RedFame

International Journal of Contemporary Education
Vol. 5, No. 1; April 2022
ISSN 2575-3177 E-ISSN 2575-3185
Published by Redfame Publishing
URL: http://ijce.redfame.com

Examining the Nature of Item Bias on Students' Performance in National Examinations Council (NECO) Mathematics Senior School Certificate Dichotomously Scored Items in Nigeria

A. Alaba Adediwura[1], Asowo A. Patricia[1]

[1]Department of Educational Foundations and Counselling, Obafemi Awolowo University, Ile-Ife, Nigeria

Correspondence: Asowo A. Patricia, Department of Educational Foundations and Counselling, Obafemi Awolowo University, Ile-Ife, Nigeria.

Received: August 10, 2020 Accepted: September 13, 2021 Online Published: November 7, 2021

doi:10.11114/ijce.v5i1.5402 URL: https://doi.org/10.11114/ijce.v5i1.5402

Abstract

This study examined the nature of item bias on students' performance in 2017 National Examinations Council (NECO) mathematics senior school certificate dichotomously scored items in Nigeria. The study adopted an ex-post-facto research design. A sample of 256,039 candidates was randomly selected from the population of 1,034,629 students who took the test. Instrument for data collection was 'Student Results' (SR). Data collected were analysed using the R language environment and an independent t-test. Results showed that the 2017 NECO Mathematics test was essentially unidimensional (-0.28 (<.20), ASSI = -0.31 (< 0.25) and RATIO = -0.31 (< 0.36). Results also showed that the nature of bias statistically encountered was a mean difference in scores bias, indicating that 86% (52 items), 79.1% (34 items), and 96% (56 items) were biased against male students, urban and public-school students, respectively. It was concluded that item bias is a notable factor that affected the validity of the NECO 2017 Mathematics test and conclusions drawn from the scores in Nigeria. Hence, it was recommended that before tests are administered for public use, examination bodies should make a careful review of tests through dimensionality assessment at the developmental stage to eliminate any perspectives that could cause test inequity among examinees.

Keywords: schooling, national examinations, Item bias, Mental true test score approach

1. Introduction

Schooling as a tool is a channel through which formal education is achieved which was planned by society to help individuals reach self-actualisation. This is envisaged to attain development through investment in human capital formation to bridge the gap between different classes of people as a result of colour, country, technology, religion, school type, social values and beliefs, ethnicity and sex differences in the society. The aim of schooling in Nigeria, as stated in the National Policy on Education (FGN, 2013), is to prepare the individual child for a better existence in society. This is done by developing a national curriculum programme that would satisfy the basic needs required for higher education. It is at the end of their non-stop training years that public certificates' examinations that are a standard prerequisite for admission into tertiary institutions are taken.

National Examinations

In Nigeria, national examinations are done to select and place students in various educational institutions for the period of their training. Among these national examinations come the ones that are implemented during the transition from secondary education to tertiary education, such as NECO, WAEC and NABTEB; leaving candidates with the choice of either writing the 3 exams or certainly one of them. However, since the test is used for all involved groups from one state to the other every year, it is frequently simple to assume that the obtained results are comparable between groups. Besides, it can be remarked on whether the structures and contents of the items in the tests provide neither advantage nor disadvantage to any of the subgroups taking the exam. If the items in the tests provide an advantage for a group because of various features such as sex, school location, school type, race, social values and beliefs, ethnicity, or technology. It might be that the exam is biased in favour of that group which would negatively affect the validity of the selections made. So, a critical look at the perception of individuals on such standard examination in Nigeria might indicate a serious nature of item bias.

Item Bias

Bandele and Aborisade (2018) explain that item bias affects the vital psychometric properties of measurements results in terms of validity and reliability. It occurs when there is a vast distinction in the performance between male and female students, private and public schools, schools in urban and rural areas, religious views, race, social values and beliefs or from state to state. Aborisade (2016) further explained that when students that have comparable ability levels are exposed to the same course content, they should have an 'equal probability of success' irrespective of the subgroup of the population to which they belong. The scholar maintained that if the examination items contain any source of difficulty that is not relevant to the construct being measured, these extraneous sources affect examinees performance. To be able to bridge the gap between them, examination agencies are expected to conduct tests that would reflect the true behaviour of individual examinees. However, if the validity and reliability of constructed items could not be ascertained before administration. Perhaps, some examination bodies do not include item bias detection in their item analysis.

The Mental True Test Score Approach

The Mental True Test Score approach (MTTSA) has been a modern-day technique that depicts a thorough picture of item functioning among examinees. It expects how passable an examinee would perform on a test. This framework proposes two distinctive units which, when jointly considered, are responsible for observing examinee response patterns in a standardised test. The first unit is the set of unobserved values of individual abilities, which each test taker possesses, denoted by θ. This ability allows measurement experts to make a concomitant comparison of performance between examinees that have taken the test. The second unit is the set of responses essential at the item level. Such properties may reflect how difficult or extreme the items are. It would tell how well the items discriminate individuals along the scale, whether the probability of correct responses monotonically relates to individual ability and many more. However, with the MTST approach, the analysis of complex interaction patterns between subgroups, individual factors and item characteristics may detect biased items that are present in a multiple-choice test (Ahmad, Mokshein, and Husin, 2018).

In Nigeria, there have been a lot of studies on test/item bias that were both internally and internationally recognised. These studies have shown biased items across various dimensions of bias using more of the traditional differential item functioning approach and less of the modern mental true score approach to measuring irrelevant construct in a test. If examination agencies do not ensure those test items are bias-free using efficient effective methods, decisions made from the test items or test scores may overestimate or underestimate students' performance. If this happens, it would significantly affect the interpretation of conclusions made from scores which may deprive candidates who desire to study a hi-tech oriented course in tertiary institutions to lose admission or place them on courses they do not request (Orluwene & Asiegbu, 2009; Nworgu, 2010; Adedoyin, 2010; Madu, 2011; Ogbebor and Onuka, (2013); Uremu & Adams, 2013; Agbir, 2014, and Bandele & Aborisade, 2018). It might also turn away the confidence parents and society have in the effectiveness of tests. Therefore, analysis of the probability of correct response for examinees and nature of bias encountered should be done to ascertain the reliability of the 2017 NECO SSCE mathematics multiple-choice test and possible conclusions drawn from the test scores in Nigeria.

Purpose of the study

The purpose of the study was to find out the nature of bias encountered on students' performance. The study, also, ascertains the dimensionality of the test and identified items that are biased based on the probability of correct responses of examinees from comparable ability levels using the MMTST approach.

