Evaluation of Item Bias Using Differential Item Functioning (DIF) Technique in NECO Conducted Economics Examination in Taraba State, Nigeria

[1]Agi Christiana Ikpoyi, [2]Hager Atisi Eremina Amakiri & [3]Amuche Blessing Ehi

[1]Department of Educational Foundations and General Studies, Joseph Sarwuan Tarka University, Makurdi Benue State Nigeria,

[2]Department of Educational Psychology, Guidance and Counselling, Ignatius Ajuru University of Education, Port Harcourt, River State, Nigeria,

[3]Department of Social Science Education, Taraba State University, Jalingo

ABSTRACT

Differential item functioning is an approach that is widely used to find out items that are bias. This study investigated items that are bias using differential item functioning approach in relation to gender, school ownership (private and public schools) and school location (urban and rural schools) using National Examinations Council (NECO) Economics questions for June/July 2022/2022. The research design employed in this study was exposit facto research design. The study sample comprised students in Taraba State, Nigeria. One Hundred (100) students were used for study. And the test contains 60 items which was administered to the students. Logistic regression was used to analysis the data. The research findings showed that out of sixty items in NECO Economics questions 11 items were biased in relation to gender, 7 items were biased in relation to school ownership and 9 items in relation to school location. The implication of these findings is that NECO economics examinations questions have presences of differential item functioning (DIF). From the result of the findings, it was then recommended that test experts and developers should explore the use of DIF approach to detect biased items.

Keyword: Differential item functioning: item bias: Economics, Examination, NECO

INTRODUCTION

In Nigeria, there exist several national examination bodies, and they include National Examination Council (NECO), West African Examination Council (WAEC), National Business and Technical Examination Board (NABTEB), and Joint Admission Matriculation Board (JAMB). These bodies cater for candidates of various backgrounds all over the country. 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, the problem of test item bias cannot be ruled out in these examinations.

A test item that is not unidimensional is not free from bias. For example, two items designed to assess multiplication skills in Economics could be as follows: (i). what is 6×7? (ii). what is the product of six and seven? Item (i) requires only knowledge in mathematical operations, while item (ii) requires, for its solution,

a specific amount of reading competence as well as knowledge of mathematical operations. When different attributes are being measured as in item (ii), the issue of item bias enters into consideration if such item is administered to two different groups and the responses of one of the groups are dependent on the secondary skill. This type of item measures different types of skills among different groups. If the test makes the members of one group look worse than their attainment on the job or in the classroom, the test is said to be biased against that group. The same notion of bias is applied to school achievement tests. For instance, children in one group consistently receive lower scores than would be expected from their observed classroom performance.

The procedures employed in the administration of the national examination could be sources of bias. The actual administration of the examinations constitutes a complex interaction among examiners' variables, examinees' variables and situational variables. In Nigeria, the concept of test item bias is among the topical issues of concern and has become a daily issue of national discourse even at the legislative assembly(Anastasia &Urbina, 2014). During the post unified tertiary matriculation examination exercise, many candidates often complain of bias in the testing process. Some tertiary institutions are accused of setting 'local and irrelevant' questions extraneous to candidates' areas of specialization.

Bias can result in systematic errors that distort the inferences made in any selection and classification. As mentioned earlier, several examination bodies exist in Nigeria, and these bodies cater to candidates of various backgrounds all over the country. 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, the problem of test item bias cannot be ruled out in these examinations. It is expedient that the examining bodies examine the degree of bias in their examinations. It has been claimed that some of the national examinations unfairly favour examinees of some particular groups, e.g., cultural or linguistic groups, to the extent that it is now believed that a particular section of the country performs most woefully in these national examinations(Emaikwu,2015). A critical look at people's perception of such a national examination in Nigeria indicates the serious nature of item bias.

According to Scheuneman (2014), a test item is described as differentially functioning when the probability of correct response is not the same for all persons of a given ability irrespective of their group membership. No individual or group answering a question should be disadvantaged in any way. Construction of assessment instruments needs to be free of bias so that students of equal ability drawn from the same population but belonging to different subgroups such as male or female, urban or rural students will have the same probability of getting an item correct. This is tampered with when an item is biased. A Test item that differentially affects groups or individuals from showing their true abilities and thereby measuring irrelevant construct is a bias test item. This type of bias item that exhibits differential item functioning (DIF) systematically underestimates or overestimates the value of the variable the items are designed to measure (Reynolds, 2014).

