UNIJERPS
Unizik Journal of Educational Research and Policy Studies VOL.16 (5); https://unijerps.org
2023 Impact Factor: **6.834;** 2022 ***Impact Factor***: 5.156; ***2021 Impact Factor: 5.641; 2020 Impact Factor: 5.355.***
Indexing and Impact Factor URL: http://sjifactor.com/passport.php?id=21363
July-December, 2023

PSYCHOMETRIC ANALYSIS AT ITEM LEVEL OF WAEC MAY/JUNE MATHEMATICS MULTIPLE CHOICE QUESTIONS USING THE ITEM RESPONSE THEORY

Solomon Chukwu Ohiri
Directorate of Academic Planning,
Alvan Ikoku Federal College of Education, Owerri Imo State
E-mail: ohiri.chukwu2018@gmail.com

ABSTRACT

The study investigated psychometric properties of the West African Examinations Council (WAEC) May/June mathematics multiple choice items from 2019 to 2021 using item response theory (IRT). Survey research design that adopted simple random sampling technique was used in selecting a sample of 2,484 senior secondary three (SS3) students from Owerri Education Zone 1 of Imo State. The study was guided by two research questions and two hypotheses. The instruments used were 2019, 2020 and 2021 WAEC mathematics multiple choice test items. The instruments (test items) were assumed to be valid and reliable by nature of standardized instrument administered by WAEC. The collected data were analyzed using jMetrik software in order to answer the research questions while ANOVA was used to test the hypotheses at α = 0.05 level of significance. Findings from the study showed that 43(86%) of the items in 2019, 41(82%) of the items in 2020 and 38(76%) of the items in 2021, satisfied the difficulty acceptable range. Again, 33(66%) of the items in 2019, 34(68%) of the items in 2020 and 32(64%) of the items in 2021 satisfied the acceptable discrimination range. The items had reasonable difficulty and discrimination levels. Based on the findings of this study, it was recommended that WAEC and other examination bodies should ensure that only quality items are developed for assessment.

Keywords: Psychometric Analysis, Item, Mathematics. Difficulty and Discrimination

Introduction

The primary concern of testing in the teaching and learning process is that the measurements derived from tests will be helpful in making valid decisions. Examiners and other test users are usually interested only in the results yielded by the test administration. Generally, they are not attuned to characteristics or technical features of the test and the test items. Still many persons who use test results do not realize that the usefulness and appropriateness of test-score interpretation is a direct result of the test's internal characteristics. The internal attributes of a test and its items are technically termed its psychometric properties or characteristics.

Psychometric characteristics of a test items refers to certain attributes inherent in the test upon which assessment of conditions is based (Moyinoluwa in Abdullahi and Darazo, 2020). Indicators of a particular test item's difficulty, its ability to discriminate among people having differing amounts of the construct being measured, and the likelihood that the correct answer could be chosen by guessing are each a psychometric property of an item. It is important to investigate and learn about tests' psychometric properties.

Psychometricians and other professionals who develop tests need to evaluate and describe how a test functions so that it can be built to a specified level of quality. Again, knowing about the psychometric properties of a test and test items, provides evidence that information obtained using such a test (instrument) can provide a sound basis for decision making.

Psychometric properties of test items include several indicators such as item difficulty, item discrimination, item mean, guessing parameter and distracter indices. In this study, only item difficulty and item discrimination are discussed.

Item difficulty parameter in IRT: This is where the item functions along the ability scale. An easy item functions among the low ability examinees and a hard item functions among the high ability examinees. This means that difficulty can be considered as the location index (Ashraf and Jaseem, 2020). This refers the proportion of examinees that correctly answered an item. This metric takes a value between 0 and 1. High values indicate that the item is easy, while low values indicate that the item is difficult. An ideal item is supposed to have a difficulty index of 0.5, but may be difficult to have items with this index. Hence, an item is acceptable if the difficulty index falls between 0.3 and 0.7. if the difficulty index is less than 0.3, it shows that the item is difficult, while any value greater than 0.7 indicates that the item is very easy.

