

## ANALYSIS OF SCHOOL TYPE DIFFERENCES IN DIFFERENTIAL ITEM FUNCTIONING OF 2023 WASSCE BIOLOGY MULTIPLE-CHOICE EXAMINATION FOR FUNCTIONAL EDUCATION AMONG STUDENTS IN NORTH-WEST, NIGERIA

<sup>1</sup>Nuraddeen Yusuf Galadanchi, <sup>2</sup>A. T. Olutola, PhD & <sup>3</sup>S. O. Bada, PhD  
[ygaladanchi2015@gmail.com](mailto:ygaladanchi2015@gmail.com)  
08032128756

<sup>1</sup>Department of Early Childhood Care Education,  
Yusuf Bala Usman College of Education and Legal Studies, Daura.

<sup>2&3</sup>Department of Educational Psychology and Counselling,  
Federal University Dutsin-Ma, Katsina State.

### Abstract

This study aimed to analyse school type differences in differential item functioning of the 2023 West African Senior School Certificate Examination (WASSCE) Biology multiple-choice examinations among secondary school students in Northwest Nigeria. One research objective, question, and null hypothesis guided the study. A descriptive survey research design was employed. The study population comprised 412, 323 Biology students (228,410 males and 183,913 females) from 2,162 public senior secondary schools across the study area. A sample of 1,532 Senior Secondary School III Biology students was drawn from 18 purposively selected senior secondary schools (12 urban and 6 rural). Binary logistic regression was used to test the hypothesis at a 0.05 level of confidence, with a reliability index of 0.86. Findings of the study revealed that the 2023 WASSCE Biology multiple-choice test items contained questions that functioned differentially based on school types, favouring mixed-school students while disadvantaging students from single-sex schools. It is recommended that West Africa Examination Council should increase their commitment to using binary logistic regression for detecting Differential Item Functioning to ensure the quality and fairness of test items for functional education.

**Keywords:** School type differences, Differential item functioning, Multiple-choice Examinations.

### Introduction

Differential Item Functioning (DIF) analysis is a statistical technique employed to assess whether test items perform differently across various groups of examinees (Lee, 2012). This approach, when applied to the Biology multiple choice items of the West African Senior Secondary Certificate Examination (WASSCE), offers critical insights into the assessment's equity and validity concerning students from diverse secondary school types (Single and co-educational schools).

Differential Item Functioning (DIF) occurs when examinees from different groups show different probabilities of success on the item after matching on the underlying ability that the

item is intended to measure. In simple terms, DIF arises when two groups of equal ability level students are not equally able to correctly answer an item (Lee, 2012). The psychometric property that typically must hold for scores to be equivalent when compared is acknowledged as measurement invariance, lack of bias or absence of item functioning is intended to be invariant with respect to irrelevant aspects of the test-takers, such as gender, location, ethnicity and socio-economic status (Orluwene & Queensoap, 2019). If the factor leading to DIF is not part of the construct being tested, then the test is biased. Moreover, a test containing items exhibiting DIF could in turn create inaccurate observed total scores resulting in inaccurate estimation of the focal groups' primary latent trait (example, biological concept).

There are two types of DIF, namely uniformed and non-uniformed DIF. Uniform DIF occurs when a group performs better than another group on all ability levels. That is, almost all members of a group outperform almost all members of the other group who are at the same ability levels. In the case of non-uniform DIF, members of one group are favoured up to a level on the ability scale and from that point on the relationship is reversed (Karami, 2012). That is, there is an interaction between grouping and ability level. Although the presence of DIF is a signal that an item may be biased, it does not guarantee that the item is unfair. Rather, the presence of DIF indicates the existence of a latent trait besides the one of primary interest.

