A Comparison of Item Bias of Mathematics Examinations Constructed by WAEC and NECO among Senior Secondary School Students in Ekiti State, Nigeria

Dr. Aborisade, Olatunbosun James
Ekiti State University, Ado-Ekiti
Ekiti State, Nigeria
bosun4success@yahoo.com

Abstract
This study investigated the item bias of mathematics examination constructed by National Examination Council (NECO) and West African Examination Council (WAEC) among difference subgroup of senior secondary school three students in Ekiti State, Nigeria with reference to comparable ability of examinees. The study was a descriptive survey aimed at determining the influence of the independent variables (gender, age, discipline and school location) on the item bias of Mathematics Examinations items by WAEC and NECO. A sample of 600 Senior Secondary School three (SSS III) students was randomly selected from 12 Senior Secondary Schools in Ekiti State, Nigeria using multistage sampling technique. Two Instruments, 50 Mathematics test items each, randomly selected from past WAEC questions and NECO were used to collect data for this study. Four hypotheses formulated were tested at 0.05 probability level. Bilog MG software statistical analysis was used to generate the difficulty level and discriminating power for each NECO and WAEC Examination items. The hypotheses were tested using Analysis of Variance (ANOVA). The result of the study showed that there was no significant difference between the item bias of examination items constructed by NECO and WAEC among different sub-groups of the examinees. It was recommended among others that there should be no discrimination against WAEC or NECO examination, items constructed by these examination bodies could be used freely among different sub-groups of the examinees since showed almost the same level of validity in the construction of their items and area of fairness as well.

Key Words: *item bias, WAEC and NECO*

Introduction

In Nigeria, examination and award of ordinary level certificate are majorly done by the West African Examination Council (WAEC) and the National Examination Council (NECO). These examination bodies construct test items on various subjects offered at the secondary school level (ordinary level) which they administer to students for certification. Students that take this examination are supposed to perform without bias to sex, age, and discipline etcetera. However, candidates who participate in the examinations conducted by these examination bodies are in different settings and therefore differently toned for personal and environmental reasons. As a result of this, the problem of test item bias cannot be ruled out in these examinations. It has been

claimed that some of the national examinations unfairly favour examinees of some particular group (Emaikwu, 2012). Test score validity is of primary importance in a Certification Programme. The National Policy on Education (FRN, 20014) stated that the National Examination tests should be as valid as possible and as fair as possible to all students.

A critical look at the perception of people on such national examination in Nigeria indicates the serious nature of item bias. A test item that is not uni-dimensional is of course not free from bias. If the test makes the members of one group look worse than their attainment would actually be on the job or in the classroom, the test is said to be biased against the group. Therefore, since item bias affects the vital psychometric properties of measurement results in terms of validity and reliability, examination bodies are expected to construct test items in such a manner that items are free from writing errors such as wordiness, irrelevancy, offensiveness, and excessive stimulations, so that when an inadequacy exists between groups examination scores, the discrepancy will be what the test purports to measure in the examinees. When important decisions are made based on test scores, it is critical to avoid bias, which may unfairly influence examinees scores. Bias is the presence of some characteristics of an item that results in differential performance for individual of the same ability but from different gender, location, discipline, age religion etcetera.

Statement of the Problem

There has been criticism against NECO. Some even say its questions are tougher than those of WAEC. Students that take these examinations are supposed to be of comparable abilities location, age, gender etcetera notwithstanding. By item response theory (IRT) Standards, test items should be invariant in nature. More often than not, psycometricians find some items interacting with the characteristics of the examinees. Therefore, the fairness of the examination items constructed by theses examination bodies should be examined for comparison.

Purpose of the Study

The purpose of the study was to examine the comparability of the item bias of the examination items constructed by WAEC and NECO among different subgroup of senior secondary school three students in Ekiti State Nigeria in order to compare the fairness of the Mathematics items constructed by these two major examination bodies.

Research Hypotheses

The following hypotheses were formulated to guide the study and tested at 0.05 level of significance.

1. There is no significant difference between the item bias of male and female students on the examinations constructed by WAEC and NECO.

2. There is no significant difference between the item bias of rural and urban students on the examinations constructed by WAEC and NECO.

3. There is no significant difference between the item bias of younger (below 18 years old) and older (above 18 years old) students on examinations constructed by WAEC and NECO.

4. There is no significant difference between the item bias of Science, Arts and Commercial Students on the examinations constructed by WAEC and NECO.

