of success on the assessment items than a less capable person should. Similarly, an examinee should always be more likely to behave better on an easier item than on a more difficult one.

As a psychometric approach, STTST provides clearer information about each item on a test. In STTST item analysis, each item characteristics are blended in order to reflect the quality of the examination. In this way, item characteristics such as item location (difficulty) and item slope or correlation (discrimination), and Chance (Pseudo-guessing) can be applied to assess the conduct of candidates on the items and the overall impact of the examination whether it evaluates the function it was designed for. To use a multiple-choice test in ranking students based on achievement level items must have the power to discriminate well or detect small differences in content learned by candidates in achievement tests. Item slope is the extent to which the items correlates between examinees who have not mastered the material taught or got the item incorrect and examinees who have mastered the concept or got the items correct.

As a standard, slope scores should range from – 1.00 to +1.00 with an ideal slope score of +1.00 as a positive coefficient indicating that high-level ability candidates tend to have higher scores on the test, while a negative coefficient would indicate that low-level ability candidates tend to cause lower scores on the exam. As an effect, items having negative slope are expected to be reviewed or discarded and items having discrimination index above 0.20 are ordinarily regarded as satisfactory for use in most examinations of academic achievement (Aggarwal, 1986; Kline, 2005). The discriminative power of a test item as explained by Onunkwo, (2002) depends on its difficulty level and largely on the credibility of distractors. Item location (difficulty) is the portion of the individual candidate that answered each item correctly; it reflects the level of easiness or difficulty of an item on a test. Item location can range from 0.0 indicating that, none of the candidates answered the item correctly to 1.0 that signifies that all the candidates answered the items correctly. Alternatively, several researchers have enlightened that an ideal level of item location for a four-option multiple-choice test should be between 60% and 80% with which an ordinary degree of difficulty can be accomplished. All the same, if the position of individual items falls outside this range, the item would have a low difficulty value that is less than 0.25. The causes could be ascribed to items that may have been misdirected; items that may be too challenging with respect to the overall power level of the candidate; items that may be equivocal or not written clearly; and there may be a multiple correct answers in the options (Hambleton & Jones, 1993; Adegoke, 2013). To place it plainly, an item that correlates effectively must have suitable grades of difficulty for comparison, and to ensure this, each of their districts must have a piece of credibility. Distractor credibility means that the distractor must be capable to pull off the attention of candidates who do not possess the required knowledge by answering an item correctly from the key or remain distractors to examinees who have the ability required to do an item correctly.

In Nigeria, several research efforts have been led and various good examples have been put through in seeking answers to the troubles of poor technique in secondary school Mathematics. Written reports have looked at assessing students' performance based on their pattern of answers to examine items in the external examination at a senior secondary school degree without paying full attention to the curriculum referenced JSSCE in Mathematics. Thus, the major focal point of this work was to investigate the psychometric properties of the Oyo State Junior Secondary School Certificate (JSSC) Mathematics multiple-choice test items using Strong True Test Score Theory (STTST) method whether the items are true and effective among candidates.

Research questions

The present paper tries to answer the questions: (1) what is the location index of the Oyo State JSSCE Mathematics multiple-choice test using the STTS theory; (2) what is the slope index of the Oyo State JSSCE Mathematics multiple-choice test using the STTS theory.

Methodology

This study adopted a descriptive survey research design. The population comprised all students that sat for the 2016 junior secondary school certificate (JSSC) Mathematics examination paper 1 in Oyo state, Nigeria. A sample of 600 students was randomly taken from a total population of 95,419 students who took the test. The tool for this subject was titled "Mathematics Test" (MT). It was a version of the Oyo State August/September (2016) JSSCE Mathematics paper 1. The 60 multiple-choice Mathematics questions covered a range of subjects in the junior secondary school (JSS) syllabus, showing that it had content validity. The reliability statistics of students' responses to the 60 multiple-choice mathematics questions were found to be 0.800 (no = 600) which was interpreted to be authentic and considered appropriate for the subject. The information collected was analysed using BILOG-MG 3.0 software bundle.

Results

Students were required to designate whether they are male or female. Their responses are summarised in Table 1.

Table 1. Distribution of respondents according to gender

Gender	Frequency	Percentage
Male	262	43.67
Female	338	56.33
Total	600	100.00

In Table 1, out of the six hundred (600) students sampled, 43.67% (262 students) were male while 56.33% (338 students) were female.

Research question one

What is the difficulty index of the Oyo State JSSCE Mathematics multiple-choice tests using STTS model?

All IRT estimations were obtained using the marginal maximum likelihood (MML) method with normal prior distribution, which is the default for BILOG-MG 3.0.

