OPEN ACCESS
Journal of Advances in Education and Philosophy
Abbreviated Key Title: J Adv Educ Philos
ISSN 2523-2665 (Print) |ISSN 2523-2223 (Online)
Scholars Middle East Publishers, Dubai, United Arab Emirates
Journal homepage: http://saudijournals.com/jaep/

Original Research Article

Psychometric Properties of 2017 West African Examination Council and National Examinations Council's Economic Senior School Certificate Examination Items

Ogunbamowo A. O[1*], Adediwura A. A[1], R.O Diyan[2]
[1]Department of Educational Foundations and Counselling, Obafemi Awolowo University, 220005, Ife, Nigeria
[2]National Examinations Council, Minna, Nigeria

***Corresponding author:** Ogunbamowo A. O
| **Received:** 10.01.2019 | **Accepted:** 20.01.2019 | **Published:** 30.01.2019
DOI:10.21276/jaep.2019.3.1.2

Abstract

The study determined the dimensionality of WAEC and NECO Economics test items and assessed the difference in each of item discrimination, difficulty and the guessing parameter of the two tests as obtained using CTT and IRT. These were with the view of determining the comparability of the two examinations under different test theories. The research design adopted for the study was descriptive. The population for the study consisted of secondary school students in Osun State and a sample of 540 students. The instruments used for the study were adopted respectively from the 2017 Economics WAEC and NECO Senior School Certificate Examination titled Economics Achievement Test 1 (EAT 1) and Test 2 (EAT 2). The results showed that the difference in the discrimination indices of NECO and WAEC Economics test items when CTT was used is not significant ($U=1.52$, $P > 0.05$). However, there is a significant difference ($U=3.029$, $P < 0.05$) in the discrimination indices when IRT was used. The results also showed that while the difference in difficulty indices of NECO and WAEC Economics test items was not significant with the use of CTT ($U=0.138$, $P > 0.05$), the difference was significant when IRT was used ($U=2.095$, $P > 0.05$). The results further showed that difference in the guessing parameter of NECO and WAEC Economics test items is not significant ($U=1.519$, $P > 0.05$). The results concluded that while the two examinations were comparable under classical test theory, they are not comparable under item response theory.

Keywords: WAEC, NECO, CTT, IRT.

INTRODUCTION

Economics is one of the senior secondary school subjects that require assessment to ascertain students' basic knowledge, skills and understanding of the concepts, and the nature of economic problems in any society. In Nigeria, as well as in most developing countries, Economics is considered as an important subject and is taught at the senior secondary school level. In order to achieve the goals or objectives of Economics at secondary school, the teaching and learning of Economics have to be properly done in the schools. This will equip the student in the fundamental areas of intellectual, vocational, cultural development and national interest. Ochuba [1], pointed out that many students perceive Economics as a very easy subject thus, may not take it seriously and this could be one of the major causes of poor performance of students in Senior Secondary School external examinations. As reported by WAEC Chief Examiner [2], students' performance is on a decline compared to previous years' results in the subject. There is therefore the desire to raise the academic achievement of students in the subject through improved teaching and learning and as such, it is pertinent to determine the psychometric properties of Economics tests by the examination bodies such as WAEC and NECO. To improve the teaching and learning of Economics in Secondary Schools, Obika [3], suggested that efforts should be made to teach Economics in an interesting and lively manner to enable students achieve maximum benefits. One important aspect which should guide any useful and effective teaching and learning of Economics, according to Obika [3] is the determination of the extent to which learning objectives are achieved through a series of assessments.

Assessment is an integral aspect of the education system. Bandele [4], therefore emphasized that "for an assessment system to be valid, it has to be fair to all examinees" and that reliable assessment tools should produce dependable, repeatable and consistent information about people. It should involve both

quantitative and qualitative description of a pupil's behaviour, the passing of value judgment concerning the desirability of that behavior [5]. Examination agencies were set up to promote education, co-ordinate educational programmes, and to monitor the quality of education in educational institutions, the purpose of which is the organization of public examinations so as to provide uniform standards to all test takers, irrespective of the type or method of instruction they have received [6]. West African Examination Councils (WAEC) and National Examinations Council (NECO) are two examining bodies, saddled with the responsibilities of conducting final assessment and evaluation of the final year students in Senior Secondary Schools in Nigeria [7]. The aim of secondary school education in Nigeria, as stated in the National Policy of Education [8], is to prepare the individual child for useful living in the society and for higher education. Since the Senior Secondary School Certificate Examination (SSCE) is the basis for group comparison across the country for the achievement of the National Policy of Education, the psychometric properties of the test items must be of acceptable standard.

