Original Article

Gender-Based Differential Item Functioning Analysis as an Innovative Pedagogical Strategy: An X-Ray of National Business and Technical Certificate Examinations

Mark E. Ugwoke[1] and Leonard C. Anigbo[2]

[1]Department of Science and Computer Education, Godfrey Okoye University, Enugu, Nigeria

[2]Department of Mathematics and Computer Education, Enugu State University of Science and Technology, Enugu,

Nigeria

Abstract

Schools across the globe have reopened after Coronavirus disease (COVID-19) pandemic with innovative pedagogical approaches including innovations in teacher-made assessments to help manage the effect of the pandemic. There are growing concerns among educators that teacher-made assessment (tests) may contain differential item functioning (DIF). The paper assessed gender-based DIF in the 2020 May/June National Business Certificate (NBC)/National Technical Certificate (NTC) Physics multiple choice test of the National Business and Technical Examinations Board (NABTEB) for possible adaptation in classroom assessment. Ex-post facto research design was used for the study. A total of 23,769 candidates that sat for the physics certificate examinations in Nigeria in 2020 were selected by census. 2020 May/June NBC/NTC Physics multiple choice test was used for data collection. The instrument belongs to a - public examination body, NABTEB and cannot be modified by external validation process. The reliability was established by NABTEB too. The students' responses to items of the instrument were obtained from NABTEB in a-person by item matrix format. The Binary Logistic Regression of SPSS 21 version was used to analyse gender-based DIF for each item. The results of the data analyses provided answers to research questions. The hypothesis was tested at 0.05 level of significance using Wald statistic. It was found that some items of the instrument contained gender-based DIF and recommended, among others, that teachers should consider DIF analysis during test development processes to enhance test fairness and quality.

Keywords: Differential item functioning**,** innovative pedagogies, national business certificate, national technical certificate, teacher-made formative and summative assessments.

Introduction

The Coronavirus disease (COVID-19) disrupted educational programmes leading to closure of schools and institutions across the globe, thus depriving millions of students and pupils of their schooling. However, schools have reopened with nations developing innovative responses including pedagogical approaches to sustain gains already made towards the goals of the 2030 Education Agenda. The innovative pedagogical approaches are essential to help manage the after effect of the pandemic. Mynbayeva et al. (2017) defined innovative pedagogies as new and creative teaching strategies adopted by teachers to facilitate effective interaction between the teachers and learners for the purpose of enhancing learners' mastery of the subject being taught. Innovative pedagogies can also be viewed from the angle of learning results acquired by the learners that are required for their effective contribution to a better future (CIVIS, 2020).

Innovative pedagogies in the classroom take different approaches including innovations in teacher-made formative and summative assessments. Formative assessment is a type of assessment teachers

carry out during or after teaching to enable them ascertain learners' progress. Formative assessment also enables teachers identify any challenges that students are facing as they learn. The most commonly used formative assessment instrument is teacher-made tests (Anigbo, 2014). The teacher-made tests are used by the teachers to ascertain the level of knowledge and skills the students have acquired in a particular subject (Adoet al., 2020; Naar, 2021). The teacher-made tests provide feedback on students' performance and are used to compare students' or groups of students' performance. According to Ugwoke (2021), many developed countries such as Australia, Canada, Denmark, Finland, Ireland, Italy, New Zealand, Switzerland, Turkey and United Kingdom have educational policies that promote the use of teacher made test to modify pedagogical approaches as learning progresses and to improve their levels of academic achievement.