There are two types of DIF, which are uniform and non-uniform DIF. Uniform DIF occurs when differences in correct response probability are found across all ability levels for a particular item. Non-uniform DIF, on the other hand, occurs when there is an interaction between the ability and group membership. An item may seem difficult for those at the higher level in one group. After a particular point, it becomes more difficult for those at a lower level in the other group. It is obvious, therefore that, the most central examination being administered to Nigerian students may not be fair to one group or the other if methods that will refine test items devoid of gender, location, ownership, school type and career pathway biases are not put into consideration, especially differential item functioning (DIF) methods It has been claimed that some of the national examinations unfairly favour examines of some particular groups, e.g. Cultural or linguistic groups to the extent that it is now believed that a particular section of the country performs most woefully in these national examinations (Garoa, 2015). A critical look at people's perception of such a national examination in

Nigeria indicates the serious nature of item bias.

No Justification to analyze for differential item functioning arises out of manifold considerations, among which but not limited to the fact that literature hinted that researchers mostly emphasize DIF studies without being mindful about the forms and effect sizes of DIF. Secondly, test decisions regarding certification should only measure this common knowledge and skills and should not be affected by test-takers location, gender of career pathway. Sequel to these problems of item bias or differential item functioning, this research works to study NECO 2021 SSCE Economics multiple-choice test items for possible differential item functioning.

STATEMENT OF THE PROBLEM

The problem of testing not providing equal opportunity for examinees has been created as a result of the test items functioning differentially. There is no question that education is a key element in improving the lives of children, families, communities and nations. When some examinees are failing while some are passing as a result of the difficulty posed by the test items, it has distorted the chance of those who failed to be promoted. This has proved the notion of how DIF could be harmful and threatening.

There is also the problem of locational differentiation. Location differentiation in this context deals with dividing students in the schools into groups based on the school location such that they have certain economic or/and social characteristics in common. The focus of education is to bridge the gap between groups. This could lead to the purpose of education being subjugated as a result of the threat being posed by the effect of DIF. The presence of DIF coincides with differential drop out in schools since the test items are proving difficult to the examinees. Examinees' failure may not be because of their inability to answer the items correctly but because of the unfairness of the test; it can result in many examinees withdrawing out of school. According to Odili (2010), the interest in the analysis of differential item functioning in test derives from the consideration that education is perceived as an instrument for achieving equity among persons. In achieving this requires test items to measure traits which are taught in schools and not those that are foreign to it.

Results of NECO may have test items that are not free from item bias which might be causing misleading decisions at the policy level as policymakers review educational policies, change their curriculum, teaching methods and assessment methods based on the results of these external examinations. Such uses or decisions are valid only to the extent that the items in the test are bias-free. However, some examination bodies do not include item bias detection in their item analysis. Could this be the case with the test items constructed by WAEC and NECO? Therefore, analysis of item bias of the items constructed by WAEC and NECO should be done to ascertain the level of validity of the examination items. Therefore, this study analyzed the probability of correct response for examinees and the nature of bias encountered and to ascertain the reliability of the 2021/2022 NECO SSCE Economics multiple-choice test.

Purpose of the Study

The purpose of the study was to examine Evaluation of Item Bias Using Differential Item Functioning (DIF) Technique in NECO Conducted Economics Examination in Taraba State

Specifically, the study is poised to determine:

1. Whether items set by NECO Economics multiple-choice examination in 2021/2022 June/July functions differently regarding gender.
2. Whether items set by NECO Economics multiple-choice examination in 2021/2022 June/July

functions differently regarding school location.
3. Whether items set by NECO Economics multiple-choice examination in 2021/2022 June/July functions differently regarding school type.

Research Questions

Based on the objectives of this research study, the study will answer the following research questions:

1. which items on NECO 2021/2022 June/July SSCE multiple-choice Economics examination functions differentially to gender?
2. which items on NECO 2021/2022 June/July SSCE multiple-choice Economics examination functions differentially to school location?
3. which items on NECO 2021/2022 June/July SSCE multiple-choice Economics examination function differentially to school ownership?

Research Hypotheses

The research hypotheses formulated to guide verification of findings. The null hypotheses were tested at significance level.

H_{O1}: NECO 2021/2022 June/July SSCE multiple-choice Economics Examination does not significantly functions differentially on gender

H_{O2}: NECO 2021/2022 June/July SSCE multiple-choice Economics Examination does not functions differentially to school location

H_{O3}: NECO 2021/2022 June/July SSCE multiple-choice Economics Examination does not function differentially to school ownership.