Item discrimination parameter in IRT: The item discrimination indicates the extent to which success on an item corresponds to success on the whole test (Ashraf and Jaseem, 2020). It describes how well an item can differentiate between examinees having the trait below the item location and those having the trait above the item location. According to Okoye (2015), an item is considered good if it is got right by the bright students and failed by the dull ones. Item discrimination refers to the power of the item to differentiate between examinees with high and low levels of knowledge or ability (Thompson, 2016). It is the correlation between item scores and total test scores called the item-total correlation. A good item records more passes in the upper one-third than in the lower one-third. The discrimination index ranges from -1 to +1. When the value is negative, it shows that more of the lower group chooses the correct option than the upper group (this is an abnormal situation, and such item is a bad one). On the other hand, items with positive values are most acceptable when the discrimination index is +1.0 and very poor in discrimination if the value is 0. An item is acceptable if the discrimination index fails between +0.3 and +1.0 (Okoye, 2015).

In the study of psychometric properties, there are two major theories, which are classical test theory and item response theory and their corresponding models. The current study attempted to ascertain the psychometric properties of test items of the West African Examinations Council (WAEC) May/June mathematics multiple choice questions from 2019-2021 using 2PL model of item response theory (IRT), in Owerri Education Zone I of Imo State of Nigeria.

A sizable number of studies on psychometric item analysis on mathematics tests have been conducted. Cobbinah and Ntumi (2022) carried out a study on the difficulty, discrimination and pseudo-guessing indices of the West African Examinations Council core mathematics objective test items in Ghana

within the remit of IRT. The study revealed that averagely, the 2020 core mathematics objective test items were within the difficulty level. In terms of discrimination level, the psychometric properties indicated that most of the items discriminated among the examinees. Ibrahim (2023) investigated on the psychometric properties analysis of mathematics of the West African Senior School Certificate Examination in Dala Education Zone, Kano State, Nigeria (2020-2022). The findings showed that there was no significant difference between the difficulty, discrimination and distracter indices of WASSCE Mathematics from 2020 to 2022. Kasali and Adeyemi (2022) carried out a study on the estimation of item parameter indices of NECO mathematics multiple choice test items among Nigerian students. The study estimated the difficulty, discrimination, and vulnerability to guessing of the 2016 National Examination Council (NECO) mathematics multiple-choice test items. The study employed an ex-post facto design with 276,338 samples. The research instruments used for the study were Optical Marks Record Sheets for the NECO June/July 2016 mathematics objectives items. The responses of the testees were scored dichotomously. Data collected were calibrated using four parameters logistic model. The results showed that most items in the 2016 NECO mathematics test had good difficulty indices. Also, the results indicated that only 21.7% of 2016 NECO Mathematics test items had a very good discrimination power, and the majority of the items had poor discrimination power. The result implies that most of the items were not effective in discriminating between examinees with the required ability and those that lack the required ability. Finally, the result revealed that the 2016 NECO test items were not vulnerable to guessing (i.e., 86.7% of items were good in terms of guessing).

Abdullahi and Darazo (2020) carried out a study on the analysis of psychometric properties of 2016 mathematics Basic Education Certificate Examination Questions (BECEQ) in Gombe State. Findings from the study showed that 52 of the 60 test items were easy; the examination discriminated well between the upper and lower ability groups of students, whereas many items showed a high probability to guessing. Aduloju and Okoh (2021) analyzed the item difficulty and discrimination indices as functions of changes in mathematics objective tests of the Benue State Examination Board Basic Education Certificate Examination (BECE) from 2015 to 2017. The results of the findings among others revealed that 5% of the items were easy across the years. However, in 2015, 40% of the items were moderate while 55% were difficult. In 2016, 22% of the items were moderate while 73% were difficult. Also, in 2017, 28% of the items were moderate and 67% were difficult. There was no significant difference in difficulty and discrimination indices of the items through the three years of study. Kinanee, Bosede and Orluwene (2017) analyzed and compared the item difficulty, discrimination and distracter indices of mathematics Junior Secondary School Certificate Examination (JSSCE) questions for students in Rivers State of Nigeria from 2014 to 2016. The results of the study showed that the test items for 2015 and 2016 had good item difficulty and discrimination indices. Test items for 2014 had 23% pass on difficulty indices criterion and 43% pass on the discrimination criterion. All the test items for the three consecutive years passed the distracter indices criteria.