One of the core issues in using test results to compare individuals and groups performance is test item bias (differential item functioning). Item bias may lead to inappropriate interpretations of test results. The Government of South Africa in 1998 enacted Employment Equity Act 55 which prohibited the use of psychological tests for recruitment of citizens into the country's service unless there are clear evidence that such tests are not biased or do not discriminate against any group of prospective employees (Beer, 2004 as cited in Ugwoke, 2021). When tests are labeled "biased" or contain DIF, it means that there fundamental issues with the items of the tests. The DIF is defined differently by different authors. A test item contains "DIF" when item takers who have equal ability to respond correctly to the item failed to answer the item correctly because they belong to different groups (e.g. gender) (Walker, 2011). According to Ojerinde et al. (2012), DIF is defined as the probability of answering an item differently by item takers of equal latent ability due to differences unrelated to the item's construct. Perhaps, Karami and Nodoushan (2011) were more assertive in their remark that an item contains DIF when item takers with the same latent ability but belong to different groups have an unequal probability of responding correctly to the item. Conversely, an item does not contain DIF when item takers with the same latent ability have equal probability of getting an item correct, regardless of their group's membership.

Analysis of DIF is important because we always interested in comparing the performance of groups of students. To conduct DIF analysis, Item Response Theory (IRT) provides a veritable framework for the analysis. The IRT relates the probability of choosing each of the response categories on a latent trait, Θ to the characteristics of items (item parameters) and characteristics of individuals (latent traits, Θ). According to Ugwoke (2021), the relationship between item parameters and latent traits and the probability of choosing an item response category on a latent trait, Θ is a nonlinear regression defined by the item response function (IRF). Ojerinde, et al. (2012) remarked that several families of item response functions are used to improve the quality of individual items during test development process. The three-parameter logistic model of the IRT consists of difficulty parameter, discrimination parameter and guessing factor. Ugwoke (2021) citing Baker (2001) opined that during testing using multiple choice tests, there is the probability of test takers getting some items correct by guessing. Thus, the probability of correct response to an item includes a small component that is due to guessing which leads to the equation for the three – parameter logistic IRT model as:

$$P(\theta) = c + (1 - c)\frac{1}{1 + e^{-a(\theta - b)}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots 1$$

where b is the difficulty parameter;

a is the discrimination parameter (power);

c is the guessing parameter; and

θ is the ability level of a particular examinee.

The c parameter is the probability of getting the item correct by guessing alone (Ojerinde et al, 2012; Ugwoke, 2021). The 3PL model was considered appropriate for this study because test takers with low trait levels may respond correctly to difficult items in the Physics multiple choice tests. Multiple choice

test is a very common objective test used in classrooms by teachers. Such test items require the students to select one answer from the list of alternatives in which only one alternative is correct, whereas other alternatives are wrong. The correct alternative is called the key while the remaining wrong alternatives are called distractors (Anigbo, 2014).

Several statistical procedures can be used to determine DIF. Sa'ad, Ali and Abdullahi (2020) highlighted some of the statistical procedures to determine DIF including Monte Carlo Simulation Approach, Haenszel (MH) technique, Simultaneous Bias Test (SIBTEST) and Swaminathan and Rogers Approach. For this study, logistic regression was employed. According to Ugwoke (2021), logistic regression procedure is widely used to detect DIF because the approach is considered simple, robust and reasonably efficient. The students' gender is an important variable in a typical classroom in Nigerian technical colleges.

The gender as used in this study is male and female students. Gender-based differences in achievements in sciences have been widely studied and reported in literatures. The factors responsible for the differences are a source of concern to researchers. National studies on tests designed for classification, admission, recruitment, or placement purposes revealed that some of the test items contained DIF (Abiam & Odok (cited in Osadebe & Agbure, 2018)). Osamede et al. (2016) investigated gender-based DIF in NABTEB's 2015 NBC/NTC Mathematics multiple choice test items and found that six (6) items favoured the male students while 11 items were in favour of the female students. Sapmaz (2019) investigated gender-based DIF in the MSPC-2018 Higher Education Institutions Examination and found that some females performed better than the males in operation skills while the males performed better on problem-solving and analytical thinking arithmetic skills than the females.

In a similar vein, Yildirim (2015) investigated gender-based DIF using 20 subset items of mathematics in the 2012 eighth-grade examination in Ankara, Turkey and found that item one of the items which was related to household items favoured girls while the remaining 19 items which were related to football games favoured the boys. Odett cited in Sapmaz (2019) studied gender-related DIF on seventh-grade mathematics items and found that males did better on fractions, percentages and measurement sub-topics while the females did better in logical and statistical problems. Abedalaziz (2010) conducted a gender-related DIF study using Logistic Regression model and Mantel Hanzael (M-H)) in mathematical ability scale test items and found that males did better than the females in spatial and deductive abilities while the females performed better than the males in numerical ability.