LITERATURE REVIEW

Concept of Test Bias

The issue of fairness is what critics labeled "bias" in testing. When the whole test is the unit of concern, then "test bias" is the issue to be examined, whereas when an individual item is the unit of concern, then "item bias" is the concept of focus. A more decorated term for item bias has been formulated, namely "differential item functioning". Differential item functioning (DIF) occurs when examinees from different groups show differing probabilities of success on the item after matching the underlying ability that the item intends to measure (Zumbo, 2014). Item bias occurs when examinees of one group are less likely to answer an item correctly than examinees of another group because of some characteristic of the test item or testing situation that is not relevant to the test purpose. Within the Item Response Theory (IRT) framework, DIF may be interpreted as follows: When an item is classified as presenting uniform DIF, the difficulty parameter changes but discrimination is the same (Camilli and Shepard, 1994). This may be seen as evidence that an irrelevant dimension is being tackled by the item and the groups differ in the distribution of this dimension (Walker, 2015). When an item is classified as presenting non uniform DIF, the difficulty parameter is the same but the discrimination is not. In this case, the interpretation of DIF would imply that the variance of the groups on irrelevant dimension is not the same or the correlation between both dimensions is different between the groups (Walker, 2015). Logistic regression has a good control on Type I error under conditions similar to those manipulated in this study. An example of hypothetical item bias is illustrated in the diagram shown thus:

Methods of Detecting Item Bias

The usual thing is to statistically detect biased items and then send them to the bias revision panel that examines them to determine if they are biased. Several statistical methods can be used to detect bias items, namely:

Item Discrimination Index

This is done by finding the discrimination index of the item for both groups. If the discrimination indexes are approximately equal, then the item is probably not biased, but if the values are not approximately equal, such items could be biased.

Factor Analysis

It can be used to evaluate the internal structure separately for the two groups. If only one factor is found in each group, then the test does not contain bias items, but if more than one factor is found in one of the groups, the test is biased.

Rank Order

This is a quick method. Here the test items are ranked in order of difficulty for each of the two groups. If the item rank differs across groups, the test is suspected to be biased.

Differential Item Functioning (DIF)

As a term, DIF statistics describe a growing and evolving collection of statistical techniques useful for detecting a systematic difference in performance by sub groups of the population at both test and item levels. There is no single strategy for investigating DIF; rather, DIF is a psychometric nomenclature described as an assemblage of empirical, statistically-based techniques targeted at a narrow but important part of item investigation. Additionally, DIF neither indicates the direction of detected differences nor connotes causation.

Differential Item Functioning (DIF) refers to differences in the functioning of items across groups, often demographic, which are matched on the latent trait or, more generally, the attribute being measured by the items or test (Camilli, 2014).

Theoretical framework

A test can be studied from different perspectives, and the items in the test can be evaluated according to different theories. Two such theories are the Classical Test Theory (CTT) and the Item Response Theory (IRT). These theories are the two major frameworks used in educational measurement to develop, evaluate and study test items. These frameworks are based on different assumptions and use different statistical approaches. This study is anchored on Item Response Theory (IRT).

Empirical studies

Amaechi, and Onah (2018) detected uniform and non-uniform gender differential item functioning in Economics multiple choice standardized test in Nigeria. One research question and one hypothesis guided the study. The design of this study is a survey which involved the inferential method. The population of the study was 4,434,979 secondary school students in the 11,875 public secondary schools in the 36 states

including Abuja the Federal Capital Territory (FCT) of Nigeria involving purposive and simple random sampling techniques. The instruments for data collection were Socio-Demographic Inventory (SDI) and a 50 item WAEC General Economics Paper I Multiple Choice Test. The instruments were revalidated by five specialists, three in Educational Measurement and Evaluation and two from Economics Education from Michael Okpara University of Agriculture Umudike and Imo State University, Owerri. The reliability of the test was reestablished using Kuder-Richardson formular 20 (KR-20) statistics with an index of 0.80. In answering the research question, IRT-Binary Logistic Regression method was used while the hypotheses were tested using Wald test associated with binary logistic regression statistics at 0.05 level of significance. The result indicated that, out of the 13 items that have DIF issues, 8 fall under the category of uniform DIF while 5 were non-uniform DIF. The uniform DIF items are 13, 15, 20, 24, 37, 39, 42 and 46; while the non-uniform DIF items are 9, 22, 30, 37 and 44. This shows that 8 items significantly displayed uniform DIF while 5 items significantly displayed non-uniform DIF. The current study relates to the study in review because both investigate item bias using DIF. The study differs in terms of method of data analysis.

Eteng-Uket (2017) investigated detecting differential item functioning using item response theory in West African Senior School Certificate English language test in south-south Nigeria. 2 research questions were formulated to guide the study. Using descriptive research survey design for the study, study population was 117845 Senior Secondary 3 students in Edo, Delta, Rivers and Bayelsa state. A sample of 1309 (604 males, 705 females) drawn through multi stage sampling technique was used for the study. Two valid instruments titled: Socio-economic status questionnaire (SSQ) and WASSCE/SSCE English language objective test (ELOT) were used to collect data for the study. The reliability indices of the instruments were estimated using the Cronbach Alpha method of internal consistency and Richard Kuderson 20 with coefficient values of .84 for the English Language objective test and .71 for the socio-economic status questionnaire respectively. Chi-square and Lord Wald test statistics statistical technique employed by Item Response Theory for Patient Reported Outcome (IRTPRO) was the technique used in data analysis which provided answers to the research questions at.05 level of significance. On analysis, the result revealed that 13 items functioned differently significant between the male and female group and significantly 23 items differentially functioned between High and low socio-economic status groups. Thus, this shows 18% status indicating large DIF and items that are potentially biased. Based on the findings, the current study relates to the study in review because both investigate item bias using DIF. The study differs in terms of method of data analysis.