Research Questions

The following research questions guided the study.

1. What are the percentage item difficulty indices of the test items of the 2019, 2020 and 2021 WAEC May/June mathematics multiple choice questions?
2. What are the percentage item discrimination indices of the 2019, 2020 and 2021 WAEC May/June mathematics multiple choice questions?

Research Hypotheses

The following research hypotheses were formulated and tested at $\alpha = 0.05$ level of significance.

1. There is no significant difference in the item difficulty indices of the 2019, 2020 and 2021 WAEC May/June mathematics multiple choice questions.
2. There is no significance difference in the item discrimination indices of the 2019, 2020 and 2021 WAEC May/June mathematics multiple choice questions.

Research Method

The research design adopted for this study was survey research design. According to Nworgu (2015), survey research design is one in which a group of people or items is studied by collecting and analyzing data from only a few people or items considered to be a representative of the entire group. This design was considered appropriate because only a part of the population was studied and findings were used to generalize for the entire population.

The study was carried out in Owerri Education Zone 1 of Imo State. The Zone is in the Eastern part of the State. It is made up of five (5) Local Government Areas which includes Ikeduru, Mbaitoli, Owerri Municipal, Owerri North and Owerri West. The Population of the study comprised all the SS3 students of the public secondary schools in Owerri Education Zone 1 of Imo State in the 2022/2023 academic session. There were 9,886 students.

The number of sampled candidates used in the study was 2,484 students. This comprised 1,178 male and 1,306 female students. Simple random sampling technique was employed for selection of the sample in the study. Two schools were obtained from each Local Government Area. This gave rise to ten schools. All SS3 students in the ten (10) sampled schools were used for the study. On the whole, a total of 2484 students made up of 1178 males and 1306 females were sampled.

The instruments used for data collection were the 2019, 2020 and 2021 May/June Multiple choice mathematics questions adopted from the West African Examinations Council (WAEC). Each of the instruments consists of 50-items. Each item consists of a question and a list of possible answers lettered A - D of which only one option is the correct answer. The test items of the instruments were scored 1 for correct option and 0 for wrong option with maximum score of 50 and minimum of 0.

The validation of the instruments was done by experts in the department of examinations and quality control of the West African Examinations Council (WAEC) and therefore require no further validation since the instruments were adopted. Thus, the items were considered appropriate in terms of subject contents and instructional objectives. On the other hand, being instruments of standardized international examination, which were conducted by the West Africa Examinations Council (WAEC), the instruments were deemed

reliable. Hence, the reliability of the instruments was not established by the researcher.

To collect pertinent data needed for the study, the instruments were administered to the SS3 students in each of the sampled schools with the help of the mathematics teachers, served as the research assistants. The researcher, through the teachers, informed the students ahead of time about the exercise and the need to be prepared because it will form part of their continuous assessment. This measure was to ensure that the students put in their best.

The researcher collected the candidates' responses on the May/June WAEC 2019, 2020 and 2021 mathematics multiple-choice questions. The data collected (students' responses on each item) were analyzed using item analysis statistics built in the jMetrik software for determination of the test item characteristics (difficulty and discrimination indices)

Presentation of Results of the Study

Research Question I: What are the item difficulty indices of the test items of the 2019, 2020 and 2021 WAEC May/June mathematics multiple-choice questions?

Table 3: Item Difficulty Indices of WAEC May/June Mathematics Multiple-Choice Questions 2019-2022.