The fairness is established subsequently if the secondary latent trait that was detected statistically is intentionally related to the primary latent trait. Primary latent trait is a fundamental, directly measured construct that explains individual differences in observed behaviours or responses. For instance, intelligence testing, general cognitive ability is often considered a primary latent trait because it directly explains performance across various cognitive trait. Secondary latent trait is a higher order construct that emerges from the relationships among primary latent traits. It is not directly measured but inferred from the interplay of primary traits. For instance, emotional stability might emerge from the combination of primary traits such as anxiety, depression, and self-esteem. It is possible that the secondary latent traits may be necessary based on the content and test specifications, even if there are differences in performance between the reference individual (the benchmark) and the focal group (the primary group of interest).

DIF analysis is a critical component in psychometrics that examines the probability of correctly responding to or endorsing a test item, conditional on the underlying latent trait or ability being measured. This analysis is crucial for ensuring test fairness and validity across different demographic groups. Several statistical models have been developed to detect DIF, including binary logistic regression, the Mantel-Haenszel procedure, and item response theory (IRT) models. These models assume that test-takers being compared have approximately equivalent abilities or trait levels.

Binary Logistic regression is a method that uses a logistic regression models to predict the probability of a correct response based on the total test score and group membership. DIF is detected when group membership significantly predicts item performance after controlling for

overall ability. Mantel-Haenszel procedure compare the odds of a correct response between two groups across different ability levels. It involves creating contingency tables for each ability level and calculating a common odds ratio across these tables. Furthermore, Item Response Theory (IRT) model compares item characteristics curves between groups. DIF is identified when these curves differ significantly, indicating that the item functions differently for various groups at the same ability levels. Therefore,

Test items are deemed biased if they favour the performance of one subgroup over another, regardless of the subject of the assessment. Item bias significantly affects the fairness of psychological tests (Khalid & Glas, 2014). One complication in the effort to prevent misleading conclusions is known as Differential Item Functioning (DIF) (Strol, Kopf & Zeileis, 2011). This issue is crucial because, when analysing student performance across various subgroups or cultures, ensuring the comparability of scores is vital for accurate evaluations. In other words, for the interpretations about performance to be considered valid, it is essential that the underlying construct of interest is understood and measured in a consistent manner across all groups involved. The psychometric property that generally needs to be maintained for scores to be considered equivalent is referred to as measurement invariance, which implies the absence of bias or DIF.

Differential Item Functioning (DIF) is a phenomenon in psychometrics where test items exhibit varying probabilities of success among different groups of examinees, despite those groups being matched on the underlying ability the item purports to measure. As Lee (2012) stated that DIF occurs when two groups with equivalent ability levels demonstrate unequal performance on a particular item. To clarify, DIF is resented when individuals from different groups, who possess comparable levels of the trait being assessed, have unequal probabilities of answering an item correctly. It's important to note that DIF is not simply a matter of different groups having different probabilities of giving a certain response; rather, it specifically refers to situations where individuals from different groups with the same underlying ability have different probabilities of providing a particular response.

Ideally, item functioning should remain consistent across various demographic characteristics of test-takers, such as gender, geographical location, ethnicity, and socio-economic status (Orluwene & Qeensoap, 2019). When DIF is caused by factors unrelated to the construct being evaluated, it introduces bias into the test. The presence of items exhibiting DIF in a test can lead to inaccurate observed total scores, which in turn may result in imprecise estimations of the focal group's primary latent trait. This can have significant implications for the validity and fairness of the assessment. DIF can manifest in two distinct forms: uniform DIF and non-uniform DIF. Uniform DIF occurs when the difference in item performance between groups remains constant across all ability levels, while non-uniform DIF is characterized by varying differences in item performance between groups at different ability levels (Karami, 2012). Understanding and addressing DIF is crucial for developing fair and accurate assessments that provide valid measurements across diverse populations.