Psychometric analysis is the science of measuring latent traits or constructs in our subjects of interests and it imply analyzing the following psychometric properties; Validity (whether a test measures what it is intended to measure), Reliability (the consistency in measuring what it intends to measure), Difficulty (easiness index) and Discrimination Index (how sharply does the test distinguish between low and high ability level of students) [9]. In WAEC and NECO, the analysis of psychometric qualities of their dichotomously scored items are mostly done with classical test theory and its thereafter kept as classified information and can be hardly accessed by the public, researchers or other educational agencies [10]. However, this study is concerned with the measurement of difficulty and discrimination indices. Psychometric theory offers two approaches or methods (Classical Test Theory and Item Response Theory) in analyzing test data. The prediction of psychological tests can be identified using parameters of item difficulty and ability of test takers using CTT and IRT. However, it is important to improve measures of validity and reliability of test items. CTT is based on the assumption that an examinee has an observed score and a true score, and it utilizes measures of item characteristics, item difficulty and item discrimination, the values of which are dependent upon the distribution of examinee proficiency within a sample. The major limitation of CTT can be summarized as circular dependency (a) the person statistics (i.e. observed score) is item sample dependent and (b) the item statistics (i.e. item difficulty and item discrimination) are examinee sample dependent. This circular dependence poses some theoretical difficulty in CTT application in some measurement situations due to the inherent advantages of Item Response Theory, it becomes absolutely compelling that emphases are to shift from Classical Test Theory to Item Response Theory in test analyses. Theoretically IRT overcomes the major weakness of CTT that is the circular dependency of CTT item / person statistics. As a result, IRT models produce item statistics independent of examinees samples and person statistics which are independent of the particular set of items administered. Thus IRT was used in the present study as the bases for analyzing the tests items

Item Response Theory can be divided into two families – unidimensional and multi-dimensional models. While uni-dimensional model measures a single trait or ability dimension, multi-dimensional models measure multiple traits. IRT models are also categorized on the bases of scored responses. The typical multiple choice items are dichotomously scored. Even if there are four or five options they are still being scored as correct or incorrect, right or wrong. A different class of models apply to polytomous outcomes where each response has different score values. An example of polytomously scored items are those rated on a scale of 1-5 or a situation where some number of steps are required to complete a particular assignment. The relationship between examinees performance and the set of traits underlying the item performance can be explained by a monotonically increasing function known as Item Characteristic Curve (ICC) or Item Characteristic Function (ICF) [11]. For items that are dichotomously scored, the ICF can be verified using the one parameter, two parameters and three parameter logistic models. Using these models, the item statistics, the item difficulty (b-parameter), item discrimination (a-parameter) and pseudo guessing, (c-parameter) can be verified for items that are dichotomously scored. The one parameter model (Rasch model) can only verify b, the two parameter model or Birnbaum model can verify b and a; while the three parameter model or Lords models can verify b, a and c, for items that are polytomously scored various models for studying the item statistics exist. This will go a long way to convince the public that the standard of the two examinations, WAEC and NECO are equivalent thereby removing bias and doubt against any of their standard as is sometimes the case.

The role of assessment or test is very vital in evaluating students in the school setting. West African Examinations Council (WAEC) and National Examinations Council (NECO) organize the Senior Secondary Certificate Examination (SSCE) in Nigeria and they are essentially used for certification. The certificates issued by the two bodies are assumed to be equivalent, but there is a concern that students' performance in Economics is not the same for the two examinations. The disparity in performances has been speculated to be due to differences in the psychometric properties of the test items. However, not much have

been done empirically using Item Response Theory (IRT) to validate the speculation as it concerns Economics test item of the two bodies. To this end the study specifically was aimed at:

The objectives of the study were to;

- Determine the dimensionality of WAEC and NECO 2017 Economics examination items
- Determine the difference in discrimination power of WAEC and NECO Economics test items using CTT and IRT;
- Establish the difference in item difficulty of WAEC and NECO Economics test items using CTT and IRT; and
- Determine the guessing parameter in WAEC and NECO Economics test items using IRT

Research Question

What is the dimensionality of WAEC and NECO 2017 Economics examination items?

Hypotheses

- There is no significance difference in discrimination power of WAEC and NECO Economics test.
- There is no significance difference in item difficulty of WAEC and NECO Economics test items.
- The difference in the guessing parameter of WAEC and NECO Economics test items is not significant.