Items	2019	Remark	2020	Remark	2021	Remark
1.	.8880	Easy	.5425	Moderate	.9040	Easy
2.	.8155	Easy	.5120	Moderate	.8495	Easy
3.	.7710	Easy	.7710	Easy	.8100	Easy
4.	.4190	Moderate	.4240	Moderate	.2260	Difficult
5.	.5195	Moderate	.5210	Moderate	.5420	Moderate
6.	.5440	Moderate	.5435	Moderate	.5630	Moderate
7.	.7240	Easy	.7260	Easy	.7550	Easy
8.	.5265	Moderate	.5285	Moderate	.5425	Moderate
9.	.5265	Moderate	.5230	Moderate	. 5420	Moderate
10.	.5410	Moderate	.5500	Moderate	. 5405	Moderate
11.	.5680	Moderate	.8835	Easy	. 5740	Moderate
12.	.5430	Moderate	.8095	Easy	. 5695	Moderate
13.	.5140	Moderate	.5120	Moderate	. 5255	Moderate
14.	.7155	Easy	.7085	Easy	. 7370	Easy
15.	.4660	Moderate	.4715	Moderate	. 4890	Moderate
16.	.4880	Moderate	.4925	Moderate	. 5045	Moderate
17.	.6515	Moderate	.6505	Moderate	. 6790	Moderate
18.	.4340	Moderate	.4330	Moderate	. 4340	Moderate
19.	.5060	Moderate	.5025	Moderate	. 2030	Difficult
20.	.4970	Moderate	.4980	Moderate	. 4975	Moderate
21.	.6290	Moderate	.6360	Moderate	. 6555	Moderate
22.	.4990	Moderate	.5020	Moderate	. 5000	Moderate
23.	.7000	Moderate	.7005	Easy	. 7150	Easy
24.	.6900	Moderate	.6855	Moderate	. 6995	Moderate
25.	.5590	Moderate	.5530	Moderate	. 5795	Moderate
26.	.4990	Moderate	.4985	Moderate	. 4935	Moderate

27.	.6845	Moderate	.6750	Moderate	. 6985	Moderate
28.	.5870	Moderate	.5930	Moderate	. 6145	Moderate
29.	.4900	Moderate	.4915	Moderate	. 5060	Moderate
30.	.5260	Moderate	.5280	Moderate	.5405	Moderate
31.	.5905	Moderate	.2875	Difficult	.2165	Difficult
32.	.5770	Moderate	. 5760	Moderate	.5830	Moderate
33.	.5600	Moderate	. 5690	Moderate	.2590	Difficult
34.	.5160	Moderate	. 2115	Difficult	.5215	Moderate
35.	.5600	Moderate	. 5550	Moderate	.5620	Moderate
36.	.5580	Moderate	. 5560	Moderate	.5680	Moderate
37.	.5055	Moderate	. 5085	Moderate	.5255	Moderate
38.	.5440	Moderate	. 5440	Moderate	.5695	Moderate
39.	.4980	Moderate	. 5025	Moderate	.5165	Moderate
40.	.4815	Moderate	. 4805	Moderate	.4995	Moderate
41.	.6340	Moderate	. 6330	Moderate	.2415	Difficult
42.	.3400	Moderate	.2440	Difficult	.3485	Moderate
43.	.5235	Moderate	.5245	Moderate	.5280	Moderate
44.	.3705	Moderate	.3680	Moderate	.3740	Moderate
45.	.6875	Moderate	.6905	Moderate	.7070	Easy
46.	.5430	Moderate	.5415	Moderate	.5625	Moderate
47.	.5495	Moderate	.5495	Moderate	.5570	Moderate
48.	.5205	Moderate	.5205	Moderate	.5320	Moderate
49.	.2900	Difficult	.5900	Moderate	.6085	Moderate
50.	.2275	Difficult	.5275	Moderate	.5385	Moderate

Table 3 showed that in 2019, five items were easy, forty–three items were moderately difficult while two items were too difficult. Also, in 2020, six items were easy, forty-one were moderately difficult, while five items were too difficult. Furthermore, table 3 also showed that in 2021, seven item were too easy, thirty-eight items were moderately difficult while five items were too difficult.

Table 4: Percentage Summary of Item Difficulty for 2019-2021.