The fact that some items in a test might be disadvantaged to one subgroup or another has become a matter of great concern to public examination bodies and test users. Public examination bodies which administer test items across students' gender anticipate that the tests are valid, reliable and fair to all (Nisbet & Shaw, 2019). The test used by the examination boards should measure in such a way that it gives every test taker or groups of test takers equal opportunity to respond the test. This is called measurement invariance which according to Şekercioğlu (2018), is concerned with the structures of a test giving equal opportunity to all test takers from different groups to respond appropriately to the test especially when the performance of the groups are being compared using the test outcomes. In other words, there is the need to establish measurement invariance among groups while carrying out DIF analysis to ascertain real source of their differences in performance. Variances in measurement lead to wrong interpretation of test results. The implication is that performance on a test should depend solely on the students' mastery of the construct measured by the test.

Incidentally, a typical physics classroom in Nigerian technical colleges consists of students from different demographic backgrounds. The NBC/NTC Physics multiple choice test was developed and

published by a public examination board in Nigeria - NABTEB. The May/June NBC/NTC examinations are conducted across the country every year through uniform administration of the tests to candidates whose demographic data such as gender are provided. Thorndike and Thorndike-Christ (2010) remarked that such tests items developed and published by State Examinations Development Centre like NABTEB, JAMB, NECO and WAEC should not be biased or detected as having DIF.

Globally, education is recognised as a means of achieving human emancipation. To this end, various governments make laws and policies towards ensuring that their respective citizens have equal access and opportunities to quality education. The National Policy on Education (Federal Republic of Nigeria [F.R.N.], 2014: 2) prescribed the goals of education to include "provision of equal access and opportunities to qualitative education for all citizens at all levels of education within and outside the formal school system". The implication is that items in a test developed and administered to test takers by public examination boards in Nigeria should not function differently among the test takers based on gender. This is very essential because Physics is a compulsory subject for any NBC/NTC holder aspiring to become an Engineer, Medical Doctor, Teacher, Nurse, among others. Physics test items should measure the knowledge and skills taught in school and are fair to all the takers irrespective of their gender.

Statement of the Problem

Despite the importance of test as a tool for innovative pedagogical approach especially in the post COVID 19 era, some educators and researchers are raising concerns on the issue of test takers who have the same abilities endorsing an item differently due to their gender. Why should test takers of the same latent trait but belonged to different gender respond differentially to a Physics item? Besides, the researchers have not found any study with enough empirical evidence on NABTEB Physics multiple choice items containing gender-based DIF. In addition, there is paucity of information on gender-based DIF in NABTEB Physics multiple choice test, as most of the studies in literature on gender-based DIF focused on Senior School Certificate (SSC) Physics multiple choice items of WAEC and NECO, thus neglecting the NABTEB NBC/NTC Physics multiple choice items. Therefore, the quest to ascertain the gender-based DIF status in NABTEB Physics multiple choice test called for the study.

Objective of the Study

The objective of the study is to integrate gender-based DIF analysis as an innovative pedagogical strategy: An x-ray of National Business and Technical Certificate examinations. Specifically, the objectives of the study are to identify:

1. Gender-based DIF in NABTEB May/June 2020 NBC/NTC Physics multiple choice test.
2. The group favoured by each item that exhibited gender-based DIF in the NABTEB May/June 2020 NBC/NTC Physics multiple choice test.

Research Questions

The following research questions guided the study:

1. Which item(s) in the NABTEB May/June 2020 NBC/NTC Physics multiple choice test exhibited gender-based DIF?
2. Which group was favoured by each that exhibited gender-based DIF in the NABTEB May/June 2020 NBC/NTC Physics multiple choice test?