Research Questions

i. How many dimensions have NECO 2017 Mathematics test among senior school students in Nigeria?
ii. What is the nature of bias statistically encountered in NECO SSCE 2017 dichotomously scored Mathematics items in Nigeria based on?
    a. Sex?
    b. School location?
    c. School type?

2. Method

The study employed an Ex-post facto research design. The population consisted of all 1,034,629 students who sat for NECO SSCE 2017 Mathematics multiple-choice examination in Nigeria. A sample of 256,039 students representing all year three senior secondary (SSSIII) students that registered and sat for the examination in twelve states was selected using a non-proportional stratified random sampling technique with geopolitical zones serving as the basis for stratification. Samples were cluster-based on sex, school location, and school type for data analysis. The instrument for the study was "Student Results' (SR). Data collected were analysed using an R language environment. Supplementary

item response theory (SIRT) was used to test dimensionality, Unidimensional item response theory (UMIRT) was used to estimate item parameters, factor analysis and t-test with the support of SPSS was performed to re-affirm test dimensionality and bias. Also, excel was used to estimates students' true scores.

3. Results

Research Question One: How many dimensions have NECO 2017 Mathematics test among senior school students in Nigeria? Here, the responses of examinees that sat for the 2017 National Examinations Council (NECO) mathematics test in Nigeria was subjected to Stout's test of essential unidimensionality and factor analysis.

Table 1. Dimensionality Assessment of 2017 NECO Mathematics dichotomously scored items in Nigeria under STEU

	Unweighted	Weighted
DETECT	-0.28	-0.28
ASSI	-0.31	-0.31
RATIO	-0.28	-0.28

Table 1 showed that the assumption of unidimensionality was not violated. This is because the criteria of adjudging essential dimensionality of a test stand on the following basis (0.20<DETECT<1.00, ASSI < 0.25, Ratio < 0.36) according to Jang and Roussos, (2007) and Zhang (2007). This table showed that the 2017 Mathematics multiple-choice test was optimal and essentially unidimensional (maximum DETECT value = -0.28 (<.20), ASSI = -0.31 (< 0.25) and RATIO = -0.31 (< 0.36). The DETECT showed that the substantive test structure based on its purpose is consistent with the statistical dimensional structure at that level. Also, the ASSI and Ratio values revealed that examinee response data displayed an approximately simple structure. This implies that one dominant dimension accounted for the variation observed in student's responses to the mathematics dichotomously scored items in Nigeria.

Table 2. Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	**Total**	**% of Variance**	**Cumulative %**	**Total**	**% of Variance**	**Cumulative %**
1	11.618	19.363	19.363	11.618	19.363	19.363
2	2.490	4.150	23.513	2.490	4.150	23.513
3	2.386	3.977	27.489	2.386	3.977	27.489
4	2.078	3.463	30.953	2.078	3.463	30.953
5	1.948	3.246	34.199	1.948	3.246	34.199
6	1.655	2.758	36.957	1.655	2.758	36.957
7	1.277	2.128	39.085	1.277	2.128	39.085
8	1.140	1.901	40.986	1.140	1.901	40.986
9	1.085	1.809	42.795	1.085	1.809	42.795
10	1.060	1.767	44.562	1.060	1.767	44.562
11	1.015	1.691	46.253	1.015	1.691	46.253
12	.988	1.646	47.899			
13	.965	1.608	49.508			
14	.935	1.558	51.066			
15	.890	1.483	52.549			
16	.869	1.448	53.996			
17	.833	1.389	55.385			
18	.827	1.378	56.763			
19	.795	1.325	58.088			
20	.787	1.312	59.400			
21	.773	1.288	60.688			
22	.770	1.284	61.972			
23	.766	1.277	63.249			
24	.748	1.247	64.496			
25	.737	1.228	65.724			
26	.723	1.205	66.928			
27	.715	1.192	68.121			
28	.706	1.176	69.297			
29	.698	1.163	70.460			
30	.689	1.148	71.609			
31	.677	1.128	72.737			
32	.672	1.121	73.857			
33	.667	1.111	74.968			
34	.660	1.100	76.068			
35	.647	1.078	77.146			
36	.643	1.071	78.217			
37	.631	1.052	79.270			
38	.623	1.038	80.308			
39	.612	1.020	81.327			
40	.606	1.010	82.337			
41	.604	1.006	83.343			
42	.592	.987	84.331			
43	.586	.977	85.308			
44	.574	.956	86.264			
45	.570	.951	87.215			
46	.561	.935	88.150			
47	.554	.924	89.073			
48	.551	.919	89.992			
49	.537	.896	90.888			
50	.534	.889	91.777			
51	.527	.878	92.655			
52	.519	.865	93.521			
53	.515	.859	94.380			
54	.506	.844	95.223			
55	.501	.836	96.059			
56	.493	.822	96.881			
57	.488	.814	97.695			
58	.478	.796	98.492			
59	.464	.773	99.265			
60	.441	.735	100.000			

Extraction Method: Principal Component Analysis.

Table 2 showed the extraction from the principal component analysis after interacting communalities revealed 11 components with eigenvalues greater than 1. This explained 19.363, 4.150, 3.977, 3.463, 3.246, 2.758, 2.128, 1.901,

1.809, 1.767, and 1.691% of variance accounted for by each component to the total variance in all the scores. For the scores from the 60 multiple-choice Mathematics items, regarding the eigenvalues greater than 1, the total percentage variance was 46.253. from the results of PCA, the items were unidimensional because the first factor (19.363) extracted exceeded the second factor (4.751) by a reasonable distance. In addition, a scree plot was used to confirm the dimensionality of the test.