Year	Easy	Moderate	Difficult
2019	5 items $\frac{5}{50} \times \frac{100}{1} = 10\%$	43 items $\frac{43}{50} \times \frac{100}{1} = 86\%$	2 items $\frac{2}{50} \times \frac{100}{1} = 4\%$
2020	6 items $\frac{6}{50} \times \frac{100}{1} = 12\%$	41 items $\frac{41}{50} \times \frac{100}{1} = 82\%$	3 items $\frac{3}{50} \times \frac{100}{1} = 6\%$
2021	7 items $\frac{7}{50} \times \frac{100}{1} = 14\%$	38 items $\frac{38}{50} \times \frac{100}{1} = 76\%$	5 items $\frac{5}{50} \times \frac{100}{1} = 10\%$

Table 4 showed that 10% of items were easy in 2019, 86% of the items were moderately difficult, while 4% of the items were too difficult. In 2020, 12% of the items were easy, 82% of the items were moderately difficult while 6 % of the items were too difficult. In 2021, 14% of the items were easy, 76% were moderately difficult while 10% of the test items were too difficult.

Research Question 2. What are the item discrimination indices of the test items of the 2019, 2020 and 2021 WAEC May/June mathematics multiple-choice questions?

Table 5: Item Discrimination Indices of WAEC May/June Mathematics Multiple-Choice Questions 2019-2021.

Item	2019	Remark	2020	Remark	2021	Remark
1.	.1201	Poor	.1376	Poor	.0514	Poor
2.	.2018	Fair	.2158	Fair	.1350	Poor
3.	.2735	Fair	.2708	Fair	.1975	Poor
4.	.1514	Poor	.1330	Poor	.1174	Poor
5.	.4287	Excellent	.4227	Excellent	.4249	Excellent
6.	.3769	Good	.3833	Good	.0514	Poor
7.	.1782	Poor	.1745	Poor	.1350	Poor
8.	.3982	Good	.3971	Good	.1342	Excellent
9.	.3757	Good	.3803	Good	.3857	Good
10.	.4173	Excellent	.3859	Good	.4604	Excellent
11.	.3772	Good	.3846	Good	.3920	Good
12.	.3953	Good	.3987	Good	.3684	Good
13.	.2842	Fair	.2866	Fair	.2896	Fair
14.	.2220	Fair	.2153	Fair	.1818	Poor
15.	.4104	Excellent	.3974	Good	.4084	Excellent
16.	.3148	Good	.3067	Good	.3081	Good
17.	.3148	Good	.3127	Good	.2863	Good
18.	.2135	Fair	.2127	Fair	.2419	Fair
19.	.3882	Good	.3936	Good	.4212	Excellent
20.	.3626	Good	.3616	Good	.3914	Good
21.	.2813	Fair	.2669	Fair	.2432	Fair
22.	.4088	Excellent	.3994	Good	.4185	Excellent
23.	.1887	Poor	.1913	Poor	.1741	Poor
24.	.2491	Fair	.2635	Fair	.2508	Fair
25.	.3715	Good	.3865	Good	.3687	Good
26.	.3284	Good	.3373	Good	.3743	Good
27.	.2612	Fair	.2786	Fair	.2505	Fair
28.	.3713	Good	.3615	Good	.3441	Good
29.	.4060	Excellent	.3934	Good	.3984	Good
30.	.3873	Good	.3849	Good	.3971	Good
31.	.4093	Excellent	.4187	Excellent	.4068	Excellent
32.	.3880	Good	.3653	Good	.4271	Excellent
33.	.3966	Good	.3653	Good	.4271	Excellent
34.	.3886	Good	.3967	Good	.4016	Excellent

35.	.4190	Excellent	.4157	Excellent	.4468	Excellent
36.	.4257	Excellent	.4284	Excellent	.3759	Good
37.	.3869	Good	.3810	Good	.3759	Good
38.	.4470	Excellent	.4466	Excellent	.4333	Excellent
39.	.4653	Excellent	.4576	Excellent	.4676	Excellent
40.	.3488	Good	.3446	Good	.3444	Good
41.	.3495	Good	.3482	Good	.3544	Good
42.	.2846	Fair	.3602	Good	.2579	Fair
43.	.3630	Good	.2820	Fair	.3915	Good
44.	.2735	Fair	.2829	Fair	.2611	Fair
45.	.2582	Fair	.0480	Poor	.2294	Fair
46.	.4439	Excellent	.4446	Excellent	.4408	Excellent
47.	.3879	Good	.3918	Good	.4119	Excellent
48.	.4005	Excellent	.4032	Excellent	.4156	Excellent
49.	.2977	Fair	.2946	Fair	.2957	Fair
50.	.3557	Good	.3515	Good	.3734	Good
Mean	0.3389		0.3332		0.3226	