Methodology

The study employed a descriptive research of the survey type. The population consisted of all the final year students in the public senior secondary school in Ekiti State, Nigeria. A total of 600 students were selected using multistage sampling technique. At the first stage, six local government areas were selected from 16 Local Government Areas in Ekiti State, Nigeria through simple random sampling technique. This was made up of two Local Government Areas from each of the three senatorial districts in the State. At the second stage, 12 public secondary schools were selected from the six local government areas by randomly selecting two public secondary schools from each of the selected local government areas. Fifty students were selected from each of the sampled schools through stratification. Stratified random sampling technique allowed for the stratification of the population into gender (male/female), age (younger/older), discipline (arts, commercial or science) and school location (rural/urban). The instruments used comprised 50 Mathematics test items randomly selected from past WAEC questions and 50 mathematics test items randomly selected from past NECO questions. Experts were asked to examine and check the adequacy of the distribution of the items selected as well as correct the item classification to ascertain face and content validity of the instruments. The instruments were trial tested using 120 students in three secondary schools outsides the sampled schools in Ekiti State. The Kudar Richardson formula 20 (KR_{20}) was used to established a reliability coefficient of 0.75 for the WAEC objective test and coefficient of 0.72 for NECO objective test. The instruments were administered on the 600 sampled students under similar conditions as given by the examination bodies.

Data Analysis

Bilog MG software statistical analysis was used to generate the difficulty levels and discriminating power for each NECO and WAEC Examination items. One of the techniques of detecting items bias is direct comparison of items difficulties (O'Neal, 1991). The hypotheses were tested using Analysis of Variance (ANOVA) by comparing the difficulty levels of the examination items constructed by the two examination bodies. According to Mellenberg (1982) a significant item by group interaction is an indicator that "some of the items are more or less difficult for some groups than others relative to the other items on the test.

Results

Question 1: What are the difficulty levels (b-parameter) and discriminating powers (a-parameter) of mathematics items constructed by WAEC?

Table 1: Summary of the a-parameter and b-parameter of WAEC examination items

Item	a-parameter (discriminating power)	B-parameter (difficulty levels)
1	0.436	4.165
2	0.985	0.948
3	1.121	1.516
4	0.478	0.567
5	0.784	0.824
6	0.118	1.827
7	0.643	0.870
8	0.401	1.367
9	0.533	0.592
10	0.746	0.972
11	0.829	1.047
12	0.248	1.225
13	0.447	0.923
14	0.726	0.617
15	1.203	0.837
16	0.112	0.211
17	0.611	1.905
18	0.516	0.952
19	0.745	1.067
20	0.463	1.860
21	0.841	1.029
22	0.460	1.987
23	0.00	0.000
24	0.658	0.797
25	0.709	1.075
26	0.736	1.604

27	0.517	1.630
28	0.339	1.348
29	0.559	1.796
30	0.419	0.481
31	0.399	1.122
32	0.820	0.794
33	0.466	0.769
34	0.877	1.493
35	0.494	2.801
36	1.263	1.248
37	0.828	0.895
38	0.086	-3.910
39	0.505	1.816
40	0.101	1.620
41	0.254	0.582
42	0.642	1.141
43	0.433	2.487
44	0.371	1.994
45	0.701	1.496
46	0.002	644.185
47	0.702	1.164
48	0.494	1.903
49	0.124	4.051
50	0.708	0.855

Table 1 above shows the results of Bilog MG software statistical analysis used to generate the difficulty levels and discriminating power for each of the WAEC examination items.

Question 2:What are the difficulty levels (b-parameter) and discriminating powers (a-

parameter) of mathematics items constructed by NECO?

Table 2: Summary of the a-parameter and b-parameter of NECO examination items

Item	a-parameter (discriminating power)	b- parameter (difficulty levels)
1	2.162	-0.159
2	0.994	1.291
3	0.427	1.853
4	1.452	0.187
5	1.528	0.404
6	1.023	0.378
7	0.126	2.382
8	0.767	1.584
9	0.414	0.486
10	0.128	5.912
11	0.180	4.295
12	1.2222	0.283
13	0.187	-0.460
14	0.219	6.010
15	0.606	0.111
16	0.664	1.852
17	0.139	3.129
18	0.572	1.886
19	0.152	6.917
20	0.472	1.812
21	0.281	1.799
22	0.510	1.064
23	0.011	48.837
24	1.070	1.737
25	0.962	0.241

26	1.091	-0.081
27	0.773	1.223
28	1.287	1.529
29	0.390	1.270
30	0.641	1.419
31	0.184	0.913
32	0.414	1.431
33	0.843	1.931
34	0.113	2.096
35	0.450	1.379
36	0.005	114.653
37	0.464	1.378
38	0.564	1.818
39	0.099	4.068
40	0.799	0.246
41	0.625	1.470
42	0.148	7.078
43	0.323	1.757
44	1.159	0.307
45	0.185	7.069
46	0.612	0.240
47	0.549	-0.128
48	0.267	6.970
49	0.508	2.395
50	0.354	2.703

Table 2 above shows the results of Bilog MG software statistical analysis used to generate the difficulty levels and discriminating power for each of the NECO examination items.