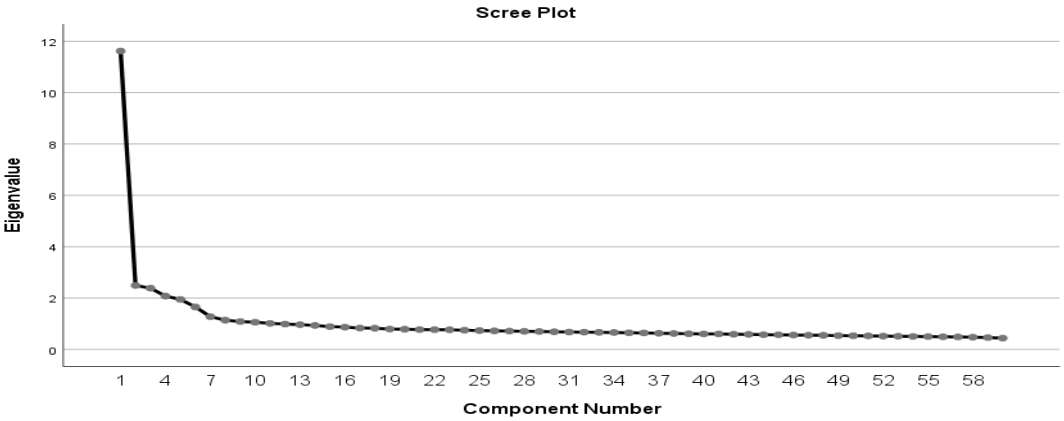

Figure 1. Scree Plot of Examinee Scores in the 2017 NECO Mathematics items

Results from the scree plot showed a visual of how the total variance associated with each factor. The steep slope revealed the largest factors associated with the loading greater than the eigenvalues of 1. Similarly, the gradual trailing off-screen showed the rest of the factors lower than eigenvalues of 1. Therefore, one distinct factor with larger eigenvalues in each case implies that the items were unidimensional.

Research Question Two: What is the nature of bias statistically encountered in 2017 NECO SSC dichotomously scored Mathematics items in Nigeria based on sex, school location and school type? Here, four levels of preliminary analyses were conducted before tentatively determining statistically the nature of bias encountered using MMTTST procedures.,

The first is DIF analysis. Here, we subject the responses of examinees from the 2017 SSC NECO Mathematics test to a Unidimensional differential item functioning (UDIF) investigation implemented in the DIF package of R language. The results revealed that 52, 43, and 58 of the items showed an incidence of DIF at 0.05 level of significance concerning sex, school location and school type. This implies that 86%, 71% and 96% of the 2017 NECO Mathematics multiple-choice test items functioned differently for candidates based on their sex, school location and school type (see appendices 1, 2 & 3). Second, examinee scores were subjected to item calibration of the R language for ability estimates and model fit. Thereafter, SPSS was used to group the ability estimates to identify examinees that have the same ability estimates. The result showed an ability estimate of 0.2 as the ability estimate with the highest frequency (13, 274) as a base for the identification of candidates that have the same ability estimate. The result implies that the examinee that falls into the same ability group had the same probability of answering the items correctly which may be an indication of administrative errors on the part of the examination agency, cheating or examination malpractice among the students (see appendix 4). Third, the probability of correct responses for all candidates with the same ability (13, 274) scores along their groups was calculated using the three-parameter logistic model that fitted the data. From the abridged table concerning sex, school location and school type, the results revealed that there were too many high scores among the candidates on many items consecutively (see appendices 4). This might imply having too many easy items on the test, the presence of cheating among students or administrative error before test administration. Finally, the probability of examinee correct responses was subjected to a comparative estimation to examine the nature of bias encountered statistically along with groups.

Table 3. Set of DIF Items that Tentatively Showed the Nature of Bias Encountered Statistically Based on Sex

Items	Sex	N	Mean	Std. Deviation	Mean Diff	t	df	Sig. (2-tailed)	Remark	Evaluation
IT1	Male	6973	.7889	.17160	-.04854	-17.758	13272	.000	Sig	Biased against male
	Female	6301	.8374	.13966						
IT3	Male	6973	.8040	.05133	-.01075	-11.401	13272	.000	Sig	Biased against male
	Female	6301	.8148	.05726						
IT4	Male	6973	.7627	.23958	-.05047	-13.115	13272	.000	Sig	Biased against male
	Female	6301	.8131	.19940						
IT5	Male	6973	.6438	.00911	-.03179	-142.923	13272	.000	Sig	Biased against male
	Female	6301	.6756	.01591						
IT7	Male	6973	.7181	.25581	-.06612	-15.927	13272	.000	Sig	Biased against male
	Female	6301	.7842	.21848						
IT8	Male	6973	.8150	.14351	-.02486	-10.317	13272	.000	Sig	Biased against male
	Female	6301	.8398	.13298						
IT9	Male	6973	.6558	.26571	-.09243	-21.123	13272	.000	Sig	Biased against male
	Female	6301	.7483	.23533						
IT10	Male	6973	.7074	.17795	-.04428	-14.762	13272	.000	Sig	Biased against male
	Female	6301	.7517	.16639						
IT11	Male	6973	.6666	.29153	-.07655	-16.152	13272	.000	Sig	Biased against male
	Female	6301	.7431	.25016						
IT12	Male	6973	.5831	.26520	-.07978	-18.441	13272	.000	Sig	Biased against male
	Female	6301	.6629	.22948						
IT13	Male	6973	.0700	.00238	-.00436	-31.594	13272	.000	Sig	Biased against male
	Female	6301	.0744	.01126						
IT15	Male	6973	.5706	.24527	-.07695	-19.068	13272	.000	Sig	Biased against male
	Female	6301	.6476	.21679						
IT16	Male	6973	.7378	.16926	-.05900	-20.299	13272	.000	Sig	Biased against male
	Female	6301	.7968	.16492						
IT17	Male	6973	.7221	.26524	-.07642	-17.734	13272	.000	Sig	Biased against male
	Female	6301	.7985	.22722						
IT19	Male	6973	.8822	.08559	-.02469	-17.283	13272	.000	Sig	Biased against male
	Female	6301	.9069	.07823						
IT20	Male	6973	.7450	.18965	-.05559	-17.712	13272	.000	Sig	Biased against male
	Female	6301	.8006	.16999						
IT21	Male	6973	.4006	.12118	-.03117	-15.778	13272	.000	Sig	Biased against male
	Female	6301	.4318	.10472						
IT22	Male	6973	.7430	.25683	-.07440	-18.458	13272	.000	Sig	Biased against male
	Female	6301	.8175	.20078						
IT23	Male	6973	.5171	.06377	-.02852	-23.760	13272	.000	Sig	Biased against male
	Female	6301	.5456	.07445						
IT25	Male	6973	.4935	.00085	-.02036	-1903.144	13272	.000	Sig	Biased against male
	Female	6301	.5138	.00007						
IT26	Male	6973	.7536	.21139	-.05550	-16.969	13272	.000	Sig	Biased against male
	Female	6301	.8091	.15853						
IT27	Male	6973	.5628	.21101	-.06929	-20.362	13272	.000	Sig	Biased against male
	Female	6301	.6321	.17739						
IT29	Male	6973	.3620	.11580	-.03581	-19.420	13272	.000	Sig	Biased against male
	Female	6301	.3978	.09421						
IT30	Male	6973	.7553	.13798	-.05050	-21.562	13272	.000	Sig	Biased against male
	Female	6301	.8058	.13106						
IT31	Male	6973	.6423	.28756	-.07972	-17.131	13272	.000	Sig	Biased against male
	Female	6301	.7220	.24393						
IT32	Male	6973	.6804	.30485	-.08710	-17.756	13272	.000	Sig	Biased against male
	Female	6301	.7675	.25481						
IT33	Male	6973	.6792	.27797	-.08287	-18.351	13272	.000	Sig	Biased against male
	Female	6301	.7621	.23813						
IT34	Male	6973	.7076	.28808	-.07834	-16.982	13272	.000	Sig	Biased against male
	Female	6301	.7859	.23776						
IT35	Male	6973	.7264	.27563	-.08066	-18.234	13272	.000	Sig	Biased against male
	Female	6301	.8071	.22882						