Hypothesis

The null hypothesis tested at 0.05 level of significance guided the study**.**

1. There is no item(s) in the NABTEB May/June 2020 NBC/NTC Physics multiple choice test that significantly exhibited gender-based DIF.

Method

An ex-post facto research design was adopted for the study. The design was used to unravel factors related to already existing state of affairs and to carry out retrospection on plausible casual factors (Cohen, Lawrence & Morrison, 2007 cited in Ugwoke, 2021). The population comprised all the 23,769 candidates that sat for NABTEB 2020 May/June NBC/NTC Physics multiple choice test while the entire 23,769 candidates were selected by census. May/June 2020 NBC/NTC Physics multiple choice test obtained from NABTEB was used for data collection. It contained 50 items and was used for the study without modifications. The instrument belongs to a standard, public examination body, NABTEB and cannot be modified by external validation process. The reliability was as established by NABTEB too and cannot also be modified by external reliability process.

The data for this study which comprised 23,769 candidates' item response matrix on 2020 May/June NBC/NTC Physics Multiple Choice tests were collected from NABTEB in soft copy. The electronic copy of the dichotomously scored response (1 for correct answer or 0 for wrong answer) which highlighted the subgroups (male and female) were obtained from NABTEB and Binary Logistic Regression of SPSS 21 version was used to analyse gender-based DIF for each item. Descriptive statistics using frequency was also used in identifying the number of correct response for each item that functioned differently by gender. The results of data analysis provided answers to research questions while hypothesis was tested at 0.05 level of significance using Wald statistic. Decision rule was any item with probability, p-value less than or equals to .05 level of significance was considered to have significantly exhibited gender-based DIF.

Results

Research Question One

Which item(s) in the NABTEB May/June 2020 NBC/NTC Physics multiple choice test exhibited gender-based DIF?

Table 1

Result of Binary Logistic Regression Analysis for gender-based DIF in the NABTEB 2020 May/June NBC/NTC Physics multiple choice test; N = 23,769; Male = Reference group; Female = Focal group

Item	B	S.E.	Wald	df	Sig.	95% C.I. for EXP(B)		Flagged DIF (F)	DIF Against
						Lower	Upper		
1	-.098	.064	2.305	1	.129	.799	1.029		
2	.045	.065	.489	1	.485	.922	1.187		
3	-.186	.048	**15.134**	1	.000	.756	.912	Flagged	Female
4	-.095	.045	4.468	1	.095	.832	.993		
5	-.020	.065	.096	1	.757	.863	1.113		
6	.066	.046	2.078	1	.149	.976	1.169		
7	.207	.077	**7.300**	1	.007	1.058	1.429	Flagged	Male
8	-.011	.049	.050	1	.824	.898	1.089		
9	-.267	.048	30.715	1	.132	.697	.842		
10	-.003	.065	.002	1	.964	.878	1.133		
11	-.231	.054	18.604	1	.556	.714	.882		
12	-.152	.044	**11.705**	1	.001	.787	.937	Flagged	Female
13	-.059	.044	1.762	1	.184	.864	1.028		
15	-.149	.042	**12.689**	1	.000	.793	.935	Flagged	Female
16	-.136	.060	**5.136**	1	.023	.776	.982	Flagged	Female
17	.133	.047	7.862	1	.515	1.041	1.253		
18	.061	.052	1.357	1	.244	.959	1.178		
19	-.040	.046	.777	1	.378	.878	1.051		
20	-.015	.050	.092	1	.762	.893	1.086		
21	-.081	.051	2.453	1	.117	.834	1.020		
22	.281	.045	**38.120**	1	.000	1.211	1.448	Flagged	Male
23	.106	.044	**5.810**	1	.016	1.020	1.211	Flagged	Male
24	-.158	.049	10.630	1	.765	.776	.939		
25	.205	.049	17.178	1	.231	1.114	1.353		
26	-.357	.061	**34.339**	1	.000	.621	.789	Flagged	Female
27	-.246	.049	25.230	1	.631	.710	.860		
28	.017	.053	.104	1	.747	.917	1.129		
29	.017	.052	.111	1	.739	.919	1.126		
30	.179	.048	**13.801**	1	.000	1.088	1.314	Flagged	Male
31	-.012	.048	.067	1	.796	.899	1.085		
32	-.036	.049	.525	1	.469	.876	1.063		
33	.056	.047	1.391	1	.238	.964	1.160		
34	.105	.052	**4.003**	1	.045	1.002	1.231	Flagged	Male
35	.068	.049	1.949	1	.163	.973	1.177		
36	.024	.047	.269	1	.604	.935	1.123		
37	-.050	.043	1.341	1	.247	.874	1.035		
38	-.129	.047	7.505	1	.636	.801	.964		
39	-.178	.053	11.301	1	.251	.755	.929		
40	.215	.051	**18.031**	1	.000	1.123	1.369	Flagged	Male
41	-.044	.065	.470	1	.493	.842	1.086		
42	.433	.049	78.638	1	.250	1.401	1.696		