MATERIAL AND METHODS

In order to achieve the objectives of this study, the study employed the ex-post facto research design. The population of this study consists of sixty-five thousand eight hundred and ninety-nine (65,899) candidates that registered and sat for Economics multiple-choice items of NECO SSCE June/July examination (2021/2022) distributed across fourteen education zones in Taraba State. Multi-stage sampling procedure with appropriate technique was used to draw 100 students to same as sample for the study. The sampling was done in stages as follows. In the first stage, two (02) local governments were randomly selected from three educational zones making six (06) local governments. In the second stage, two (02) of each schools randomly selected from each local government making (12) schools respectively. For the third stage, systematic sampling was used in selecting the candidates for the study. Stage One: large groups or clusters are identified and selected. These clusters contain more population units than are needed for the final sample. Stage two: population units are picked from within the selected cluster (using possible probability sampling methods) for a final sample.

The NECO 2021 Economics multiple-choice test (NEMT) was the instrument for data collection. The students were provided with an answer sheet by the researcher. The experts were requested to check both the face and content validity of the instrument and to rate the instrument. The reliability of an instrument is

indicative of its consistency or the relatedness of items within a measure. For this study, the reliability of NEMT was analyzed using the Split-half method and the results of the reliability analyses revealed that the Spearman-Brown Coefficient (Equal Length) was .85. Logistic regression was used to analyze the data. It involved the following steps: Identify Reference and Focal groups of interest usually two at a time. Design the DIF study to have samples which are large as possible. Choose DIF statistics which are appropriate for the data. Carry out the statistical analyses. Interpret DIF statistics/results and delete items or make item changes as necessary.

Research Question One

Which items on NECO 2021/2022 June/July SSCE multiple-choice Economics examination functions differentially to gender?

Table 1 Logistic Regression to Detect Gender Bias

Item	B	S.E	Sig	Exp (B)	Lower
1	0.157	0.225	0.483	1.17	0.754
2	0.243	0.238	0.308	1.275	0.799
3	0.095	0.19	0.616	1.1	0.758
4	-0.076	0.19	0.691	0.927	0.639
5	-0.235	0.231	0.309	0.791	0.503
6	0.311	0.211	0.142	1.364	0.902
7	-0.177	0.19	0.454	. .837	0.577
8	-0.339	0.191	.004*	0.712	0.49
9	0.417	0.195	0.343	1.517	1.035
10	0.92	0.197	0.639	1.097	0.746
11	0.242	0.218	0.268	1.273	0.831
12	-0.227	0.19	.033*	0.797	0.549
13	0.663	0.201	0.531	1.941	1.31
14	1.039	0.361	.004*	2.826	1.393
15	0.249	0.202	0.219	1.283	0.863
16	-0.959	0.266	.000*.	0.383	0.227
17	-0.023	0.191	0.905	0.977	0.672
18	-0.319	0.191	0.094	0.727	0.5
19	0.241	0.199	0.226	. 1.272	0.861
20	0.317	0.193	0.101	1.373	0.941
21	0.163	0.247	0.509	1.177	0.725
22	0.164	0.354	.001*	0.897	0.6785

23	-0.543	0.307	0.077	0.581	0.318
24	0.218	0.261	0.402	1.244	0.747
25	-0.494	0.325	0.129	0.61	0.323
26	-0.131	0.202	0.507	0.877	0.59
27	0.083	0.196	0.672	1.087	0.74
28	-458	0.266	0.085	0.632	0.375
29	-0.111	0.271	0.682	0.895	0.527
30	0.046	0.19	0.808	1.047	0.721
31	0.299	0.197	0.129	1.349	0.916
32	0.122	0.256	0.635	1.129	0.683
33	0.166	0.191	0.386	1.181	0.811
34	-0.141	0.216	0.513	0.868	0.568
35	0.204	0.198	0.29	1.233	0.836
36	0.242	0.223	0.278	1.273	0.823
37	-0.14	0.201	0.486	0.869	0.587
38	0.374	0.287	0.192	1.454	0.829
39	0.257	0.201	0.202	1.293	0.871
40	-0.326	0.198	0.1	0.722	0.489
41	0.086	0.191	0.653	1.89	0.75
42	0.136	0.278	0.626	1.145	0.664
43	-1.488	0.459	.001*	0.226	0.092
44	0.46	0.218	.034*	1.585	1.035
45	0.065	0.215	0.761	1.068	0.7
46	0.461	0.201	.021*	1.586	1.07
47	-0.209	0.28	0.455	0.811	0.469
48	0.263	0.207	0.203	1.301	0.867
49	0.414	0.191	.031*	1.513	1.039
50	-0.506	0.228	.027*	0.603	0.386
51	0.103	0.272	0.705	1.109	0.65
52	0.106	0.245	0.666	1.112	0.688
53	0.134	0.216	0.536	1.143	0.749
54	-0.071	0.193	0.711	0.931	0.638
55	-0.161	0.207	0.437	0.851	0.567