Table 6 showed that in the year 2019, twelve items had excellent discrimination, twenty-two items had good discrimination, twelve items had fair discrimination and four items with poor discrimination indices. Again in 2020, eight items had excellent discrimination indices, twenty-six items were with good discrimination indices, eleven items had fair and five items very poor discrimination indices. In 2021, sixteen items had excellent discrimination indices, sixteen with good discrimination indices, nine items with fair discrimination indices, while nine items had poor discrimination indices.

Table 7: Percentage summary of item discrimination indices for 2019-2021.

Year	Excellent	Good	Fair	Poor
2019	12 items $\frac{12}{50} \times \frac{100}{1} = 24\%$	22 items $\frac{22}{50} \times \frac{100}{1} = 44\%$	12 items $\frac{12}{50} \times \frac{100}{1} = 24\%$	4 items $\frac{4}{50} \times \frac{100}{1} = 8\%$
2020	8 items $\frac{8}{50} \times \frac{100}{1} = 16\%$	26 items $\frac{26}{50} \times \frac{100}{1} = 52\%$	11 items $\frac{11}{50} \times \frac{100}{1} = 22\%$	5 items $\frac{5}{50} \times \frac{100}{1} = 10\%$
2021	16 items $\frac{16}{50} \times \frac{100}{1} = 32\%$	16 items $\frac{16}{50} \times \frac{100}{1} = 32\%$	9 items $\frac{9}{50} \times \frac{100}{1} = 18\%$	9 items $\frac{9}{50} \times \frac{100}{1} = 18\%$

Table 7 showed that in 2019, 24% of the items discriminate excellently, 44% of the items were with good discrimination, 24% of the items had fair discrimination while 8% of the items had poor discrimination. In 2020, 16% of the items discriminated excellently, 52% of the items had good discrimination, 22% had fair discrimination and 10% of the items had poor discrimination indices. Furthermore, in 2021, 32% items discriminated excellently, 32% of the items had good discrimination, 18% had fair

discrimination while 18% of the items had poor discrimination indices.

Hypotheses

Ho_1: There is no significant difference in the item difficulty indices of the 2019, 2020 and 2021 WAEC May/June mathematics multiple choice questions

Table 8. ANOVA Table of Differences in Difficulty Indices in WAEC Mathematics 2019-2021.

Source of variation	Sum of square	DF	Mean square	F	P-value	Remark
Between group	.007	2	.003	.300	.741	NS
Within group	1.653	147	.011			
Total	1.659	149				

Table 8 showed that the p-value of 0.741 is greater than the alpha (α) value of 0.05. Hence, the test statistics is not significant and the null hypotheses is not rejected. This implies that there is no statistical significant difference between item difficulty indices in WAEC May/June mathematics multiple choice questions administered within the year 2019 and 2021.

HO_2: There is no significant difference in the item discrimination indices of the 2019, 2020 and 2021 WAEC May/June mathematics multiple choice questions.

Table 9. ANOVA Table of Differences in Discrimination Indices of WAEC Mathematics 2019-2021.

Source of variation	Sum of square	DF	Mean square	F	P-value	Remark
Between group	.000	2	.000	.007	.993	NS
Within group	1.239	147	.008			
Total	1.239	149				

Table 9 revealed that the P value of 0.993 is greater than the alpha (α) value of 0.05. Hence, the test statistic is not significant and the null hypotheses is not rejected. This implies that there is no statistical significant difference between item discrimination indices in the WAEC May/June mathematics multiple choice questions administered within 2019 and 2021.