METHODOLOGY

The descriptive survey design was adopted for the study. The descriptive survey has capacity to gather different set of data at the same time from the sample that will be considered for the study to justify current condition and practice and make more plans for improving them. Therefore, this study adopted the descriptive survey design because it sought to obtain information from a representative sample of the population. The population for the study consisted Senior School Students in Osun State. There are 410 Senior Secondary Schools in Osun State. The students' population comprised a total number of 137,083 that comprise of 115,681 from public and 21,402 from private schools with a total number of 69,372 males and 67,711 females.

The study sample consisted of 540 students selected using multistage sampling techniques from three senatorial districts of the state. From each of the three senatorial districts in the State, two Local Government Areas (LGAs) were selected using simple random sampling technique to make a total of six LGAs. From each of the six LGAs that were selected, three schools were selected randomly to make a total of 18 schools. From each school 30 Senior Secondary Two (SSII) Students were selected using convenience sampling technique.

Two adapted instruments titled Economics Achievement Test 1 (EAT 1) and Economics Achievement Test 2 (EAT 2) were used for the collection of relevant data for the study. The EAT 1 was adapted from 2017 WASSCE Economics objective test items while EAT 2 was adapted from SSCE Economics test items conducted by NECO. The items were assumed to be reliable and valid being adapted from standardized tests.

The researcher with the permission of the school Principals and assistance of the Economics teachers in the selected schools administered the EAT 1 and EAT 2 on students offering Economics. The research assistants were teachers from selected schools, with a minimum qualification of Bachelor Degree (B.Sc. Ed. And B.Sc.). The test administration was conducted under strict examination conditions. The duration of data collection for the study lasted for a period of three weeks.

RESULTS

Research Question

What is the dimensionality of WAEC and NECO 2017 Economics examination items?

To assess the number of dimension underlying the two tests, the responses of the examinees to EAT 1 and EAT 2 tests were respectively subjected to nonlinear factor analysis.

Table-1 presents the results of the dimensionality analysis of NECO and WAEC tests respectively.

Table-1: Dimensionality of NECO Economics test

	Dimension	GFI	RMSR criterion	RMSR	DIFF	Reduction	Percentage
	1	0.89	0.180334	0.011675			
NECO	2	0.90		0.011050	0.000625	0.053517	
	3	0.91		0.010553	0.000498	0.045022	4.5
WAEC	1	0.90	0.180334	0.012194			
	2	0.91		0.011394	0.000799	0.0655513	6.6

Table-1 shows that for NECO, when 1-dimension was hypothesized to underlie the data set, the data showed a Goodness of Fit index, GFI > 0.89, the minimum bench mark for which a model is considered acceptably good and the Root mean square of residuals, RMSR, 0.011675 was lesser than the criterion (the

reciprocal of 4 times the square root of the number of students that took the test), indicating a good fit. Taking together, the result showed that NECO economics test violated assumption of unidimensionality. To determine the optimal number of dimensions underlying the test, 2 dimensions were hypothesized to underlie the test data. The result presented in Table-1 shows that GFI value, 0.90 was equal to the minimum bench mark, 0.90 for good fit indication. Furthermore, the RMSR value 0.011675 was lower than the RMSR criterion, indicating a good fit. However, to determine whether there exists another additional dimension underlying the data set, the percentage reduction in the RMSR value from the hypothesized 2 to 3-dimension model was examined. The table showed that the percentage reduction (5.4%) was less than the 10% reduction in RMSR bench mark. These result showed that the test violated the assumption of unidimensionality and that two dimensions underlie the test data. Consequently, the data set was model using Multidimensional item response theory (MIRT) model.

Table-1 also showed that for WAEC, when 1-dimension was hypothesized to underlie the data set, the data showed a good fit (Goodness of Fit index, GFI = 0.90), root mean square of residuals, RMSR, 0.180334 was lesser than the criterion. This result showed that 1-dimension is good enough to explain the variation observed in the performance of the examinees. However, to determine whether additional dimension underlie the test data, 2 dimensions was hypothesized to underlie the data and the 2-dimension model was fitted to the data and compared to the 1-dimension model. Although the GFI for the 2-dimension model was greater than 0.90 and the RMSR was less than the criterion, the percentage in reduction of RMSR from 1-dimension model to 2-dimension model was less than 10%. These results showed that 1 fitted the data better than the 2-dimension model. The result showed that one dominant dimension underlie the performance of the examinees on the WAEC Economics test. The implication of the result is that the Economics test fulfilled unidimensionality assumption. Consequently, the data set was model using the traditional item response theory (IRT) model.