IT36	Male	6973	.5755	.27042	-.08760	-19.840	13272	.000	Sig	Biased against male
	Female	6301	.6631	.23451						
IT37	Male	6973	.5354	.26733	-.07733	-17.112	13272	.000	Sig	Biased against male
	Female	6301	.6127	.25164						
IT38	Male	6973	.6976	.20329	-.06318	-18.249	13272	.000	Sig	Biased against male
	Female	6301	.7608	.19457						
IT40	Male	6973	.5831	.26847	-.07183	-16.274	13272	.000	Sig	Biased against male
	Female	6301	.6549	.23681						
IT41	Male	6973	.6754	.19261	-.05174	-15.990	13272	.000	Sig	Biased against male
	Female	6301	.7271	.17880						
IT42	Male	6973	.5072	.25361	-.08275	-19.414	13272	.000	Sig	Biased against male
	Female	6301	.5900	.23558						
IT43	Male	6973	.6925	.26461	-.07589	-17.730	13272	.000	Sig	Biased against male
	Female	6301	.7684	.22418						
IT44	Male	6973	.5591	.06170	-.03920	-35.230	13272	.000	Sig	Biased against male
	Female	6301	.5983	.06649						
IT45	Male	6973	.7114	.21861	-.06417	-17.319	13272	.000	Sig	Biased against male
	Female	6301	.7755	.20695						
IT46	Male	6973	.7653	.25862	-.07702	-18.785	13272	.000	Sig	Biased against male
	Female	6301	.8423	.20784						
IT47	Male	6973	.5548	.16748	-.06063	-22.209	13272	.000	Sig	Biased against male
	Female	6301	.6155	.14467						
IT48	Male	6973	.6954	.20393	-.06106	-17.566	13272	.000	Sig	Biased against male
	Female	6301	.7565	.19550						
IT49	Male	6973	.7706	.16609	-.05866	-21.693	13272	.000	Sig	Biased against male
	Female	6301	.8292	.14305						
IT50	Male	6973	.6711	.18821	-.04565	-14.366	13272	.000	Sig	Biased against male
	Female	6301	.7167	.17670						
IT51	Male	6973	.8212	.16457	-.04041	-15.924	13272	.000	Sig	Biased against male
	Female	6301	.8616	.12217						
IT52	Male	6973	.8296	.19878	-.04918	-15.436	13272	.000	Sig	Biased against male
	Female	6301	.8788	.16452						
IT53	Male	6973	.7725	.14983	-.04283	-16.555	13272	.000	Sig	Biased against male
	Female	6301	.8153	.14775						
IT54	Male	6973	.7669	.21608	-.05275	-14.748	13272	.000	Sig	Biased against male
	Female	6301	.8196	.19373						
IT55	Male	6973	.3592	.07253	-.04667	-36.575	13272	.000	Sig	Biased against male
	Female	6301	.4059	.07439						
IT56	Male	6973	.7175	.15271	-.03217	-12.089	13272	.000	Sig	Biased against male
	Female	6301	.7496	.15352						
IT57	Male	6973	.7393	.21689	-.06091	-17.116	13272	.000	Sig	Biased against male
	Female	6301	.8002	.19042						
IT59	Male	6973	.6598	.23601	-.05954	-15.205	13272	.000	Sig	Biased against male
	Female	6301	.7193	.21277						
IT60	Male	6973	.5038	.12499	-.03756	-18.303	13272	.000	Sig	Biased against male
	Female	6301	.5414	.10986						

Significant at $p \leq 0.05$

Table 3 showed the disparity between individual examinee performances on the item with mean score differences at face value. The table revealed that (86% of items) 52 scores (1, 3, 4, 8, 9, 10, 11, 12, 13,15, 16, 17, 19, 20, 21, 22, 23, 25, 26, 27, 29, 30, 31, 32, 33, 34, 35, 36, 37, 40, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 59 and 60) were biased against male students aside eight items with logits greater than .05. Likewise, female scores showed they had the skills needed to answer difficult items with possible higher computation skills and on easier items that require basic applications. From the scores, the number of significant items by group interaction upon the assumption that there was an equivalent distribution of abilities across groups. It specified that the test is measuring irrelevant factors that are no longer related to the construct being assessed. This implies that the mean difference in score bias has contributed largely to the non-equivalency of ability distribution among males and females. This might have affected the construct and content validity of the test.

Table 4. Set of DIF Items that tentatively showed the Nature of Bias encountered statistically based on School Location