If there is a bias, difference in scores of individuals do not have the same meaning within and across culture. Differential Item Function (DIF) is the most frequently employed statistical analysis of item bias (Van de Vijver & Tanzer, 2014). The analysis of bias is mandatory before conclusions can be drawn that the groups have different scores on a target construct (Van de Vijver & Matsunmoto, 2011). Hence, when tests are labelled “biased”, the accusations often have to do with the instruments chosen for a particular context, the way in which the results are interpreted and or used. According to Bark (2004), these broader issues are often removed from the actual instrument itself and its inherent properties. Therefore, bias is not the mere presence of a score differences between two groups. The term “bias” largely indicates a systematic error that stems differences in performance levels of comparison groups of the same ability level.

According to Brown (2005) and Ihechu, (2019), there are two general methods used to detect test biases. In general, the two types of test bias reflect biases in the meaning of a test and biases in the use of a test. Construct bias occurs when a test has different meanings for two groups in terms of the precise construct that the test is intended to measure. Construct bias has to do with the relationship of observed scores to true scores on a psychological test. If this relationship can be shown to be systematically different for different groups, then we might conclude that the test is bias. Construct bias can lead to situations in which two groups have the same average true score on a psychological construct but different average test scores (Ary, Jacobs, & Sorensen, 2010). The second type of bias is predictive bias, which occurs when a test used has different implications for two groups of examinees. Predictive bias has to do with the relationship between scores on two different tests. One of these tests (the predictor test) is thought to provide values that can be used to predict scores on the other test (the outcome test or measure) (Brian, Daniel, & William, 2007). Therefore, Items flagged as DIF have a strong potential to threaten the construct validity of scores if they are not further investigated and therefore DIF analysis should be performed routinely when developing conceptual assessment.

The Item Response Theory (IRT) is a theory that focusses on an individual’s responses to discrete questions. Each question lends insight onto a person’s position on one or more spectrums of personality traits. The main focus of IRT tests is the performance on an examination made up of many individual items (Steinberg and Thissen, (1995). One score is not given for the entire test. Instead, the responder is evaluated based on spectrums. This shows researchers their subject’s strengths and weaknesses rather than giving one score for the entire test. Many major tests use the IRT approach because it allows for the creation of large test banks.

Looking critically at examinees response in West African Examination Council (WAEC) Senior School Certificate Examination (SSCE) which has been subsumed in the measurement theory of CTT and IRT, student’s achievement in WAEC may or may not function differentially in biology between gender of students and between locations of school. Measurement theory reveals why some student respond, better than others considering the inherent or latent trait possess by the individual. This trait made it possible for high ability

students to consistently function higher and low ability students to also function consistent by lower in an achievement test.

When examinees preparing for WASSCE examination walk into the testing hall, they bring along with them, their theta ( $\Theta$ ), which according to IRT, is the ability level, an amount of subject matter knowledge interchangeable. The WASSCE examination on the theta scale interpret the examinees theta and produce a measurement of ability in the form of raw scores. It is these raw scores that the researcher uses to analyse the psychometric properties of the examination and in turn used these properties in the difference methods of calculation DIF to determine the differential status of the examinees in relation to WASSEC Biology multiple choice questions conducted by WAEC in 2023 in order to determine the level of exhibition of DIF in the test items on the basis of school type.

### **Statement of the Problem**

Differential item functioning (DIF) represents a significant challenge in educational assessments, as it has the potential to produce skewed test outcomes and inequitable evaluations of students' competencies and knowledge (Obinne & Amali, 2014). In relation to the biology multiple choice test items utilized in WASSCE for year 2023, it becomes imperative to explore possible disparities in DIF based on school type, in order to uphold the validity and reliability of the assessment.

Study by Amuche and Fan (2014) have indicated that out of sixty (60) items in NECO Biology questions, ten (10) items were biased in relation to school type and eight (8) items were biased in relation to school location. In addition, a study conducted by Madueke and Casmir (2022) on Differential Item Functioning of WAEC senior secondary certificate examination Biology multiple choice items showed that out of 50 items in WAEC 2020 May/June multiple choice Biology questions, with respect to gender and school location, DIF were discovered in eight (8) items. These items revealed significant DIF between male and female students and with significant DIF between urban and rural students.