Hypothesis 1

There is no statistically significant difference between the item bias of male and female students on the examinations constructed by WAEC and NECO.

Table 3: 2X2ANOVA showing item bias of students by gender and type of examination

Source	*SS*	*Df*	*MS*	*F*	*P*
Corrected Model	1996.440[a]	3	665.480	1.085	.357
Sex	4.147	1	4.147	.007	.935
Type of Exam	765.187	1	765.187	1.247	.265
Sex * Type of Exam	1227.106	1	1227.106	2.000	.159
Error	120268.554	196	613.615		
Corrected Total	122264.994	199			
Total	236287.220	200			

P>0.05 (Not Significant)

Table 3 reveals that there is no statistically significant difference between the item bias of male and female students on the examinations constructed by WAEC and NECO ($F_{1,196}$=2.000, P>0.05). The null hypothesis is not rejected. This implies that there is no statistically significant difference between the item bias of male and female students on the examinations constructed by WAEC and NECO.

Hypothesis 2

There is no statistically significant difference between the item bias of rural and urban students on the examinations constructed by WAEC and NECO.

Table 4: 2 X 2 ANOVA showing item bias of students by location and type of examination

Source	*SS*	*Df*	*MS*	*F*	*P*
Corrected Model	6089.629[a]	3	2029.876	2.714	.046
School Location	271.212	1	271.212	.363	.548

Type of Exam	5547.204	1	5547.204	7.418	.007
School Location * Type of Exam	271.212	1	271.212	.363	.548
Error	146576.913	196	747.841		
Corrected Total	152666.542	199			
Total	282527.910	200			

P>0.05 (Not Significant)

Table 4 reveals that the difference between the item bias of rural and urban students on the examinations constructed by WAEC and NECO is not statistically significant at 0.05 level ($F_{1,196}$=0.363, P<0.05). The null hypothesis is not rejected. This implies that there is no statistically significant difference between the item bias of rural and urban students on the examinations constructed by WAEC and NECO.

Hypothesis 3

There is no statistically significant difference between the item bias of younger and older students on examinations constructed by WAEC and NECO.

Table 5: 2 X 2 ANOVA showing item bias of students by age and type of examination

Source	***SS***	***df***	***MS***	***F***	***P***
Corrected Model	6698.136	3	2232.712	4.000	.009
Age	6580.339	1	6580.339	11.788	.001
Type of Exam	64.752	1	64.752	.116	.734
Age * Type of Exam	53.045	1	53.045	.095	.758
Error	109415.655	196	558.243		
Corrected Total	116113.791	199			
Total	205900.060	200			

P>0.05 (Not Significant)

The result in Table 5 shows that there is no significant difference between the item bias of younger and older students on examinations constructed by WAEC and NECO ($F_{1,196}$=0.095, P>0.05). The hypothesis is not rejected. This implies that there is no statistically significant difference between the item bias of younger and older students on examinations constructed by WAEC and NECO.

Hypothesis 4

There is no statistically significant difference between the item bias of Science, Arts and Social Science students on the examinations constructed by WAEC and NECO.

Table 6: 3 X 2 showing item bias of students by discipline and type of examination

Source	*SS*	*df*	*MS*	*F*	*P*
Corrected Model	27551.550	5	5510.310	18.731	.000
Discipline	27156.866	2	13578.433	46.158	.000
Type of Exam	89.424	1	89.424	.304	.582
Discipline * Type of Exam	305.260	2	152.630	.519	.596
Error	86487.154	294	294.174		
Corrected Total	114038.704	299			
Total	176581.521	300			

P>0.05 (Not Significant)

Table 6 reveals that there is no statistically significant difference between the item bias of Science, Arts and Commercial students on the examinations constructed by WAEC and NECO ($F_{2,294}$=0.519, P>0.05). The null hypothesis is not rejected. This implies that there is no statistically significant difference between the item bias of Science, Arts and Social Science students on the examinations constructed by WAEC and NECO.