Hypothesis 1: There is no significance difference in each of the discrimination and difficulty index of WAEC and NECO 2017 Economics test items.

To test this hypothesis, the responses of the examinees to the respective test items were subjected to item analysis under IRT item calibration. Unidimensional IRT was used in the calibration of WAEC test and Multidimensional item response theory was used for the calibration of the NECO test. Thereafter, the resulting item parameters of NECO and WAEC under CTT were compared and item parameters of the IRT calibration of the NECO and WAEC test were also compared. The results are presented as follows:

Table-2: Discrimination indices of NECO and WAEC Economics test items

		Discrimination Index				Difficulty Index		
	NECO		WAEC		NECO		WAEC	
Item	a	Remark	a	Remark	b	Remark	b	Remark
1	2.59	Poor	1.73	Good	1.32	Good	1.31	Good
2	0.72	Good	1.18	Good	1.29	Good	2.09	Good
3	1.91	Good	0.70	Good	-0.20	Good	1.51	Good
4	0.95	Good	1.79	Good	1.19	Good	2.03	Good
5	3.22	Poor	0.91	Good	0.08	Good	1.01	Good
6	1.31	Good	1.58	Good	1.70	Good	0.99	Good
7	1.03	Good	1.16	Good	2.96	Good	1.24	Good
8	1.08	Good	1.69	Good	0.86	Good	1.74	Good
9	2.83	Poor	0.57	Good	1.15	Good	0.57	Good
10	0.98	Good	2.16	Poor	1.16	Good	1.52	Poor
11	1.50	Good	0.60	Good	0.89	Good	2.68	Good
12	4.01	Poor	1.39	Good	1.66	Good	1.36	Good
13	1.10	Good	1.55	Good	0.01	Good	2.39	Good
14	1.22	Good	1.50	Good	-0.95	Good	2.19	Good
15	2.07	Poor	1.48	Good	-0.71	Good	1.87	Good
16	1.29	Good	0.37	Good	1.29	Good	15.08	Poor
17	3.06	Poor	0.88	Good	1.41	Good	1.98	Good
18	1.70	Good	0.52	Good	1.34	Good	3.51	Poor
19	2.30	Poor	0.64	Good	0.29	Good	3.88	Poor
20	2.41	Poor	0.99	Good	-1.08	Good	1.68	Good
21	0.14	Good	0.87	Good	48.75	Poor	1.59	Good
22	1.77	Good	1.41	Good	0.57	Good	1.44	Good
23	0.70	Good	1.08	Good	11.01	Poor	1.20	Good
24	1.50	Good	1.94	Good	0.75	Good	0.65	Good

25	3.31	Poor	2.08	Poor	1.90	Good	1.31	Poor
26	4.90	Poor	0.94	Good	1.32	Good	1.11	Good
27	0.39	Good	1.53	Good	2.13	Good	0.96	Good
28	2.70	Poor	1.65	Good	0.34	Good	1.10	Good
29	1.76	Good	0.45	Good	27.38	Poor	3.49	Poor
30	1.40	Good	3.46	Poor	0.88	Good	1.59	Poor
31	0.60	Good	2.76	Poor	3.28	Poor	1.44	Poor
32	1.01	Good	0.24	Good	1.64	Good	6.50	Poor
33	1.84	Good	0.85	Good	0.78	Good	1.77	Good
34	1.36	Good	0.55	Good	2.23	Good	0.38	Good
35	7.56	Poor	3.31	Poor	1.80	Good	1.79	Poor
36	0.80	Good	2.39	Poor	1.06	Good	1.56	Poor
37	10.70	Poor	2.11	Poor	1.31	Good	1.57	Poor
38	1.82	Good	3.56	Poor	3.53	Poor	1.60	Poor
39	1.41	Good	0.32	Good	-0.37	Good	11.47	Poor
40	4.57	Poor	1.95	Good	1.45	Good	1.69	Good
41	4.08	Poor	0.76	Good	1.32	Good	1.81	Good
42	2.76	Poor	2.08	Poor	1.79	Good	1.69	Poor
43	0.98	Good	0.97	Good	3.10	Poor	2.66	Good
44	3.81	Poor	0.75	Good	1.88	Good	0.86	Good
45	1.04	Good	2.96	Poor	1.23	Good	1.48	Poor
46	1.73	Good	0.31	Good	1.28	Good	3.40	Poor
47	1.96	Good	1.80	Good	2.37	Good	1.62	Good
48	0.63	Good	0.43	Good	10.07	Poor	2.70	Good
49	6.38	Poor	0.34	Good	1.71	Good	4.19	Poor
50	1.68	Good	0.98	Good	1.13	Good	1.42	Good
51	1.11	Good			2.30	Good		
52	3.55	Poor			1.16	Good		
53	0.56	Good			56.40	Poor		
54	1.32	Good			1.19	Good		
55	3.11	Poor			2.33	Good		
56	2.09	Poor			1.76	Good		
57	1.87	Good			1.75	Good		
58	2.93	Poor			1.46	Good		
59	1.40	Good			0.65	Good		
60	5.95	Poor			1.17	Good		
$\bar{x}$	2.28		1.36		3.72		2.33	
SD	1.86		0.85		9.95		3.08	