Within the IRT framework, the following parameters were calculated: (a) item slope (discrimination) parameter (a value); (b) item location (difficulty) parameter (the b value); (c) chance (Pseudo- guessing/Noise) parameter (c value).

The computed parameter estimates are shown in Table 2.

Table 2. Estimates of item parameters

	1PL	2PL		3PL				1PL	2PL		3PL		
Item	b	A	b	a	b	c	Item	B	a	b	a	b	c
1	0.347	0.554	0.158	0.778	-0.476	0.239	31	2.655	0.248	2.948	0.623	0.299	0.278
2	3.406	0.637	1.692	1.182	-0.255	0.500	32	1.368	0.354	1.089	0.892	0.075	0.184
3	-0.307	0.487	-0216	1.134	0.095	0.283	33	4.766	0.161	7.988	0.815	-0.189	0.228
4	0.785	0.440	0.502	1.230	0.945	0.193	34	1.483	0.259	1.580	1.093	-0.267	0.089
5	-0.819	0.394	-0.616	0.605	0.748	0.186	35	1.417	0.342	1.165	0.994	0.262	0.191
6	1.929	0.319	1.696	0.411	-1.215	0.221	36	2.697	0.318	12.385	0.751	-0.608	0.142
7	2.824	0.309	2.565	0.509	-0.497	0.194	37	1.287	0.347	1.043	3.175	0.917	0.327
8	1.617	0.212	2.079	0.631	-0.249	0.239	38	1.368	0.316	1.209	1.145	-0.154	0.110
9	0.527	0.432	0.335	0.623	-1.455	0.193	39	2.240	0.466	1.410	0.898	-0.864	0.128
10	2.978	0.314	2.663	0.483	0.241	0.256	40	1.159	0.258	1.240	2.330	1,255	0.238
11	0.481	0.481	0.272	1.203	1.215	0.285	41	2.433	0.206	3.222	0.980	0.764	0.218
12	0.302	0.406	0.197	0.650	-0.079	0.195	42	1.600	0.282	1.572	1.824	0.449	0.136

13	2.394	0.342	1.979	0.821	1.074	0.282	43	2.203	0.345	1.803	0.963	0.612	0.233
14	1.080	0.200	1.473	1.367	-0.692	0.141	44	1.207	0.100	3.222	1.009	- 0.060	0.093
15	1.368	0.351	1.097	1.390	-0.820	0.182	45	0.034	0.363	0.013	0.430	0.647	0.257
16	1.127	0.729	0.464	1.171	0.701	0.500	46	0.693	0.440	0.440	-	-	-
17	2.374	0.510	1.388	1.348	1.297	0.346	47	1.583	0.122	3.473	1.221	0.621	0.187
18	2.738	-	-	0.782	-0.986	0.222	48	1.823	0.473	1.126	1.034	0.880	0.239
19	1.401	0.500	0.818	0.684	0.104	0.306	49	2.824	0.153	4.965	1.083	0.356	0.119
20	1.982	0.214	2.526	1.109	-0.405	0.085	50	3.162	0.50	1.880	1.205	0.772	0.248
21	1.336	0.611	0.652	1.014	0.555	0.235	51	1.583	0.121	3.503	1.423	0.755	0.059
22	4.135	0.368	3.215	0.897	0.599	0.494	52	1.384	0.339	1.147	-	-	-
23	0.377	0.477	0.209	1.663	1.257	0.300	53	2.889	0.587	1..517	2.337	0.902	0.146
24	1.033	0.706	0.433	1.925	1.021	0.124	54	1.858	0.354	1.484	0.878	- 0.221	0.075
25	1.191	0.260	1.263	-	-	-	55	1.739	0.148	3.155	1.286	1.840	0.180
26	2.760	0.223	3.388	3.238	1.368	0.355	56	3.835	0.230	4.569	1.413	- 0.094	0.259
27	3.331	0.184	4.918	0.941	1.130	0.311	57	2.413	0.211	3.123	0.638	0.115	0.080
28	1.805	0.291	1.723	0.758	0.167	0.294	58	1.064	0.317	0.935	1.133	0.273	0.069
29	-1.208	0.417	-0.860	0.784	0.709	0.163	59	2.036	0.358	1.614	0.879	0.387	0.094
30	1.550	0.551	0.834	0.572	0.747	0.280	60	1.771	0.329	1.510	0.960	0.206	0.244

Table 2 showed the item parameter estimates obtained from the one, two, & three-parameter logistic models. Utilizing the one parameter logistic model (1PLM), all the 60 items fitted the model. Applying the two-parameter, logistic model (2PLM) 59 items all fitted the model, while item 18 mis-fitted the model because the item was less than -0.15. In addition, using the three-parameter logistic model (3PLM), 57 items fitted the model, while three items mis-fitted the model because the items were less than -0.15. These items are item 25, 46, and item 57. The results implies that the two-parameter logistic model best fitted the items considering the measurement of true variability in the difficulty and discrimination levels of the student.