	Location	N	Mean	Std. Deviation	Mean Diff	t	df	Sig. (2-tailed)	Remark	Evaluation
IT1	Rural	4728	.8416	.11844	.04703	16.327	13272	.000	sig	Biased against Urban
	Urban	8546	.7946	.17739						
IT2	Rural	4728	.5420	.04960	.03431	37.919	13272	.000	sig	Biased against Urban
	Urban	8546	.5077	.05009						
IT4	Rural	4728	.8244	.17840	.05812	14.500	13272	.000	sig	Biased against Urban
	Urban	8546	.7662	.24157						
IT5	Rural	4728	.6465	.00311	-.01671	-56.375	13272	.000	sig	Biased against Rural
	Urban	8546	.6632	.02025						
IT6	Rural	4728	.8302	.06813	.00896	5.874	13272	.000	sig	Biased against Urban
	Urban	8546	.8213	.09181						
IT7	Rural	4728	.7906	.19437	.06302	14.560	13272	.000	sig	Biased against Urban
	Urban	8546	.7275	.26015						
IT8	Rural	4728	.8533	.11870	.04134	16.450	13272	.000	sig	Biased against Urban
	Urban	8546	.8120	.14852						
IT9	Rural	4728	.7490	.22571	.07931	17.442	13272	.000	sig	Biased against Urban
	Urban	8546	.6697	.26377						
IT10	Rural	4728	.7570	.14428	.04744	14.706	13272	.000	sig	Biased against Urban
	Urban	8546	.7095	.19413						
IT11	Rural	4728	.7378	.23349	.05373	10.800	13272	.000	sig	Biased against Urban
	Urban	8546	.6841	.29475						
IT12	Rural	4728	.6669	.22042	.07276	16.064	13272	.000	sig	Biased against Urban
	Urban	8546	.5941	.26479						
IT13	Rural	4728	.0706	.00294	-.00249	-20.377	13272	.000	sig	Biased against Rural
	Urban	8546	.0731	.00812						
IT14	Rural	4728	.6991	.23345	.07030	15.087	13272	.000	sig	Biased against Urban
	Urban	8546	.6288	.26926						
IT15	Rural	4728	.6200	.21639	.01720	4.071	13272	.000	sig	Biased against Urban
	Urban	8546	.6028	.24182						
IT16	Rural	4728	.7973	.14212	.05384	17.343	13272	.000	sig	Biased against Urban
	Urban	8546	.7435	.18545						
IT17	Rural	4728	.7987	.21067	.06288	14.008	13272	.000	sig	Biased against Urban
	Urban	8546	.7358	.26595						
IT18	Rural	4728	.4909	.09369	-.00914	-5.001	13272	.000	sig	Biased against Rural
	Urban	8546	.5000	.10462						
IT20	Rural	4728	.8046	.14861	.05270	16.001	13272	.000	sig	Biased against Urban
	Urban	8546	.7519	.19767						
IT21	Rural	4728	.4261	.09778	.01517	7.222	13272	.000	sig	Biased against Urban
	Urban	8546	.4109	.12482						
IT24	Rural	4728	.7797	.16207	.04649	13.636	13272	.000	sig	Biased against Urban
	Urban	8546	.7332	.20103						
IT25	Rural	4728	.4833	.00106	-.03154	-933.288	13272	.000	sig	Biased against Rural
	Urban	8546	.5149	.00218						
IT27	Rural	4728	.6127	.17218	.02368	6.611	13272	.000	sig	Biased against Urban
	Urban	8546	.5890	.21043						
IT28	Rural	4728	.3859	.06813	.04900	26.121	13272	.000	sig	Biased against Urban
	Urban	8546	.3369	.11861						
IT29	Rural	4728	.3637	.10255	-.03045	-15.696	13272	.000	sig	Biased against Rural
	Urban	8546	.3941	.10941						
IT30	Rural	4728	.7864	.11371	.01220	4.817	13272	.000	sig	Biased against Urban
	Urban	8546	.7742	.15230						
IT34	Rural	4728	.7776	.23588	.04844	10.075	13272	.000	sig	Biased against Urban
	Urban	8546	.7292	.28018						
IT37	Rural	4728	.5773	.25083	.00713	1.497	13272	.134	NS	IMPACT
	Urban	8546	.5702	.26880						
IT39	Rural	4728	.7834	.16479	.19270	40.934	13272	.000	sig	Biased against Urban
	Urban	8546	.5907	.29959						
IT40	Rural	4728	.6357	.23626	-.00633	-1.453	13272	.146	Sig	Biased against Rural
	Urban	8546	.6420	.24274						
IT41	Rural	4728	.7166	.17926	.18963	40.159	13272	.000	sig	Biased against Urban
	Urban	8546	.5270	.29604						
IT44	Rural	4728	.5754	.04158	-.03258	-28.092	13272	.000	sig	Biased against Rural
	Urban	8546	.6080	.07351						
IT45	Rural	4728	.7678	.18773	.04253	10.782	13272	.000	sig	Biased against Urban
	Urban	8546	.7252	.23251						

IT46	Rural	4728	.8405	.17838	.04378	10.785	13272	.000	sig	Biased against Urban
	Urban	8546	.7967	.24557						
IT48	Rural	4728	.7599	.17411	.17765	36.903	13272	.000	sig	Biased against Urban
	Urban	8546	.5823	.30462						
IT50	Rural	4728	.7289	.15687	.08887	24.564	13272	.000	sig	Biased against Urban
	Urban	8546	.6400	.21970						
IT52	Rural	4728	.8785	.12888	.02765	8.715	13272	.000	sig	Biased against Urban
	Urban	8546	.8509	.19597						
IT53	Rural	4728	.7995	.13858	.21056	44.692	13272	.000	sig	Biased against Urban
	Urban	8546	.5890	.30712						
IT54	Rural	4728	.8237	.15359	.03370	9.492	13272	.000	sig	Biased against Urban
	Urban	8546	.7900	.21577						
IT55	Rural	4728	.3760	.05352	-.05535	-41.773	13272	.000	sig	Biased against Urban
	Urban	8546	.4314	.08195						
IT56	Rural	4728	.7487	.14412	.22139	51.870	13272	.000	sig	Biased against Urban
	Urban	8546	.5273	.27321						
IT58	Rural	4728	.7998	.14630	.03409	9.878	13272	.000	sig	Biased against Urban
	Urban	8546	.7657	.21088						
IT59	Rural	4728	.7132	.18624	.01670	4.239	13272	.000	sig	Biased against Urban
	Urban	8546	.6965	.23281						
IT60	Rural	4728	.5287	.10131	-.02926	-14.343	13272	.000	sig	Biased against Rural
	Urban	8546	.5579	.11830						

Significant at $p \leq 0.05$

Table 4 illustrates the disparity between individual examinee performances on the item with mean score differences at face value among the comparison groups. The table showed that (79.1% of items) 34 scores (1, 2, 4, 6, 7, 7, 8, 10, 11, 12, 14, 15, 16, 17, 20, 24, 27, 28, 30, 34, 39, 41, 45, 46, 47, 48, 50, 52, 53, 54, 55, 56, 58 and 59) were biased against students from urban schools. Similarly, (18.6% of items) 8 scores (5, 13, 18, 25, 29, 44 and 60) were biased against candidates from rural schools. It also revealed that score 37 (2.3% of items) showed item impact, which means that the item measured the intended construct of the test. the results revealed that scores of students from urban areas showed they were at a disadvantage on difficult items that require higher calculation skills but were on an advantage at easier items that require the application of one to more simple basic mathematical operations based on the location of their schools. Likewise, students' scores from rural areas showed it advantaged them on difficult items that require higher calculation skills but maybe careless at easier items that require simple mathematical skills. It is possible from the score that candidates from urban schools lack good mastery of subject content and confident skills when the test was administered, which had put them in a disadvantaged position. The result further revealed that the relative difficulty of scores on the test based on school location has not remained constant considering the number of significant scores of group interaction. From the scores, it is unlikely that it met the assumption of the equivalent distribution of abilities across groups. The result implies that the nature of bias encountered based on school location was a mean difference in scores bias. Thus, a reflection of differential construct and content validity biases might have occurred along with the groups.