Table-2 showed the discrimination indices of the NECO and WAEC tests items estimated under IRT frameworks. The column labelled "Remark" under IRT frameworks indicated the efficacy of the item discrimination parameter as judged by Hambleton *et al.*, [11] criteria. According to the criteria, items having discrimination estimate outside the range 0 to 2 is considered poor. Table-2 also showed that 10 items representing 20.0% of the 50 WAEC Economics test items were poor while for the NECO 23 items representing 38.3% of the 60 test items were poor. On the average the WAEC test items had a better average discrimination index (Mean = 1.36, SD = 0.85) than the NECO test items (Mean = 2.28, SD = 1.86). To test whether the difference observed in the discrimination of the NECO and WAEC test items was significant Mann-Whitney U test was conducted. The result is presented in Figure-1.

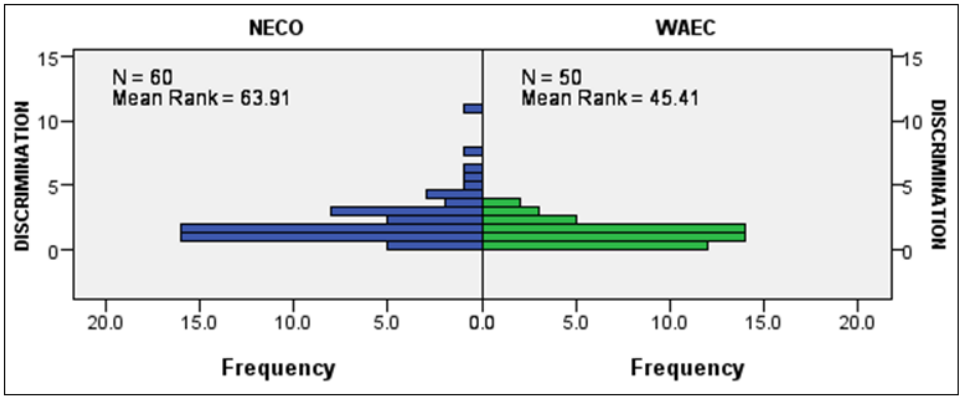

Fig-1: Distribution of discrimination indices of WAEC and NECO test items under IRT

Table-3: Mann-Whitney U test of the NECO and WAEC discrimination indices using IRT

Null Hypothesis	Test	Sig	Decision
The distribution of the discrimination NECO and WAEC test items are same	Independent-samples Mann-Whitney U test 3.029	0.002	Reject the null hypothesis

Table-3 showed that the distribution of the discrimination indices of NECO and WAEC Economics tests items as presented in Figure-1 and Table-3 were different from one another significantly (U=3.029, $p < 0.05$). Hence the Hypothesis which states that "There is no significance difference in discrimination power of WAEC and NECO" was rejected. This implies that the discrimination indices of WAEC and NECO test' items were different from one another.

Table-2 also shows the difficulty indices of the NECO and WAEC Economics tests items. The column labelled "Remark" was used to judge the adequacy of the difficulty indices. The table showed that 18 items representing 36.0% of the 50 items of the WAEC test items were poor while for the NECO test, 8 representing 13.3% of the 60 items of the test were poor. On the average the NECO test items were of more appropriate difficulty level (Mean = 3.72, SD = 0.14) than the WAEC test items (Mean = 2.33, SD = 3.08). To test whether the difference observed in the difficulty of the NECO and WAEC test items was significant Mann-Whitney U test was conducted. The result is presented in Figure-2 and Table-4.