56	0.255	0.248	0.305	1.29	0.793
57	0.168	0.211	0.425	1.183	0.783
58	-0.014	0.246	.007*	0.986	0.609
59	0.564	0.207	0.677	1.758	1.171
60	-0.06	0.195	0.76	0.942	0.643

Table 1 shows the items in relation to gender (Male and female), identified by logistic regression. Out of sixty items in NECO Economics questions DIF was present in eleven items. These items are item 8, 12, 14, 16, 22, 43, 44, 46, 49, 50, and item 58.

Research Question Two

Which items on NECO 2021/2022 June/July SSCE multiple-choice Economics examination functions differentially to school location (rural and urban)?

Table 2: Logistic Regression of Sixty NECO Item for School Location

Item	B	S.E	Sig	Exp (B)	Lower	Upper
1	0.017	0.243	0.965	1.017	0.55	1.76
2	-0.236	0.246	.000*	0.788	0.512	1.28
3	1.22	0.2	0.34	3.388	2.29	5.012
4	-0.403	0.191	0.35	0.669	0.46	0.973
5	0.194	0.233	0.406	1.214	0.769	1.917
6	-0.217	0.209	.002*	0.805	0.535	1.212
7	-0.84	0.194	0.254	0.432	0.295	0.632
8	-0.339	0.191	0.075	0.712	0.49	1.035
9	-0.618	0.199	.002*	0.539	0.365	0.796
10	-0.37	0.197	0.059	0.69	0.469	1.015
11	-0.506	0.217	.019*	0.603	0.394	0.921
12	0.098	0.19	0.604	1.103	0.76	1.602
13	0.107	0.199	0.591	1.113	0.754	1.643
14	-0.254	0.315	0.419	0.776	0.419	1.437
15	-0.116	0.201	0.562	0.86	0.6	1.32
16	-0.432	0.249	0.084	0.65	0.398	1.059
17	-0.611	0.193	.002*	0.543	0.372	0.793
18	0.37	0.191	0.053	1.447	0.995	2.105
19	0.122	0.198	0.538	1.13	0.766	1.667
20	-0.017	0.193	0.928	0.983	0.674	1.434
21	-0.332	0.252	0.188	0.717	0.438	1.176
22	0.424	0.234	0.07	1.528	0.966	2.417
23	0.08	0.293	0.785	1.083	0.61	1.925
24	-0.333	0.266	0.212	0.717	0.425	1.209
25	-0.087	0.314	0.781	0.916	0.495	1.695
26	0.276	0.202	0.171	1.318	0.888	1.957

27	0.199	0.196	0.311	1.22	0.83	1.792
28	0.02	0.258	0.938	1.02	0.615	1.693
29	0.324	0.269	0.228	1.382	0.816	2.341
30	-0.316	0.191	0.097	0.729	0.502	1.059
31	0.029	0.191	0.883	1.029	0.701	1.511
32	-0.143	0.259	0.58	0.867	0.522	1.439
33	-0.054	0.191	.012*	0.948	0.651	1.379
34	-0.329	0.218	0.131	0.72	0.47	1.103
35	0.209	0.198	0.29	1.233	0.836	1.817
36	0.093	0.223	0.678	1.097	0.706	1.698
37	0.181	0.2	0.366	1.198	0.81	1.773
38	0.54	0.289	0.062	1.715	0.975	3.02
39	-0.068	0.202	0.737	0.934	0.629	1.388
40	0.333	0.197	0.091	1.395	0.948	2.05
41	-0.133	0.191	0.487	0.876	0.602	1.273
42	0.213	0.278	0.447	1.237	0.718	2.132
43	-0.644	0.377	0.088	0.525	0.251	1.1
44	-0.153	0.218	0.483	0.858	0.559	1.316
45	-0.405	0.219	0.065	0.667	0.434	1.025
46	-0.1	0.201	0.619	0.905	0.611	1.341
47	-1.069	0.314	.001*	0.343	0.185	0.636
48	-0.08	0.208	0.701	0.923	0.614	1.387
49	-0.463	0.192	0.061	0.629	0.432	0.918
50	-0.251	0.224	0.263	0.778	0.502	1.207
51	-0.195	0.276	0.476	0.822	0.479	1.411
52	-0.575	0.255	.024*	0.563	0.341	0.929
53	-0.053	0.216	.002*	0.948	0.621	1.449
54	0.598	0.194	0.365	1.819	1.244	2.661
55	0.223	0.206	0.28	1.249	0.834	1.871
56	-0.054	0.25	0.827	0.947	0.581	1.544
57	0.124	0.211	0.558	1.132	0.749	1.71
58	-0.258	0.248	0.299	0.773	0.475	1.257
59	0.265	0.206	0.198	1.304	0.87	1.953
60	– .2 51	0.196	0.201	0.778	0.53	1.143