Table 4.1.2.3. Set of DIF Items that tentatively showed the Nature of Bias encountered statistically Based on School Type

	Type	**N**	**Mean**	**Std. Deviation**	**Mean Diff**	**t**	**df**	**Sig. (2-tailed)**	**Remark**	**Evaluation**
IT1	Private	6079	.8395	.12203	.04590	17.404	13272	.000	Sig	Biased against public
	Public	7195	.7936	.17233						
IT2	Private	6079	.5421	.07067	.03151	31.794	13272	.000	Sig	Biased against public
	Public	7195	.5105	.04184						
IT3	Private	6079	.8349	.06439	.03863	39.473	13272	.000	Sig	Biased against public
	Public	7195	.7963	.04817						
IT4	Private	6079	.8264	.16896	.06673	17.914	13272	.000	Sig	Biased against public
	Public	7195	.7597	.24540						
IT5	Private	6079	.7283	.05439	.10294	158.369	13272	.000	Sig	Biased against public
	Public	7195	.6254	.00828						
IT6	Private	6079	.8382	.06654	.02342	16.679	13272	.000	Sig	Biased against public
	Public	7195	.8148	.09078						
IT8	Private	6079	.8442	.13890	.03120	12.756	13272	.000	Sig	Biased against public
	Public	7195	.8130	.14165						
IT9	Private	6079	.7246	.21965	.04705	10.800	13272	.000	Sig	Biased against public
	Public	7195	.6775	.27316						
IT10	Private	6079	.7409	.15876	.02485	8.250	13272	.000	Sig	Biased against public
	Public	7195	.7160	.18399						
IT11	Private	6079	.7538	.21897	.08855	18.902	13272	.000	Sig	Biased against public

	Public	7195	.6653	.30479						
IT12	Private	6079	.6533	.20332	.05737	13.409	13272	.000	Sig	Biased against public
	Public	7195	.5959	.27632						
IT13	Private	6079	.0806	.01483	.01274	70.900	13272	.000	Sig	Biased against public
	Public	7195	.0678	.00325						
IT14	Private	6079	.6724	.23278	.03701	8.227	13272	.000	Sig	Biased against public
	Public	7195	.6354	.27790						
IT15	Private	6079	.6377	.20878	.05212	12.938	13272	.000	Sig	Biased against public
	Public	7195	.5856	.24862						
IT16	Private	6079	.7707	.15915	.01552	5.280	13272	.000	Sig	Biased against public
	Public	7195	.7551	.17643						
IT17	Private	6079	.7876	.21010	.05255	12.363	13272	.000	Sig	Biased against public
	Public	7195	.7351	.26930						
IT18	Private	6079	.5274	.08339	.04834	29.000	13272	.000	Sig	Biased against public
	Public	7195	.4791	.10496						
IT19	Private	6079	.9142	.06717	.03443	24.257	13272	.000	Sig	Biased against public
	Public	7195	.8798	.09185						
IT20	Private	6079	.7973	.14658	.04477	14.675	13272	.000	Sig	Biased against public
	Public	7195	.7525	.19599						
IT21	Private	6079	.4291	.09815	.02267	11.627	13272	.000	Sig	Biased against public
	Public	7195	.4065	.12239						
IT22	Private	6079	.8221	.18500	.07248	18.471	13272	.000	Sig	Biased against public
	Public	7195	.7496	.25435						
IT23	Private	6079	.5847	.06429	.08450	73.928	13272	.000	Sig	Biased against public
	Public	7195	.5002	.06671						
IT24	Private	6079	.7859	.17454	.06125	18.777	13272	.000	Sig	Biased against public
	Public	7195	.7247	.19737						
IT25	Private	6079	.5801	.00008	.11514	36557.583	13272	.000	Sig	Biased against public
	Public	7195	.4649	.00024						
IT26	Private	6079	.8257	.14033	.07315	23.354	13272	.000	Sig	Biased against public
	Public	7195	.7525	.20737						
IT27	Private	6079	.6477	.16685	.08293	24.778	13272	.000	Sig	Biased against public
	Public	7195	.5647	.21113						
IT28	Private	6079	.3751	.09887	.02775	16.119	13272	.000	Sig	Biased against public
	Public	7195	.3473	.09876						
IT29	Private	6079	.4175	.09370	.05848	32.111	13272	.000	Sig	Biased against public
	Public	7195	.3590	.11288						
IT30	Private	6079	.8068	.13681	.04644	19.286	13272	.000	Sig	Biased against public
	Public	7195	.7604	.13942						
IT31	Private	6079	.7190	.23214	.06713	14.479	13272	.000	Sig	Biased against public
	Public	7195	.6519	.29180						
It32	Private	6079	.7578	.24617	.06524	13.322	13272	.000	Sig	Biased against public
	Public	7195	.6926	.30752						
IT33	Private	6079	.7575	.22468	.06761	15.073	13272	.000	Sig	Biased against public
	Public	7195	.6899	.28225						
IT34	Private	6079	.7909	.21602	.07938	17.470	13272	.000	Sig	Biased against public
	Public	7195	.7116	.29339						
IT35	Private	6079	.8199	.20442	.09346	21.504	13272	.000	Sig	Biased against public
	Public	7195	.7265	.28199						
IT36	Private	6079	.6680	.23256	.08281	18.716	13272	.000	Sig	Biased against public
	Public	7195	.5852	.27078						
IT37	Private	6079	.5890	.24985	.03677	7.995	13272	.000	Sig	Biased against public
	Public	7195	.5523	.27537						
IT38	Private	6079	.7239	.21791	.00148	.411	13272	.681	sig	Item Impact
	Public	7195	.7225	.19740						
IT39	Private	6079	.7792	.19404	.04915	14.972	13272	.000	Sig	Biased against public
	Public	7195	.7301	.18354						
IT40	Private	6079	.6566	.22829	.06541	14.855	13272	.000	Sig	Biased against public
	Public	7195	.5912	.27171						
IT41	Private	6079	.6940	.18108	-.00295	-.898	13272	.369	Sig	Item Impact
	Public	7195	.6969	.19449						
IT42	Private	6079	.5768	.23659	.05419	12.524	13272	.000	Sig	Biased against public
	Public	7195	.5227	.25786						
IT43	Private	6079	.7409	.22633	.02700	6.278	13272	.000	Sig	Biased against public
	Public	7195	.7139	.26308						
IT44	Private	6079	.6149	.05819	.05790	55.340	13272	.000	Sig	Biased against public
	Public	7195	.5570	.06160						