Table 3 revealed that 83.33% (50 items) on the 2016 Oyo State JSSCE Mathematics multiple-choice items under 1PLM were easy items, 13.33% (8 items) were acceptably difficult. Under the 2PLM, 6.78% (4 items) were considered difficult, 22.03% (13 items) were considered ideal items and 71.19% (42 items) were weighed to be easy items. Under the 3PLM, 15.79% (9 items) were considered difficult, 40.35% (23 items) were considered ideal, while, 43.86% (25 items) were weighed to be easy items. The results imply that since there were more easy items on the test across models, the test could no longer differentiate between students that have actually learnt the subject content well and those who do not.

Table 3. Item Location Index (b) of 2016 Oyo State JSSCE MT Items for 1PLM, 2PLM, and 3PLM

Level of Location indices	1PLM	2PLM	3PLM	Decision
< 0.20	3, 45	1, 12, 15, 45	3, 12, 19, 20, 28, 32, 33, 44, 59	Very Difficult items
0.20 – 0.69	1,9,11,12,14,23,46,47	2 – 6, 9, 11, 14,16, 21, 23, 24, 46	1, 2, 7, 8, 10,14, 20 -22 31, 34 -36, 42, 43, 45, 47, 49 ,53, 54, 58 – 60	Medium/ideal items
0.70 - 0.90	2,4 - 8, 10, 13, 15 - 22, 24 – 44, 48 - 60,	7, 8, 10, 13, 17, 19, 20, 22, 25 – 44, 47 – 60	4 -6, 9, 11, 13, 15-18, 23 - 27, 29, 30, 37 - 41, 50 - 52, 55	Easy items

Source: Researcher's analysis, 2017

Research question two

What is the slope index of the Oyo State JSSCE Mathematics multiple-choice test using STTS model?

Table 4. Slope index (a) of 2016 Oyo State JSSCE MT Items for 2PLM, and 3PLM

Level of Slope indices	**Items**		**Decision**
	2PLM	**3PLM**	
0.01 – 0.34	6 -10, 13, 14, 19, 21, 25 - 28, 31, 33 - 44, 47, 49, 51, 52, 54 - 56, 58, 60		Very low/ poor items
0.35 – 0.64	1 - 5, 9, 11, 12, 15, 17, 20, 22, 23, 29, 30, 32, 39, 45, 46, 48, 50, 53, 59,	6 - 10, 31, 45, 57, 59, 60	Low /good items
0.65 - 1.34		1 - 5, 11 - 13, 16 - 22, 32 - 36, 38, 39, 41, 43, 44, 47, 49, 50, 58	Moderate/very good items
1.35 – 1.69	16, 24,	14, 15, 23, 26 - 30, 48, 51, 55, 56,	High items
>1.70		24, 37, 40, 42, 53	Very high items

Source: Researcher's analysis, 2017

Table 4 revealed that the 2016 Oyo State JSSCE Mathematics multiple-choice items under the 2PLM 3.38% (2 items) had a very high slope, 38.98% (23 items) sloped well and 57.63% (34 items) sloped poorly. Under the 3PLM, 8.77% (5 items) had a very high slope, 22.81% (13 items) had a high slope, 50.88% (29 items) sloped moderately and 17.54% (10 items) sloped poorly respectively. This implied that there were more poor items under the 2PLM but poor items were reduced under the 1PLM which showed that students had a higher chance of guessing the items right. The results implies that the item parameter estimates of Oyo State JSSC Mathematics Examination under the STTS model were easy items (item location) and could no longer discriminate well among low and high-level ability candidates (item slope).

Discussion of findings

Findings from the study showed that the item parameter estimates of students' responses to Oyo State JSSC Mathematics Examination based on the strong true test score model were easier items (item location) and incapable of discriminating between low ability candidates and high ability candidates (item slope). The test reflected that the items displayed a high presence of pseudo-guessing under the 3PLM, which made the items capable of working well among low ability examinees. This in turn, was a disadvantage to high ability learners because low ability candidates had eminent probability of getting extremely difficult items correctly on the test. However, it is understandable in measurement that, such items are frequently needed to adequately attain sample course content and objectives across comparable levels.