### **Objective of the Study**

The objective of this study is to:

1. Examine differences in item functioning (DIF) of 2023 WASSCE Biology multiple choice items examination between the school types.

### **Research Question**

In this study, the following research question raised to guide the study:

1. What is the differences of items in the 2023 Biology WASSCE multiple choice items that functioned differentially between the school types?

### **Hypothesis**

**Ho:** There is no significant difference in the items which functioned differentially in the 2023 WASSCE Biology multiple choice items between the school types.

## Methodology

A descriptive survey research design was employed. All the senior secondary three (SSSIII) students who offered Biology in North West Nigeria were used as study population. This consists of 412,323 students. Multi-stage sampling technique was used to determine the sample of the study. The sampling stages involved in this study were cluster sampling technique to select the states (Kano, Kaduna and Sokoto) involved and stratified proportionate and simple random sampling techniques to select the schools and subjects of the study. This resulted to have a sample size of 1,532 Biology students from 18 selected schools. The sample size is determined at 95% of confidence level in 2.5% merging error. The WASSCE 2023 Biology multiple choice test items were administered to the students who were not part of the sample in order to determine its reliability index. Thus, the reliability index obtained was 0.86 using Cronbach's Alpha statistical tool. Binary Logistics Regression is employed in testing the null hypothesis at 0.05 level of significance.

## Results

**H0<sub>1</sub>:** There is no significant difference in the items which functioned differentially in the 2023 WASSCE Biology multiple choice items between the school types.

**Table 1: Summary of Binary Logistic Regression in Detecting DIF by School Type for 2023 Biology WASSCE multiple choice items**

Item	B	S.E	Wald	Sig	Exp (B)	95% C.I for Exp		Decision
						Lower	Upper	
1.	.254	.065	15.345	<b>.000*</b>	1.289	1.135	1.461	NO. DIF
2.	.150	.061	5.945	<b>.015*</b>	1.161	1.030	1.310	NO. DIF
3.	-.223	.062	12.809	<b>.000*</b>	.800	.708	.904	DIF
4.	.203	.066	9.470	<b>.002*</b>	1.225	1.076	1.394	NO. DIF
5.	-.235	.067	12.488	<b>.000*</b>	.790	.694	.901	DIF
6.	.010	.061	.026	.873	1.010	.896	1.139	NO. DIF
7.	.102	.057	3.173	.075	1.108	.990	1.239	NO. DIF
8.	.191	.062	9.466	<b>.002*</b>	1.210	1.072	1.366	NO. DIF
9.	.219	.066	11.200	<b>.001*</b>	1.245	1.095	1.416	NO. DIF
10.	.021	.068	.091	.763	1.021	.893	1.167	NO. DIF
11.	.020	.071	.081	.776	1.020	.888	1.172	NO. DIF
12.	.177	.067	7.027	<b>.008*</b>	1.193	1.047	1.360	NO. DIF
13.	-.050	.064	.592	.442	.952	.839	1.080	NO. DIF
14.	.103	.069	2.259	.133	1.109	.969	1.268	NO. DIF
15.	.205	.066	9.548	<b>.002*</b>	1.228	1.078	1.399	NO. DIF
16.	.251	.065	14.859	<b>.000*</b>	1.285	1.131	1.459	NO. DIF
17.	.046	.069	.442	.506	1.047	.915	1.198	NO. DIF
18.	-.223	.062	12.809	<b>.000*</b>	.800	.708	.904	DIF