IT45	Private	6079	.7748	.18063	.05905	16.077	13272	.000	Sig	Biased against public
	Public	7195	.7158	.23334						
IT46	Private	6079	.8452	.19735	.07136	17.783	13272	.000	Sig	Biased against public
	Public	7195	.7739	.25492						
IT47	Private	6079	.6281	.14764	.07009	25.759	13272	.000	Sig	Biased against public
	Public	7195	.5580	.16306						
IT48	Private	6079	.7439	.18626	.03844	11.003	13272	.000	Sig	Biased against public
	Public	7195	.7055	.21188						
IT49	Private	6079	.8255	.15782	.04579	16.487	13272	.000	Sig	Biased against public
	Public	7195	.7797	.16077						
IT50	Private	6079	.6810	.15695	-.00920	-2.883	13272	.004	Sig	Biased against public
	Public	7195	.6902	.20286						
IT51	Private	6079	.8734	.10765	.05308	21.994	13272	.000	Sig	Biased against public
	Public	7195	.8203	.16006						
IT53	Private	6079	.8099	.14225	.03215	12.308	13272	.000	Sig	Biased against public
	Public	7195	.7778	.15619						
IT54	Private	6079	.8135	.16746	.03989	11.463	13272	.000	Sig	Biased against public
	Public	7195	.7736	.22343						
IT55	Private	6079	.4578	.08801	.11674	89.107	13272	.000	Sig	Biased against public
	Public	7195	.3410	.06236						
IT56	Private	6079	.7362	.16779	.00903	3.265	13272	.001	Sig	Biased against public
	Public	7195	.7272	.15086						
IT57	Private	6079	.8012	.17255	.05582	16.046	13272	.000	Sig	Biased against public
	Public	7195	.7454	.22004						
IT58	Private	6079	.8074	.15745	.06526	19.632	13272	.000	Sig	Biased against public
	Public	7195	.7422	.21499						
IT59	Private	6079	.6982	.20021	.02246	5.750	13272	.000	Sig	Biased against public
	Public	7195	.6758	.24264						
IT60	Private	6079	.5670	.11195	.06801	33.717	13272	.000	Sig	Biased against public
	Public	7195	.4990	.11893						

Significant at $p \leq 0.05$

Table 5 showed the disparity between individual examinee performances on the items, with mean differences in the score at face value among the comparison groups. The table showed that 56 scores (96% of items) (1, 2, 3, 4, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 39, 40, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 58, 59 and 60) were biased against candidates from public schools; while (3% of items) score 38 and 41 showed item impact. This showed that there was a similar bundle of items assembled in the tests that made comparison groups respond equally to these items because item impact reflected the intended purpose of the test. It was apparent that it disadvantaged examinees from public schools on difficult items that required higher calculation skills, perhaps; they were also careless on easier items that require the application of one to more simple basic mathematical operations. Scores of examinees from private schools also indicated that they could answer both difficult and easier items that require higher calculation skills and on item estimates of simple mathematical skills. In comparing scores along with the groups, we might conclude that students from private schools have mastered the subject content well and have confident skills when the test was administered. The results imply that school type is a factor that is irrelevant to the purpose of the test that brought mean differences in score across groups, thereby reflecting differential construct validity and content validity biases of items.

4. Discussion

The study under the MTTST method examined the nature of item bias on students' performance in the 2017 NECO SSC dichotomously scored Mathematics test in Nigeria. Findings from the study showed that the 2017 NECO SSC dichotomously scored Mathematics test was essentially unidimensional. The results of factor analysis further revealed that the 2017 NECO SSCE Mathematics items are evidence of unidimensionality as the first eigenvalue factor exceeded that of the second factor with a reasonable distance. This means that the test was accurate in interpreting the behaviour of examinees on the test. Therefore, the assumption holds to a good extent that only one dominant dimension accounted for the variations observed in student's responses to the 2017 NECO Mathematics multiple-choice test items in Nigeria.

These findings support the submission of Ojerinde and Ifewulu (2012) who had earlier investigated test dimensionality using the 2010 UTME Mathematics examination with the outcome that specified that the test fulfilled the assumption of unidimensionality using the MTTST approach. These findings also corroborated the results of the study of Orluwene and Asiegbu (2009) who earlier investigated test dimensionality of Rivers State JSSCE Business Studies using the Item Response Theory approach. Its outcomes indicated that the test items met the assumptions of unidimensionality and local independence.