The findings agreed with the studies of Nkpone (2001), Adedoyin and Mokobi (2013) that the more difficult an item is on a test, the higher an examinee's ability must be in order to answer the item correctly. Items with high b values are easy items, that is, values of b greater than 1 which indicated that most examinees including those with low ability would have moderate chance of answering the items correctly. Correspondingly, items with low b values or items below -1 are hard items that indicated very difficult items and on it low ability examinees are unlikely to answer it correctly. The results also revealed that the 2016 Oyo State JSSCE Mathematics multiple-choice test had a discriminating mean of 0.35. This finding was consistent with the findings of Moyinoluwa, (2015) that the acceptance discrimination power for a multiple-choice item should range between 0.4 and above. The scholar observed that when the lower ability level was extended to 0.35, 70% of items on the mathematics test discriminated better between the high and lower achievers. The significance of these findings is that the interpretation of the chance low ability candidates got in answering difficult items correctly either through guessing or test wiseness could affect the conclusion reached. Equally, the psychometric properties of the test could not efficiently reflect whether candidates have satisfactorily learned the subject curricular being tested on before the administration of the test or possibly the items on the test are of poor item quality.

Conclusion

The 2016 Oyo State JSSCE Mathematics multiple-choice test using the Strong True Test Score Theory (STTST) was of moderate psychometric quality based on its measurement indices. It showed that most examinees including low ability examinees had at least a moderate probability of answering the items along the MT correctly considering individual examinee item location on the test. This signified that the MT could not discriminate well between examinees that understood the subject content and those who do not.

Recommendations

From the findings and conclusion of the study, the following recommendations are drawn: (1) Test developers and examination agencies should adopt a confidence scoring pattern of Mathematics multiple-choice test with the inclusion of a five option formats. This could improve the psychometric properties of Mathematics test items and assist in correcting the issue of blind guessing and test wiseness among examinees with the functional power of item distractors. (2) Measurement experts and subsequent studies should examine the proportion and effect of item bias on students' performance using modern frameworks to provide information on the validity and reliability of items drawn from mathematics blueprint in ensuring test security.

` **REFERENCES**

Adedoyin O.O. & Mokobi, T. (2013). Using irt psychometric analysis in examining the quality of junior certificate mathematics multiple-choice examination test items. *Int. J. Asian Soc. Sci., 3,* 992-1011.

Adegoke, B.A. (2013). Comparison of item statistics of Physics achievement test using classical test and item response theory frameworks. *J. Educ. & Practice, 4*(22), 87- 96.

Agarwal, Y.P. (1986). *Statistical methods, concepts, applications, and computations*. New Delhi: Sterling.

Betiku, O.F. (2001) Causes of mass failure in Mathematics examination among students: A commissioned paper presented at Government Secondary School Karu, Abuja Science Day, 1st March.

Hambleton, R.K. & Jones, R.W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educ. Measurement: Issues & Practice, 12*(3), 38 -47.

Kline, T.J.B. (2005). Classical test theory: assumptions, equations, limitation, and item analyses (pp. 91-106). In: Kline, T.J.B. (Ed.). *Psychological testing: a practical approach to design and evaluation.* Calgary: Sage.

Moyinoluwa, T.D. **(**2015**).** Analysing the psychometric properties of mathematics in public examinations in Nigeria**.** *Res. Humanities & Soc. Sci., 5*(7), 15-21.

Nkpone, H.L. (2001). *Application of latent traits models in the development and standardisation of physics achievement test for senior secondary students: PhD thesis.* Nsukka: University of Nigeria.

Obameata, J.O. (2000). *Principles of essay and multiple-choice test construction*. Ibadan: Sterling.

Omole, D.O.K. (2007). Comparative study of students' performance in school-based assessment and certificate examination at the upper basic education level in FCT. *Nigerian J. Educ. Res. & Evaluation, 7*(1), 50-56.

Onunkwo, G.I.N. (2002). *Fundamentals of education measurement and evaluation*. Owerri: Cape.

Opara I.M. (2012). Psycho-social predictors of students' achievement in mathematics: imperatives for enhancing the quality of educational assessment in mathematics. *J. Educ. Rev., 5*(2), 229-235.

Oyedeji, S. O. (2016). Educational assessment in Nigerian secondary schools: historical analysis, challenges, & prospects. *Int. Adv. J. Teaching & Learning, 2*(4), 44-45.

✉ Asowo Ayobode Patricia
Department of Educational Foundations and Counselling,
Obafemi Awolowo University
Ile-Ife, Nigeria
E-Mail: ayobodeasowo@gmail.com