43	-.167	.063	7.052	1	.882	.748	.957		
44	-.303	.052	34.031	1	.741	.667	.818		
45	.174	.046	14.215	1	.562	1.087	1.302		
46	-.206	.056	**13.724**	1	.000	.729	.907	Flagged	Female
47	-.069	.048	2.072	1	.150	.850	1.025		
48	.179	.054	10.863	1	.321	1.075	1.331		
50	-.014	.048	.079	1	.779	.898	1.084		

Table 1 showed statistics of 2020 May/June NBC/NTC Physics multiple choice test items that contained gender-based DIF. It revealed that the Physics test which consisted 50 items had 12 items that contained gender-based DIF. Specifically, the item numbers 3, 7, 12, 15, 16, 22, 23, 26, 30, 34, 40 and 46 contained gender-based DIF. Item number 36 was detected as constant during analysis and removed from the result.

Research Question Two

Which group was favoured by each item that exhibited gender-based DIF in the NABTEB May/June 2020 NBC/NTC Physics multiple choice test?

Table 2

Result of Binary Logistic Regression Analysis for gender-based DIF in the NABTEB 2020 May/June NBC/NTC Physics multiple choice test; N = 23,769; Male = Reference group; Female = Focal group

Item	B	S.E.	Wald	df	Sig.	95% C.I. for EXP(B)		Flagged DIF (F)	DIF Against
						Lower	Upper		
3	-.186	.048	**15.134**	1	.000	.756	.912	Flagged	Female
7	.207	.077	**7.300**	1	.007	1.058	1.429	Flagged	Male
12	-.152	.044	**11.705**	1	.001	.787	.937	Flagged	Female
15	-.149	.042	**12.689**	1	.000	.793	.935	Flagged	Female
16	-.136	.060	**5.136**	1	.023	.776	.982	Flagged	Female
22	.281	.045	**38.120**	1	.000	1.211	1.448	Flagged	Male
23	.106	.044	**5.810**	1	.016	1.020	1.211	Flagged	Male
26	-.357	.061	**34.339**	1	.000	.621	.789	Flagged	Female
30	.179	.048	**13.801**	1	.000	1.088	1.314	Flagged	Male
34	.105	.052	**4.003**	1	.045	1.002	1.231	Flagged	Male
40	.215	.051	**18.031**	1	.000	1.123	1.369	Flagged	Male
46	-.206	.056	**13.724**	1	.000	.729	.907	Flagged	Female

The table 2 shows that out of the 12 items detected to exhibit DIF, 6 items which are 7, 22, 23, 30, 34 and 40 exhibit DIF against the male test takers while 6 items which are 7, 7, 12, 15, 16, 22, 13, 26, 30, 33, 40, and 46 exhibit DIF against the female Physics test takers. Items 14 and 49 were detected as constants during analysis and consequently removed from the result.

Hypothesis

There is no item(s) in the NABTEB May/June 2020 NBC/NTC Physics multiple choice test that significantly exhibited gender-based DIF.

Table 3

Wald test statistic showing the significance of gender-based DIF in NABTEB May/June 2020 NBC/NTC Physics multiple choice test.