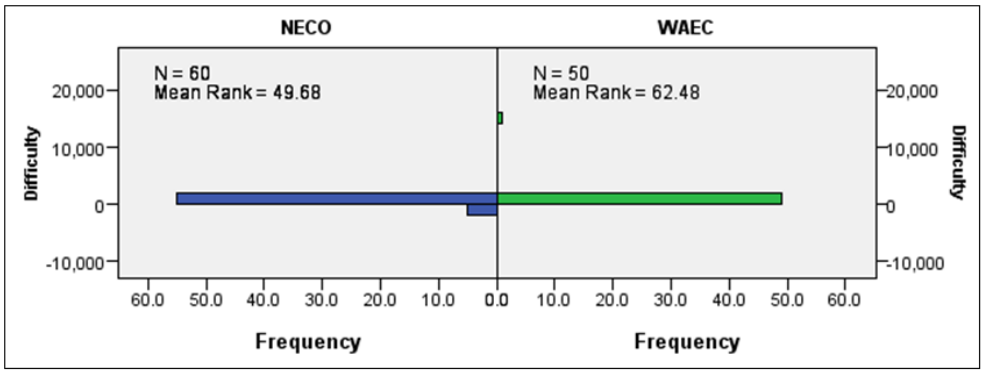

Fig-2: Distribution of Difficulty indices of WAEC and NECO test items under IRT

Table-4: Mann-Whitney U test of the NECO and WAEC difficulty indices under IRT

Null Hypothesis	Test	Sig	Decision
The distribution of the difficulty NECO and WAEC test items are same	Independent-samples Mann-Whitney U test 2.095	0.04	Reject the null hypothesis

Table-4 showed that the distribution of the difficulty parameter of NECO and WAEC tests under CTT presented in Figure-2 were different from one another significantly ($U = 2.095$, $p < 0.05$). Hence the Hypothesis which states that "There is no significance difference in difficulty power of WAEC and NECO" was rejected. This implies that the difficulty indices of WAEC and NECO test' items were different from one another.

Hypothesis 2: The difference in the guessing parameter of WAEC and NECO using IRT is not significant.

Table-5 presents the vulnerability to guessing estimates of the NECO and WAEC test items.

Table-5 showed the vulnerabilities to guessing estimates of NECO and WAEC tests items estimated. The column labelled "Remark" was used to judge the adequacy of the guessing estimates. According to Hambleton *et al.,* [11], items guessing value equal to or greater than 0.35 is considered outrageous. The table shows that all the items of the NECO test had guessing value that were minimal while for the WAEC test, 2 representing 4.0% of the 50 items of the test. On the average the WAEC test items were more vulnerable to guessing ($\bar{x}$ = 0.20, SD = 0.07) than the NECO test items ($\bar{x}$ = 0.18, SD = 0.06). To test whether the difference observed in the vulnerability to guessing of the NECO and WAEC test items was significant Mann-Whitney U test was conducted. The result is presented in Figure-3 and Table-6.

Table-5: Vulnerability to guessing estimates of NECO and WAEC Economics tests' items

	NECO		WAEC	
Items	c	Remark	C	Remark
1	0.13	minimal	0.11	minimal
2	0.15	minimal	0.31	minimal
3	0.16	minimal	0.21	minimal
4	0.20	minimal	0.12	minimal
5	0.16	minimal	0.15	minimal
6	0.20	minimal	0.21	minimal
7	0.17	minimal	0.13	minimal
8	0.11	minimal	0.23	minimal
9	0.23	minimal	0.19	minimal
10	0.13	minimal	0.15	minimal
11	0.11	minimal	0.14	minimal
12	0.31	minimal	0.13	minimal
13	0.14	minimal	0.18	minimal
14	0.18	minimal	0.16	minimal
15	0.16	minimal	0.19	minimal
16	0.12	minimal	0.25	minimal
17	0.23	minimal	0.11	minimal
18	0.12	minimal	0.22	minimal
19	0.10	minimal	0.14	minimal
20	0.22	minimal	0.16	minimal
21	0.16	minimal	0.21	minimal
22	0.13	minimal	0.24	minimal
23	0.15	minimal	0.16	minimal
24	0.15	minimal	0.15	minimal
25	0.31	minimal	0.13	minimal
26	0.09	minimal	0.19	minimal
27	0.20	minimal	0.15	minimal
28	0.15	minimal	0.13	minimal
29	0.14	minimal	0.16	minimal
30	0.14	minimal	0.33	minimal
31	0.17	minimal	0.22	minimal
32	0.23	minimal	0.21	minimal
33	0.27	minimal	0.18	minimal
34	0.13	minimal	0.21	minimal
35	0.10	minimal	0.29	minimal
36	0.17	minimal	0.22	minimal
37	0.21	minimal	0.10	minimal
38	0.23	minimal	0.40	outrageous
39	0.23	minimal	0.28	minimal
40	0.25	minimal	0.23	minimal
41	0.11	minimal	0.21	minimal
42	0.26	minimal	0.29	minimal