Table 2 shows the items in relation to school location (rural and urban), identified by logistic regression method using. Out of sixty items in NECO Economics questions DIF was present in nine items. These items are item 2, 6, 9, 11, 17, 33, 47, 52 and 53.

Research Question Three

Which items on NECO 2021/2022 June/July SSCE multiple-choice Economics examination function differentially to school ownership (public and private)?

Table 3: Logistic Regression of Sixty NECO Item for School Ownership

Item	B	S.E	Sig	Exp (B)	Lower	Upper
1	0.017	0.243	0.965	1.017	0.55	1.76
2	-0.236	0.246	0.292	0.788	0.512	1.28
3	1.22	0.2	.002*	3.388	2.29	5.012
4	-0.403	0.191	0.35	0.669	0.46	0.973
5	0.194	0.233	0.406	1.214	0.769	1.917
6	-0.217	0.209	0.732	0.805	0.535	1.212
7	-0.84	0.194	0.254	0.432	0.295	0.632
8	-0.339	0.191	.003*	0.712	0.49	1.035
9	-0.618	0.199	0.483	0.539	0.365	0.796
10	-0.37	0.197	.000*	0.69	0.469	1.015
11	-0.506	0.217	0.019	0.603	0.394	0.921
12	0.098	0.19	0.604	1.103	0.76	1.602
13	0.107	0.199	0.591	1.113	0.754	1.643
14	-0.254	0.315	0.419	0.776	0.419	1.437
15	-0.116	0.201	0.562	0.86	0.6	1.32
16	-0.432	0.249	0.084	0.65	0.398	1.059
17	-0.611	0.193	0.502	0.543	0.372	0.793
18	0.37	0.191	0.053	1.447	0.995	2.105
19	0.122	0.198	0.538	1.13	0.766	1.667
20	-0.017	0.193	0.928	0.983	0.674	1.434
21	-0.332	0.252	0.188	0.717	0.438	1.176
22	0.424	0.234	.001*	1.528	0.966	2.417
23	0.08	0.293	0.785	1.083	0.61	1.925
24	-0.333	0.266	0.212	0.717	0.425	1.209
25	-0.087	0.314	0.781	0.916	0.495	1.695
26	0.276	0.202	0.171	1.318	0.888	1.957
27	0.199	0.196	0.311	1.22	0.83	1.792
28	0.02	0.258	0.938	1.02	0.615	1.693
29	0.324	0.269	.002*	1.382	0.816	2.341
30	-0.316	0.191	0.097	0.729	0.502	1.059
31	0.029	0.191	.003*	1.029	0.701	1.511
32	-0.143	0.259	0.58	0.867	0.522	1.439
33	-0.054	0.191	0.489	0.948	0.651	1.379
34	-0.329	0.218	0.131	0.72	0.47	1.103
35	0.209	0.198	0.29	1.233	0.836	1.817
36	0.093	0.223	0.678	1.097	0.706	1.698
37	0.181	0.2	0.366	1.198	0.81	1.773
38	0.54	0.289	0.062	1.715	0.975	3.02
39	-0.068	0.202	0.737	0.934	0.629	1.388

40	0.333	0.197	0.091	1.395	0.948	2.05
41	-0.133	0.191	0.487	0.876	0.602	1.273
42	0.213	0.278	0.447	1.237	0.718	2.132
43	-0.644	0.377	0.088	0.525	0.251	1.1
44	-0.153	0.218	0.483	0.858	0.559	1.316
45	-0.405	0.219	0.065	0.667	0.434	1.025
46	-0.1	0.201	0.619	0.905	0.611	1.341
47	-1.069	0.314	0.521	0.343	0.185	0.636
48	-0.08	0.208	0.701	0.923	0.614	1.387
49	-0.463	0.192	0.061	0.629	0.432	0.918
50	-0.251	0.224	0.263	0.778	0.502	1.207
51	-0.195	0.276	0.476	0.822	0.479	1.411
52	-0.575	0.255	0.674	0.563	0.341	0.929
53	-0.053	0.216	0.652	0.948	0.621	1.449
54	0.598	0.194	.000*	1.819	1.244	2.661
55	0.223	0.206	0.28	1.249	0.834	1.871
56	-0.054	0.25	0.827	0.947	0.581	1.544
57	0.124	0.211	0.558	1.132	0.749	1.71
58	-0.258	0.248	0.299	0.773	0.475	1.257
59	0.265	0.206	0.198	1.304	0.87	1.953
60	– .2 51	0.196	0.201	0.778	0.53	1.143

Table 3 shows the items in relation to school ownership (private and public), identified by logistic regression method using. Out of sixty items in NECO Economics questions DIF was present in nine items. These items are item 3, 8, 10, 22, 29, 31, and 54.