Discussion

The result of hypothesis one showed that there was no statistically significant different between the item bias of the Mathematics items constructed by WAEC and NECO in relation to gender. The result of the study is in conformity with the findings of Aborisade (2017) that items constructed by NECO are gender bias. However, Adedoyin (2010) in his study on investigating gender in public examination found that out of 16

test items fitted the 3PL item response theory statistical analysis, 5 items were gendered biased.

The result of hypothesis two showed that there was no statistically significant different between the item bias of the Mathematics items constructed by WAEC and NECO in relation to school location. The result of the study is in conformity with the finding of Ajayi (1999) which found out that there was no statistically significant difference between students academic achievement of rural and urban secondary school students. The study however disagree with the findings of Ogbebor and Onuka (2013) in their study that investigated items that are bias using National Examinations Council (NECO) Economics questions for 2010 reported that using Logistic regression statistics, detected items that have DIF against sub group such as urban and rural school students.

The result of hypothesis three showed that there is statistically no significant different between the item bias of younger and older students on examinations constructed by WAEC and NECO. The results of hypothesis three contravenes with the finding of Marquie and Baracat (2000) which reported that in line with the popular belief that the increased age people are less confident and more cautious, it has long been contended that older people show a more conservative response bias than younger ones.

The findings of hypothesis four showed that there is no statistically significant difference between the item bias of Science, Arts and Social Science students on the examinations which is not in conformity with report that Mathematics being compulsory subject in our secondary schools makes it difficult for our students in Arts and Social Science subjects to drop it for external examinations (Madu, 2012).

Conclusion

It is concluded from the findings of this study that items constructed by WAEC and NECO exhibit item bias but there was no significant difference between the item bias of the Mathematics examinations constructed by these examination bodies irrespective of the gender, school location, Age and discipline of the students.

Recommendations

Based on the findings from this study, it is recommended among others that

(i) both WAEC and NECO should ensure that items constructed by them are bias-free by including item bias analysis in the item analysis procedure of their examinations.

(ii) WAEC and NECO could be used freely among different sub-groups of the examinees since the examination bodies have almost the same standard in terms of the fairness of the examinations items constructed by their outfits.

REFERENCES

Aborisade, O. J. (2017). *Identification of gender biased items in the mathematics examinations constructed by West Africa Examination Council (WAEC) in Nigeria*. A Book of Reading

in honour of Prof. Micheal Funso Alonge, Ekiti State University Press, Ado-Ekiti. 375-388.

Adedoyin, O. O. (2010). Using IRT approach to detect gender biased items in public examinations: A case study from the Botswana Junior Certificate Examination in Mathematics. *Educational Research and Reviews 5 (7),* 385-399.

Ajayi, I. A (1999). Unit Cost *of Secondary Education and Student Academic Performance in Ondo State, Nigeria (1991-1995)*. Ph.D Dissertation. University of Ibadan, Ibadan.

Awoniran, C. O. (1987): Identification of Biased Items in West African School Certificate Biology Achievement Test. An unpublished M.Ed thesis of Obafemi Awolowo University, Ile-Ife.

Doolittle, A. E & Cleary, T. A (1987). Gender-based Differential Item Performance in Mathematics achievement items. *Journal of Educational Measurement* 24(2), 157-166

Emaikwu, S. O. (2012). Issues in Test Item Bias in Public Examinations in Nigeria and Implications for Testing. *International Journal of Academic Research in Progressive Education and Development* 1 (1) 175-187.

Federal Republic of Nigeria (2004): National Policy on Education, 4th ed., Lagos, NERDC.

Hessen, D.J. et al (2012).Age and Gender Differences in Depression across adolescence: Real or Bias?.*Journal of Child Psychology and Psychiatry*, Blackwell Publishing,

Madu, B.C (2012). Analysis of Gender-Related Differential Item Functioning in Mathematics Multiple Choice Items Administered by West African Examination Council (WAEC). *Journal of Education and Practice* 3(8); 71-78

Mellenberg, G. J (1982). *A Posteriori analyses of biased items. In R.A Berk(Ed). Handbook of methods for detecting test bias.* Baltimore, M.D. The Johns Hopkins University Press.

Obinne, A. D. E., Nworgu, B. G., &Umobong, M. E. (2013). An Investigation into differential item iunctioning of tests conducted by the two major examination bodies in Nigeria. *Advances in Educational Research 2 (1),* 001-008.

Ogbebor, U & Onuka, A. (2013). Differential Item Functioning method as an item bias indicator. *Educational Research, 4(4)*; 367-373.

O'Neal, M.R.(1991). A Comparison of Methods for Detecting Item Bias /http://eric.ed.gov.