43	0.10	minimal	0.16	minimal
44	0.18	minimal	0.19	minimal
45	0.17	minimal	0.36	outrageous
46	0.13	minimal	0.24	minimal
47	0.14	minimal	0.28	minimal
48	0.14	minimal	0.23	minimal
49	0.18	minimal	0.22	minimal
50	0.14	minimal	0.15	minimal
51	0.26	minimal		
52	0.23	minimal		
53	0.21	minimal		
54	0.26	minimal		
55	0.15	minimal		
56	0.28	minimal		
57	0.22	minimal		
58	0.23	minimal		
59	0.23	minimal		
60	0.19	minimal		
$\bar{x}$	0.18		0.20	
SD	0.06		0.07	

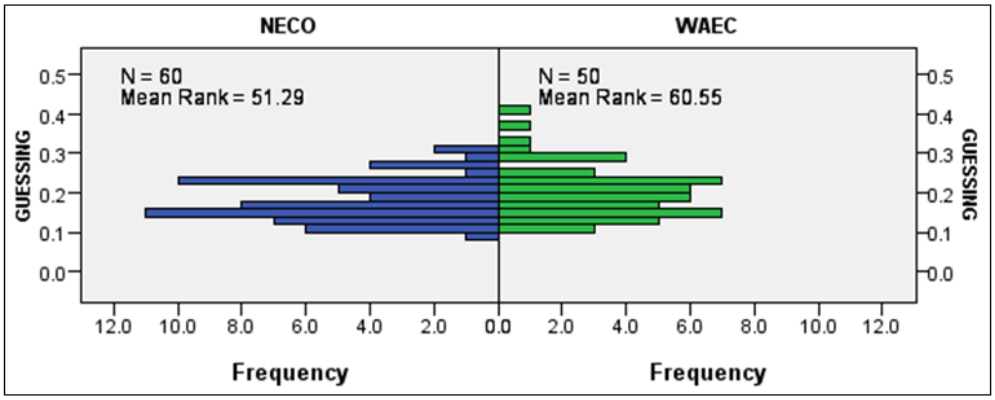

Fig-3: Distribution of Guessing Parameter of WAEC and NECO test items

Table-6: Mann-Whitney U test of the NECO and WAEC guessing parameter

Null Hypothesis	Test	Sig	Decision
The distribution of the vulnerability to guessing NECO and WAEC test items are same	Independent-samples Mann-Whitney U test 1.519	0.13	Do not reject the null hypothesis

Table-6 showed that the distributions of the vulnerability to guessing of NECO and WAEC tests presented in Figure-3 were not significantly different from one another (U = 1.519, p < 0.05). Hence the Hypothesis which states that "The difference in the guessing parameter of WAEC and NECO using IRT is not significant" was not rejected. This implies that the guessing parameter of WAEC and NECO test' items were not different from one another.

DISCUSSION OF FINDINGS

This study determined the difference in discrimination power of 2017 WAEC and NECO Economics test items using IRT. It also established the difference in item difficulty of WAEC and NECO Economics examination items. This study further estimated the guessing parameter of WAEC and NECO Economics items.

The result showed that NECO Economics items violated the assumption of unidimensionality and that two dimensions underlie the test data, however, the WAEC Economics items fulfilled the unidimensionality assumption using IRT model. The findings of the study showed that the discrimination indices of NECO and WAEC Economics items were different from one another significantly, which implies that the discrimination indices of WAEC and NECO Economics were different from one another. On the average the WAEC Economics items had a better discrimination index compared to the NECO Economics items. This finding was similar to the result conducted by [12]. The response of the examinees to the respective test items were subjected to item analysis under IRT item