Discussion of Results

Based on the results of the analyses, the discussions were as follows: Research question one on difficulty indices showed that in 2019, five items (10%) out of the fifty items were easy, forty-three items (86%) out of the fifty items were moderately difficult, while two items (4%) out of the fifty items were too difficult. In 2020, six items (12%) out of the fifty items were easy, forty-one (82%) out of the fifty items were moderately difficult while three items (6%) out of the fifty items were too difficult. In 2021, seven items (14%) out of the fifty items were easy, thirty-eight items (76%) out of the fifty items were moderately difficult while five items (10%) out of the fifty items were too difficult. These findings aer in agreement with Cobbinah and Ntumi (2022) who revealed that averagely, the 2020 core mathematics objective test items of the West African Examinations Council in Ghana were within the difficulty level. This study is related to the related to the current study in the areas of subject (mathematics) and theoretical framework but differ in the area of study and sample size.

The findings of research question two on discrimination indices of the test items showed that in the year 2019, twelve items (24%) had excellent discrimination, twenty-two items (44%) had good discrimination, twelve items (24%) had fair discrimination and four items (8%) had poor discrimination indices. Again in 2020, eight items (16%) had excellent discrimination indices, twenty-six items (52%) were with good discrimination indices, eleven items (22%) had fair and five items (10%) had very poor discrimination indices. In 2021, sixteen items (32%) had excellent discrimination indices, sixteen items (32%) were with good discrimination indices, nine items (18%) were with fair discrimination indices, while nine items (18%) had poor discrimination indices. The mean values of the item discrimination for each of the three consecutive years were 0.3389, 0.3332 and 0.3226 respectively. The three values are greater than 0.3, hence the item discrimination for each of the examinations conducted in the three years were good on the average. According to Okoye (2015), items with discrimination indices less than 0.3 should be discarded, replaced or revised, if they will be used in future examinations. This study is in agreement with the findings of Cobbinah and Ntumi (2022) who submitted that the 2020 core mathematics objective test items of the West African Examinations Council discriminated among the examinees.

Conclusion and Recommendation

Item Response Theory (IRT) and related models are important to the practice of educational and psychological measurement because they provide a framework for considering issues and addressing technical problems. Using IRT, the study concluded that most of the items in the 2019, 2020 and 2021 West African Examinations Council (WAEC) May/June Mathematics multiple choice questions were suitable because their difficulty and discrimination indices were within the acceptable range. Items that failed to meet the acceptable level of difficulty and discrimination indices should be eliminated or reviewed before their inclusion in future examinations. Based on the findings of this study, it was recommended that WAEC and other examination bodies should ensure that only quality items are developed for assessment.

References

Abdullahi, I. and Darazo, F.I. (2020). Analyses of psychometric properties of 2016 mathematics Basic Education Certificate Examination Questions (BECEQ) in Gombe State. *Nigerian Journal of Educational Research and Evaluation, 19*, 67-77

Aduloju, M.O. and Okoh, O. (2021). Analysis of item difficulty and discrimination indices as functions of changes in mathematics objective test of the Benue State Examination Board Basic Education Certificate Examination (BECE) from 2015 to 2017. *Nigerian Journal of Educational Research and Evaluation, 20*, 163-181.

Ajeigbe, T. O. and Afolabi, E. R. I. (2013). A study of item response and anchor bias in Economics objective tests among senior secondary school students in Osun State. *Academic Journal of Interdisciplinary Studies, 2*(3), 173-179.

Ashraf, Z.A. and Jaseem, K. (2020). Classeical and modem methods in item analysis of test tools. *International Journal of Research and Review*, *7*(5), 397-403.

Cobbinah, A. and Ntumi, S. (2022). Difficulty, discrimination and pseudo-guessing indices of the West African Examinations Council core mathematics multiple choice items: Practical implications of using item response theory. *Journal Research in Education Sciences, 13*(5), 51-60.

Effiom, A.P. (2021). Test fairness and assessment of differential item functioning of mathematics achievement test for senior secondary students in Cross River State, Nigeria, using item response theory. *Global Journal of Educational Research, 20*, 55-62.

Eleje, L. I., Onah, F. E. and Abanobi, C. C. (2018). Comparative study of classical test theory and item response theory using diagnostic quantitative economics skill test item analysis results. *European Journal of Educational and Social Sciences*, *3*(1), 57-75.