Preceding the nature of bias statistically encountered, some preliminary analyses carried out, established a relationship between the test and what was aimed to wholly measure. First, there were indications that the items that reflected DIF in the test represent very hard to very easy items in the subject content of the 2017 NECO Mathematics multiple-choice test blueprint. It implies that not only very difficult items are susceptible to DIF but with easier items as well, despite careful development of items by measurement experts. The study of Madu (2011) provided evidence on gender differences in mathematics multiple-choice test items as it varies according to the content area even when the substance is nearly linked to the course of study. The findings of this study are consistent with the study by Ogbebor and Onuka (2013), Abedalaziz (2011) and Nworgu (2010) who reported the incidence of gender, location and school type DIF in mathematics, Economics and Biology respectively. It means, there were other irrelevant factors such as model misspecification, school location, age, sex, culture, peer group, ethnicity, social values and beliefs that have affected the magnitude to which DIF occurs in these examinations. Findings revealed 0.2 was established as the highest ability estimate of candidates with the same probability of answering the items correctly. It means that successful students with the chance of a maximum score on the test proceeded for further studies, while the chances of unsuccessful examinees continuing their studies are restricted. On another thought, the possibility of examinees to have the same ability estimates on the test by answering the items correctly may also suggest incidences of administrative errors on the part of the examination agency, guessing, easiness of items, test wiseness, ethnicity, test language, item format, cheating, or examination malpractices among the students. If examination agencies do not ensure that test items are free from such factors, decisions made from the scores might overestimate or underestimate students' performance in the process. This implies that any misinterpretation of the examinee latent ability on the scale would negatively affect the validity of the decisions made which may lead to a biased result or incorrect conclusions about individuals from the groups. The findings from the estimation of the probability of correct response of examinees having the same ability estimate based on sex, school location and school type showed the total true score of each candidate on the test. It revealed that male students, students from urban and public schools were disadvantaged on items that required higher calculation skills and perhaps, on easier items that require the application of one to more simple basic mathematical operations. These findings are in disparity with Allahnana *et al.* (2018) study on gender interest. The researcher observed male students outshined in mathematics achievement tests than their female counterparts. Likewise, the scholar believed that less complex rural lifestyles are more than what is experienced at city centres where cultural diversity affects students' educational achievement. The results of the study of Agbir (2004) were consistent with these findings who discovered that rural area students performed better on realistic abilities in chemistry than their city centres counterparts did. It was clear in this study that candidates from rural areas could perform better than candidates from urban areas, considering the provision of social amenities and opportunities. However, for this construct does not have reliably measure its intended purpose effectively, the factors that brought differences in the performances of students might have affected the reliability of the examination scores. As noted by Geisinger *et al* (2013), these factors could be indications of time between testing administrations, the similarity of content, changes in subjects over time that is introduced by physical complaints, emotional problems, location of the school, type of school, sex, fatigue, starving, while the test-based agents are poor test instructions, administrative errors, subjective scoring, test wiseness, guessing, cheating, model misspecification, and expectations of subjects regarding different elements of the examination.

5. Conclusion and Recommendations

The findings of the study showed that the 2017 NECO Mathematics multiple-choice examination designed for the certification of candidates for higher education exhibits a mean difference in score bias. Based on the findings of this study, the following recommendations were made:

(1) Examination agencies and test developers should be encouraged to consistently have test materials reviewed by experts trained in identifying culturally and linguistically diverse subgroups.

(2) Test experts should incorporate a base for identifying candidates that have the same ability estimate with the highest frequency for bias analysis. This would provide information on the examinee that falls into the same ability group.

(3) Examination agencies and test experts should also subject scores of candidates with the same ability scores by estimating their probability of correct response along with their groups with the model that fitted the data. This would provide information on candidates actual true scores and know whether the items are too easy or too difficult for the student.

(4) Curriculum developers may consider the inclusion of more topics in curriculum design that could measure achievement connected with knowledge and skills students need outside the school.

References

Abedalaziz, N. (2011). Detecting dif using item characteristic curve approaches. *International Journal of Educational and Psychological Assessment*, *8*(2), 1-15.

Aborisade, O. J. (2016). Comparative analysis of item bias of the mathematics examination constructed by WAEC and NECO in Nigeria. An unpublished PhD Thesis of the Ekiti State University, Ado- Ekiti, Nigeria.

Adedoyin, O. O. (2010). Using IRT approach to detect gender biased items in public examination: A case study from the Botswana junior certificate examination in mathematics. *Educational Research and Reviews*, *5*(7), 385-399.

Agbir, E. (2004). *The meaning of things.* New York: Cambridge University Press.

Ahmad, H., Mokshein, S. E., & Husin, M. R. (2018). Detecting item bias in an anatomy & physiology test for nursing students using item response theory. *International Journal of Academic Research in Progressive Education and Development*, *7*(1), 97-109. https://doi.org/10.6007/IJARPED/v7-i1/3904

Allahnana, K. M., Akande, M. T., Vintseh, I. M., Usman, Alaku, E. A., & Monica, E. A. (2018). Assessment of gender and interest in mathematics achievement in Keffi local government area of Nasarawa State, Nigeria. *International Journal of Operational Research in Management, social sciences and Education. IJORMSSE. 5*(9), 10-13.

Bandele S. O., & Aborisade, O. J. (2018). Examining the item bias of mathematics examinations constructed by WAEC and NECO in Nigeria. *International Journal of Quantitative and Qualitative Research Methods, 6*(2), 1-7.

Federal Republic of Nigeria (2013). National Policy on Education, 2014 Edition, Lagos, NERDC.

Geisinger, K. F., Bracken, B. A., Carlson, J, F., Hansen, J. C., Kuncel, N. R., Reise, S. P., & Rodriguez, M. C. (2013). APA handbook of testing and assessment in psychology. Volume 3: Testing and assessment in school psychology and education. Washington, DC: APA. https://doi.org/10.1037/14049-000

Jang, E. E., & Roussos, L. A. (2007). An investigation into the dimensionality of TOEFL using conditional covariance-based nonparametric approach. *Journal of Educational Measurement, 44*(1), 1-21. https://doi.org/10.1111/j.1745-3984.2007.00024.x

Madu, B. C. (2011). Using transformed item difficulty procedure to assess ggender-relateddifferential item functioning of multiple-choice mathematics items administered in Nigeria Research on Humanities and Social Sciences. ISSN 2224-5766 (Paper) ISSN 2225-0484(Online) *2*(6), 41-55.

Nwogu, M. (2010). Achievement in mathematics and science: Do mothers' beliefs matter 12 years later? *Journal of Educational Psychology, 96*(1), 97-109. https://doi.org/10.1037/0022-0663.96.1.97

Ogbebor, U., & Onuka, A. (2013). Differential item functioning of economics question papers of national examinations council in Delta State, Nigeria, *Nigerian Journal of Educational Research and Education*, *12*(1), 45-60.

Ojerinde, D., & Ifewulu, B. C. (2012). *Item Unidimensionality Using 2010 Unified Tertiary Matriculation Examination Mathematics Pre-Test*. a paper presented at the 2012 International Conference of IAEA. Kazakstan.

Orluwene, G., & Asiegbu, C. N. (2009). Detecting item bias in Rivers State JSSCE business studies using Item Response Theory (IRT) approach. *European Journal of Educational and Development Psychology, 4*(1), 42-52.

Uremu, O., & Adams, O. (2013). Differential items functioning method as an item bias indication. *Journal of Educational Research*. 4(4), 367-373.

Zhang, J. (2007). Conditional covariance theory and detect for polytomous items. *Psychometrika, 72*, 69-91. https://doi.org/10.1007/s11336-004-1257-7

Copyrights