DISCUSSION OF FINDINGS

The findings show that items in relation to gender (Male and female), identified by logistic regression have DIF. Out of sixty items in NECO Economics questions DIF was present in eleven items. These items are item 8, 12, 14, 16, 22, 43, 44, 46, 49, 50, and item 58. From the findings, it is also observed that these items that showed DIF are due to the structure of the questions and stem, thus these could be the characteristics that affected the test takers response to getting the item correctly. Nworgu, (2011), revealed that current research evidence has implicated test used in national and regional examination as functioning differently with respect to different subgroups. This means that students' scores in such examinations are determined largely by the group to which an examinee belongs and not by ability. Adedoyin (2010) in his study on investigating gender biased items in public examinations; he found that out of 16 test items that fitted the 3PL item response theory statistical analysis, 5 items were gender biased.

From the findings, it is observed that these items that showed DIF are due to the structure of the questions and stem, thus these could be the characteristics that affected the test takers response to getting the item correctly. The findings of this study agrees with the work of Pedrajita, (2009) when he used Logistic regression to detect test items bias in Chemistry Achievement", the result from the study revealed that there is school type bias in the Chemistry Achievement test that was administered to the testees out of 22 items that were biased 11 items favoured public schools while eleven also favored private schools. It means that the performance of the examinees on the items do not only depend on their ability in the subject

(Economics), but also on the ownership of their schools. This finding is in congruence with Chukwudi (2019) who affirmed that 38 out of the 60 items of the 2017 BECE Mathematics multiple choice test were biased with respect to school ownership. 31 items were biased against public schools, while 7 items were biased against private schools.

CONCLUSION AND RECOMMENDATIONS

Based on the forgoing findings the following conclusions were made. There were presence of gender, school ownership and school location bias in NECO June/July 2021 Economics questions. Based on the findings study recommended that:

Test experts and developer should explore the use of differential item functioning method, particularly the use of logistic regression to detect both uniform and no uniform biased items.

Measurement practitioners should make use of logistic regression for developing a valid, reliable gender fair test school type fair test with biased items revised or replaced.

Test developers and examination bodies should take into account multiple background risk variables of gender school type and school location simultaneously when collating items for administration.

Evaluators and educational practitioners who are engaged in the development of assessment tools should use logistic regression for bias correction when dealing with both uniform and no uniform biased items.

REFERENCES

1. Abonyi, O. S. (2011). *Instrumentation in behavioural research*: A practical approach. : TIMEX Publishing Company.
2. Adebule, O. S. (2019). Reliability and levels of difficulty of objective test of ten senior secondary schools in five local government areas of Akure, Ondo State. *Educational Research and Review, 4(11), 585-587.*
3. Adebule, S. O. (2015). A comparative analysis of difficult and discriminating indices of objective test items in a Mathematics achievement test. An unpublished M.Ed Thesis, Ondo State University, Ado Ekiti
4. Adebule, S. O. (2014) Relationship between difficulty and discriminating indices of Multiple Choice and True False test items in a Mathematics Achievement test, J. Res. Dev. 3(7): 26-30
5. Adedoyin, I. E., Nenty, H. J.& b Chilisa A.R (2018). Fundamental of measurement and evaluation in education. Calabar. University Press UNESCO (2017). http://www.unesco.org/
6. Adedoyin, O. O., & Makobi, T. (2013). Using IRT psychometric analysis in examining the quality of junior certificate Mathematics multiple choice examination test items. International *Journal of Asian Social Sciences, 3(4), 992-1011.*
7. Adebukola (2019). Development and validation of an introductory technology achievement test. Unpublished M.Ed. thesis, University of Nigeria, Nsukka.
8. Adedlaziz, N. (2015). A Gender-Related Differential Item Functioning. *International Journal of Educational and Psychological Association 5, 101-116*
9. Agommuoh P.C (2016) Content Validity of May/June West African Senior School Certificate Examination (WASSCE) *Journal of Education and practice*
10. Agu, N. N., Onyekuba, C., &Anyichie, A. C. (2021). Measuring teachers' competencies in constructing classroom-based tests in Nigerian secondary schools: Need for a test construction skill inventory.
11. Aiken, L. R. Jnr. (2017): Intelligence variables and mathematics achievement. *Journal of Research in Educational Psychology. 9 (20), 200 – 215.*

12. Ajai, J. T. &Amuchie, C. I. (2015). *Educational research methods and statistics*. Academica House Publishers Nig. Ltd Jos, Plateau State.