calibration. Unidimensional IRT was used in the calibration of WAEC test because the items measured a single latent trait and Multidimensional IRT was used for the calibration of the NECO test because the items measured multiple latent traits. The data generated showed that 75% of the test items set by WAEC passed the realistic standard for discrimination index. This result is corresponding to the view of [13]. The findings also revealed that the NECO Economics items were of more appropriate difficulty level compared to the WAEC items. This finding is similar to the findings of Idowu *et al.*, [14], they found out that IRT offers a sound alternative to CTT approach. This is because CTT is based on the process of dependability (an item or test is (sample) dependent) as opposed to measurement, it deals with individual total score and not their ability at the individual item level. Also the finding is similar to the study result of Ojerinde *et al.*, [15], it was seen that the IRT method is sample dependent. In this study, using IRT the difficulty level of WAEC and NECO Economics items were different from one another, this implies that the WAEC Economics items were more difficult compared to the NECO Economics items, however, the items in a test should not be too easy nor too difficult. The finding of this study agrees with Obinne [16] that negative difficulty estimates show that the items are easy while positive difficulty estimates show that the items are hard. The findings which discovered that the items were chosen centered on the b-value, which range of -3 to +3 matches with Baker [17] that hypothetically, difficulty values can range from - 00 to + 00, in practice, difficulty values usually are in the range of - 3 to + 3.

From the findings of this study, it was revealed that NECO Economics items had guessing value 'c' that was minimal. On the average the WAEC Economics items were more vulnerable to guessing compared to the NECO Economics items. The lower c-value indicates that the probability of getting an answer correctly by random guessing is low, the higher c-value indicates that the probability of getting an answer by random guessing is high. This finding is in line with the view of Obinne [16], who opined that WAEC items were more prone to guessing than those of the NECO items.

CONCLUSION

The study based on the findings concluded that the use of CTT is not good enough for the determination of the two examinations item characteristics. It is therefore suggested that WAEC and NECO should endeavor to make use of IRT in the analysis of items right from the process of standardizing the items to the point of making final decision on students' grades after the final examination.

REFERENCES

1. Ochuba O. C. (1994). The development of economics curriculum for secondary school. *Journal of teacher education*, 4(2), 6-10.
2. WAEC Chief Examiner. (2009). West African Examination Council Report 6th issue, January.
3. Obika, T. (1990). The process of development of instrument for assembly economics. *Journal of Science Teacher Today*, 3(1), 11-15.
4. Bandele, S. O. (2004). Educational research in perspective Ibadan, Niyi Commercial and Printing Ventures.
5. Udofia, N. A., & Udoh, K. I. (2017). Comparative Analysis of WAEC and NECO Senior Secondary School Mathematics Examination. *Mathematical Theory and Modelling*, 7(6).
6. Kpolovie, P. J. (2002). Test, Measurement and Evaluation in Education. Port Harcourt: Emhai Printing and Publishing.
7. Bandele, S. O., & Adewale, A. E. (2013). Comparative Analysis of the Item Difficulty Levels of WAEC, NECO and NABTEB Mathematics Achievement Examinations. *Mediterranean Journal of Social Sciences*, 7(2).
8. Federal Republic of Nigeria (FRN). (2004). *National Policy on Education* (4th edition).
9. Adonu, I. I. (2014). Psychometric Analysis of WAEC and NECO Practical Physics Tests Using Partial Credit Model. *PhD thesis. University of Nigeria, Nsukka*.
10. Korashy, A. F. (1995). Applying the Rasch Model to Selection of Items for a Mental Ability Test. *Educational and Psychological Measurement*, 55(5): 753-763.
11. Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory* (Vol. 2). Sage.
12. Olutola, A. T. (2015). Item difficulty and discrimination indices of multiple choice biology tests. *Liceo Journal of Higher Education Research*, *11*(1), 16-31.
13. Ossai-Onah, O. V., Nwobasi, R. N., & Uwa, E. O. (2013). Information needs and seeking behaviour of students in two universities in IMO state, Nigeria. *Library Philosophy and Practice*.
14. Eluwa, O. I., Eluwa, A. N., & Abang, B. K. (2011). Evaluation of mathematics achievement test: A comparison between classical test theory (CTT) and item response theory (IRT). *Journal of Educational and Social Research*, 99.
15. Ojerinde, D., Popoola, K., Ojo, F., & Onyeneho, P. (2012). Introduction to item response theory: Parameter models, estimation and application. *Abuja: Marvelouse Mike Press Ltd*.
16. Obinne, A. (2008). Psychometric properties of senior certificate biology examinations conducted by West African Examinations council and National Examination council: application of item response theory. *Unpublished Ph. D Thesis Nsukka: University of Nigeria Nsukka*.

17. Baker, F. B. (2001). *The basics of item response theory*. For full text: http://ericae.net/irt/baker.