Item	B	S.E.	Wald	Df	Sig.	Level of significance	Decision
3	-.186	.048	**15.134**	1	.000		
7	.207	.077	**7.300**	1	.007		
12	-.152	.044	**11.705**	1	.001		
15	-.149	.042	**12.689**	1	.000		$H_{04:}$
16	-.136	.060	**5.136**	1	.023		Rejected
22	.281	.045	**38.120**	1	.000	.05	
23	.106	.044	**5.810**	1	.016		
26	-.357	.061	**34.339**	1	.000		
30	.179	.048	**13.801**	1	.000		
34	.105	.052	**4.003**	1	.045		
40	.215	.051	**18.031**	1	.000		
46	-.206	.056	**13.724**	1	.000		

The data in table 3 showed that the NABTEB 2020 May/June NBC/NTC Physics test item numbers: 3, 7, 12, 15, 16, 22, 23, 26, 30, 34, 40, and 46 have p-values of .000, .007, .001, .000, .023, .000, .016, .000, .000, .045, .000, and .000 respectively. The p-value for each of the 12 items is less than 0.05 alpha levels; therefore, the null hypothesis was rejected. The researchers therefore, concluded that some items in the NABTEB 2020 May/June NBC/NTC Physics multiple choice test significantly exhibited gender-based DIF.

Discussion

Table 1 revealed that 12 out of 50 items in the NABTEB 2020 NBC/NTC Physics multiple choice test exhibited gender-based DIF. This finding is in line with that of Sapmaz (2019) who investigated gender-based DIF and identified three (3) out of 40 items that had DIF. However, whereas the study was based on Mathematics, this study assessed gender-based DIF in Physics test.

Table 2 revealed that six items that contained gender-based DIF were against the males while the remaining six items were against the females. This result was corroborated by Yildirim (2015) who investigated gender-based DIF and found that one of the items which exhibited DIF favoured the females while the remaining 19 items favoured the males. However, Tshering (2006) who performed gender-based DIF analysis of 910 dichotomized items of English Reading Comprehension and Vocabulary for the Dutch Secondary School Education noted that four (4) items of the test functioned differentially across gender, but did not indicate which of the gender (male test takers or female test takers) the DIF items favoured. Differential item functioning affects test validity. Ugwoke (2021) traced gender-based response differences in science to item format effect.

The result of the study was equally corroborated by Osamede et al. (2016) who investigated gender-based DIF in NABTEB 2015 NBC/NTC Mathematics multiple choice test and found that 17 items

(34%) contained DIF. However, whereas Mathematics multiple choice test was used by the researchers, this study used 2020 NBC/NTC Physics multiple choice test. The result also showed that the hypothesis was rejected which implied some items in the NABTEB 2020 May/June NBC/NTC Physics test significant exhibited gender-based which is in line with that of Zimba and Chafutwa (2019) who re-examined gender differential item functioning (DIF) in Malawi School Certificate of Education (MSCE) Computer Studies examination items using a Classical Test Theory (CTT)-based Mantel-Haenszel (MH) DIF detection procedure. The researchers detected three (3) out of 34 items representing 8.82% that had significant DIF.

Conclusion and Recommendations

Concluding, some items in the NABTEB 2020 May/June Physics multiple choice test contained gender-based DIF. The researchers therefore, strongly recommended that:

1. Teachers across the globe should be trained on DIF analysis to enhance the development of quality formative and diagnostic tests; and
2. Test development officers of testing agencies and examination bodies across the country carry out DIF analysis as a routine task during test development processes to eliminate biased items and enhance test quality.

References

Abedalaziz, N. (2010). Detecting gender related DIF using Logistic Regression and Mantel-Haenszel. *Approaches Procedia Social and Behavioral Sciences*, 7(C), 406–413

Adom, D, Mensah, J. A. & Dake, D.A. (2020). Test, measurement, and evaluation: Understanding and use of the concepts in education. *International Journal of Evaluation and Research in Education* (IJERE), 9(1), 109-119. http//:doi: 10.11591/ijere.v9i1.20457

American Educational Research Association, American Psychological Association & National Council on Measurement in Education ([AERA, APA & NCME]. 2014). *Standards for educational and psychological testing.* American Educational Research Association.