13. Ajidgba, U. A (2019). Public Examination Bodies for Secondary education in Nigeria: WAEC and NECO;

14. Adolphus O.M (2016) Assessment of Gender Disparity in Achievement Test Item Format among Students' of Economics in Senior Secondary Schools. *Asian Journal of Educational research*

15. Akem, J. A. (2016). *Evaluation techniques in schools and colleges.* A handbook for teachers. Markudi: Selfers pub.

16. Alade, O. M. & Omoruyi, I. V. (2014). Table of specification and its relevance in educational development assessment, *European Journal of Educational and Development Psychology. 2(1), 1-17.*

17. Akanwa U.N., Agommuoh P.C. & Ihechu K.J (2016) Differential Item Functioning Method as an Items Bias Indicator For Big Data Assessment In The 21ST Century. *Journal of the Nigerian Academy of Education Vol. 16, No.2*

18. Ali, A. (2012) Educational measurement and evaluation. Awka Meks Unique Publishers.

19. Allen, M. J. & Yen, W. M (2012). Introduction to measurement theory. Long Groove, II Wareland Press.

20. Akanwa U.N., Agommuoh P.C. &Ihechu K.J (2019) Differential Item Functioning Method as an Item Bia Indicator for Big Data Assessment in the 21ST Century. *International Research Journals.*

21. Amuche, C. & Fan, A .F. (2013) Assessed Item Bias using Differential Item Functioning Technique in NECO BIOLOGY Conducted Examination in Taraba State Nigeria. *American International Journal of Research in Humanities, Arts and Social Sciences*

22. Amuche, C. I. & Igba, G. J. (2016). Influences of content validity of Teacher-made-Tests on Physics students' academic achievement in Taraba state. *Journal of Science and Technology, Mathematics and Entrepreneurial Education.*

23. Anastasi, A., & Susana, U. (2014). *Psychological testing*. 7th PHI Learning Private Ltd.

24. Anastasi, A. & Urbina, S. (2006). *Psychological testing.* 8th Edition. Upper Saddle River, NJ: Prentice-Hall.

25. Bloom, B. S. (1956). *Taxonomy of educational objectives:* Cognitive Domain. New York: David Mckay Co., Inc.

26. Borisade O. J. (2019) Investigating 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, *International Journal of Quantitative and Qualitative Research Methods*

27. Cassel, R. (2013). Confluence is a primary measure of test validity and it includes the creditability of test taker. *College Student Journal, 37:348-353.*

28. Camilli, A. (2016) *An Introduction to Psychological Assessment and* Psychometrics http://dx.doi.org/10.4135/9781446221556

29. Camilli G., Shepard L. A. (1994), *Methods for identifying biased test items,* SAGE Publications.

30. Clauser B., Mazor K., Hambleton R. K. (1993), "The Effects of Purification of Matching Criterion on the Identification of DIF Using the Mantel-Haenszel Procedure", *Applied Measurement in Education 6(4), pp. 269-279.*

31. Cohen, R. J. (2019). Psychological testing and assessment: An introduction to tests and (7th ed.). New York: McGraw-Hill.

32. Cohen, L., Manion, L. and Morrison, K. (2013) Research Methods in Education (6th ed.) London: Routledge.

33. UNESCO (2020), Multistage Sampling retrieved from *http://uis.unesco.org/englossary-term/multi-stage-sampling on 28/04/2020*

34. Wilmut, J., & Yakasai, I. M. (2016). A brief review of the assessment of student achievement in Kaduna, Kano and Kwara states of Nigeria.

35. Yang, F. M., & Kao, S. T. (2014). Item response theory for measurement validity. Shanghai Archives of Psychiatry, 26(3), 171-177. doi: 10.3969/j.issn.1002-0829.2014.03.010

36. Yu, C. H. (2013). A simple guide to the item response theory (IRT) and Rasch modeling. Retrieved from: http://www.creative-wisdom.com

37. Walker, C. (2015). What's the DIF? Why DIF analyses are an important part of the instrument in development and validation. *Journal of Educational Assessment, 29, 364-376*

38. Walker, C. M., Beretvas, S. N., & Ackerman, T. A. (2001). An examination of conditioning variable used in computer adaptive testing for DIF. *Applied Measurement in Education,*

39. L., Hambleton R.K., Robin, F. (2003), "Detection of Differential Item Functioning in Large-Scale State Assessments: *A Study evaluating a Two-Stage Approach", Educational and Psychological Measurement 63(1), pp. 51-64*