Erguven, M. (2014). Two approaches to psychometric process: Classical test theory and item response theory. *Journal of Education, 2*(2), 23-30

Federal Government of Nigeria (2013). *National policy on Education (4th Ed.).* Lagos: NERDC Press.

Ibrahim B.M. (2023). Psychometric properties analysis of mathematics West African Senior School Certificate Examination in Dala Education Zone, Kano State, Nigeria (2020-2022). *International Journal of Advanced Academic Research, 9*(1), 199-212.

Ikegwuruka, J. (2015). *Item bias in mathematics achievement test of WAEC question paper (2010-2012).* Unpublished Master's Thesis. Department of Physical Science Education, Imo State University, Owerri.

Ikegwuruka, J. (2021). *Effects of test format on students' mathematics achievement and item parameter*. Unpublished Ph.D. Dissertation. Department of Physical Science Education, Imo State University Owerri.

Kasali, J. and Adeyemi, A.A. (2022). Estimation of item parameter indices of NECO mathematics multiple choice test items among Nigerian students. *Journal of Integrated Elementary Education*, *2*(1), 43 – 54.

Kinanee, J., Bosede, J.N. and Orluwene, G.W. (2017). A comparative study on item characteristics of 2014 – 2016 mathematics objective tests in Junior Secondary School of Certificate Examination questions in Rivers State. *International Journal Mathematics Trends and Technology, 52*(8), 545 – 551.

Krishnan, V. (2013). The early child development instrument: An item analysis using classical test theory on Alberta's data. *Early Child Development Mapping Project (ECMap) Alberta*, Community University Partnership, Faculty of Extension, University of Alberta, Canada.

Liu, M. (2017). *Differential Item functioning in large scale mathematics assessment: Comparing the capabilities of the Rasch Tree Method to Traditional approaches*. Unpublished Ph.D Dissertation, Department of Foundations of Education, University of Toledo.

Nkwocha, P.C. (2019). *Basics of education measurement and evaluation (Revised ed.)*. Owerri: Mercy Divine Publishers.

Nworgu, B.G. (2015). *Educational research: Basic issues and methodology* (3rd Ed.). Nsukka: University Trust Publishers.

Okoye, R.O. (2015). *Educational and psychological measurement and evaluation* (2nd ed.). Awka: Erudition Publishers.

Oloda F.S.S. (2021). Gender differential item functioning of test items of 2015 joint senior secondary II mathematics promotion examination in Ondo State, Nigeria. *Global Journal of Human-Social Science*, *21*(7), 64-71.

Oloda, F.S. and Fakindele, G.O. (2017). Self-assessment in mathematics as correlate of performance of students in basic technology and basic science in Ondo State, Nigeria: Implication for curriculum development. *Abacus Journal of Mathematical Association of Nigeria*, *42*(2), 51-64

Osadebe, P.U. and Agbure, B. (2019). Assessment of differential item function in social studies multiple choice questions in Basic Education Certificate Examination. *European Journal of Education Studies*, *6*(8), 312-344

Otoum, A.; Khalaf, H.B.; Bajbeer, A. and Hamad, H.B. (2015). The level of test-wiseness for the students of Arts and Science Faculty at Sharoural and its relationship with some variables. *Journal of Education and Practice*, *6*(29), 102-113.

Sapmaz, Z.M. (2019). *Detection of gender-related differential item functioning (DIF) in the mathematics subjects in Turkey*. Unpublished Master's Thesis, Department of Educational and Psychological Studies, University of South Florida.

Shanmugam, S.K.S. (2020). Gender-related differential item functioning of mathematics computation items among non-native speakers of English. *The Mathematics Enthusiast*, *17*(1), 107-140

Thompson, N.A. (2016). *Introduction to classical test theory with CITAS*. Minnesota: Assessment System Corporation.

Xinming, A. and Yiu-Fai, Y. (2014). Item response theory: What it is and how you can use the IRT procedure to apply it. *Controlled Clinical Trials*, *24*, 1-14.

Yu, C. (2013). A simple guide to the item response theory (IRT) and Rasch modeling. *Journal of Education and practice*, *24* (11), 1-30.