Anigbo, L. C. (2014). *Teachers' Handbook on Measurement and Evaluation*. Executive Press Resources Limited, Enugu.

Bottomley, Andrew; Alexander de Graef,; Groenvoid, M. Gundy, C. Koller, M.; Petersen, M. A.; & Spranger, M. G. (2010). The quality of life group and quality of life cross-cultural meta-analysis group. Differential item functioning (DIF) analyses of health-related quality of life instruments using logistic regression. *Health and Quality of Life Outcomes*; 8(81)

CIVIS (2020). Innovative pedagogies: Ways into the process of learning transformation https://civis.eu/storage/files/innovative-pedagogies-ways-into-the-process-of-learning-transformation.pdf

Karami, H. & Nodoushan, M. A. S. (2011). Differential item functioning (DIF): Current problems and future directions. *International Journal of Language Studies* (IJLS), 5(3), https://files.eric.ed.gov/fulltext/ED521872.

Mynbayeva, A., Sadvakassova, Z. & Akshalova, B. (2017). Pedagogy of the Twenty-First Century: Innovative Teaching Methods. In O. B. Cavero, & N. Llevot-Calvet (Eds.). New Pedagogical Challenges in the 21st Century – Contributions of Research in Education. IntechOpen. http//:doi: 10.5772/ intechopen.72341Scott, N. W.; Fayers, P. M.; Neil, K.

Naar, D. (2021). What are the advantages and disadvantages of teacher-made test? https://www.reference.com/world-view/advantages-disadvantages-teacher-made-tests-dfe7c2c05a85d3eb

Nisbet, I. & Shaw, S. (2019). Is assessment fair? Pre-Conference Workshop, IAEA, Baku 22nd September, 2019

Ojerinde, D.; Popoola, K.; Ojo, F. & Onyeneho, O. P. (2012). *Introduction to item response theory parameter models, estimation and application (2nd ed).* Ogbomosho: Marvellous Mike Press Ltd

Osadebe, P.U. & Agbure, B. (2018). Assessment of differential item functioning in social studies multiple choice questions in basic education certificate examination. *European Journal of Education Studies* 4(9)

Osamede, K., Omorogiuwa, & Iro-Aghedo, E. P. (2016). Determination of differential item functioning by gender in the National Business and Technical Examinations Board (NABTEB) 2015 mathematics multiple choice examination. *International Journal of Education, Learning and Development*. 4(10), 25-35.

Sa'ad, N., Ali, A. A. & Abdullahi, A. (2020).Differential item functioning of 2014 NECO English Language Examination in North Senatorial District of Kano State, Nigeria. *Kano Journal of Educational Psychology (KaJEP, 2(2).*

Sapmaz, Z.M.(2019). *Detection of gender-related differential item functioning in Mathematics subsets in Turkey* [Published Graduate Theses and Dissertations]. Scholar Commons, University of South Florida.

Şekercioğlu, G. (2018). Measurement invariance: Concept and implementation. *International Online Journal of Education and Teaching* (IOJET), 5(3), 609-634.

Skinner, C, E. (2002). *Educational Psychology*, (4th ed). Prentice-Hall.

Ugwoke, M. E. (2021, August). Assessment of differential item functioning in physics multiple choice tests of the National Business and Technical Examinations Board. *Unpublished Ph.D. Thesis,* Enugu State University of Science and Technology, Enugu, Nigeria.

UNESCO (2022). UNESCO's education response to COVID-19 https://www.unesco.org/en/covid-19/education-response, 4th November.

Walker, C. M. (2011). What's the DIF? Why differential item functioning analyses are an important part of instrument development and validation. *Journal of Psycho-educational Assessment*, 29(4), 364–376.

Yildirim, H. (2015). An investigation of item bias of mathematics subtest in the 2012-year Level Determination Exam. *Master`s thesis*. Gazi University. Educational Science Institutes. Department of Measurement and Evaluation in Education. (February)