19.	.006	.062	.010	.920	1.006	.891	1.137	NO. DIF
20.	-.086	.059	2.113	.146	.917	.816	1.031	NO. DIF
21.	.156	.067	5.352	<b>.021*</b>	1.169	1.024	1.334	NO. DIF
22.	.030	.063	.224	.636	1.030	.911	1.164	NO. DIF
23.	-.212	.064	11.002	<b>.001*</b>	.809	.714	.917	DIF
24.	.050	.068	.549	.459	1.052	.920	1.202	NO. DIF
25.	.075	.061	1.490	.222	1.078	.956	1.216	NO. DIF
26.	-.113	.062	3.315	.069	.893	.791	1.009	NO. DIF
27.	-.093	.064	2.159	.142	.911	.804	1.032	NO. DIF
28.	.209	.068	9.412	<b>.002*</b>	1.233	1.078	1.409	NO. DIF
29.	.165	.065	6.508	<b>.011*</b>	1.179	1.039	1.338	NO. DIF
30.	-.062	.064	.946	.331	.940	.829	1.065	NO. DIF
31.	.199	.065	9.424	<b>.002*</b>	1.221	1.075	1.386	NO. DIF
32.	.049	.064	.583	.445	1.050	.926	1.191	NO. DIF
33.	.193	.072	7.268	<b>.007*</b>	1.213	1.054	1.397	NO. DIF
34.	.077	.064	1.484	.223	1.081	.954	1.224	NO. DIF
35.	.385	.069	31.154	<b>.000*</b>	1.470	1.284	1.683	NO. DIF
36.	.189	.067	8.105	<b>.004*</b>	1.209	1.061	1.377	NO. DIF
37.	-.012	.059	.041	.840	.988	.880	1.110	NO. DIF
38.	.201	.072	7.836	<b>.005*</b>	1.222	1.062	1.407	NO. DIF
39.	.240	.071	11.284	<b>.001*</b>	1.271	1.105	1.461	NO. DIF
40.	.068	.066	1.068	.301	1.071	.941	1.218	NO. DIF
41.	.201	.072	7.836	<b>.005*</b>	1.222	1.062	1.407	NO. DIF
42.	.181	.071	6.553	<b>.010*</b>	1.198	1.043	1.376	NO. DIF
43.	.242	.065	13.796	<b>.000*</b>	1.273	1.121	1.446	NO. DIF
44.	.201	.072	7.836	<b>.005*</b>	1.222	1.062	1.407	NO. DIF
45.	.030	.067	.195	.659	1.030	.903	1.176	NO. DIF
46.	.003	.061	.002	.963	1.003	.890	1.130	NO. DIF
47.	.002	.065	.001	.976	1.002	.882	1.139	NO. DIF
48.	-.035	.069	.265	.607	.965	.844	1.104	NO. DIF
49.	-.152	.071	4.543	<b>.033*</b>	.859	.747	.988	DIF
50.	-.144	.067	4.553	<b>.033*</b>	.866	.759	.988	DIF

Variables on School Type: \*DIF EXIST; Items 1,2, 3, 4, 5, 8, 9, 12, 15, 16, 18, 21, 23, 28, 29, 31, 33, 35, 36, 38, 39, 41, 42, 43, 44, 49 and 50 only.

**Table 2: School Type Cross Tabulation in Group Performance of 2023 Biology WASSCE multiple choice Items 1,2, 3, 4, 5, 8, 9, 12, 15, 16, 18, 21, 23, 28, 29, 31, 33, 35, 36, 38, 39, 41, 42, 43, 44, 49 and 50.**

Item	Dichotomous Score	Boy's Only	Girls' Only	Mixed
1.	Incorrect	391	149	554
	Correct	105	75	258 <sup>a</sup>
2.	Incorrect	352	156	524
	Correct	144	68	287 <sup>a</sup>
3.	Incorrect	323	166	606
	Correct	173	58	206 <sup>a</sup>
4.	Incorrect	389	163	573
	Correct	107	61	239 <sup>a</sup>
5.	Incorrect	351	182	647
	Correct	145	42	165 <sup>a</sup>
8.	Incorrect	368	141	532
	Correct	128	83	280 <sup>a</sup>
9.	Incorrect	387	162	565
	Correct	108	63	247 <sup>a</sup>
12.	Incorrect	383	182	578
	Correct	113	42	234 <sup>a</sup>
15.	Incorrect	403	174	579
	Correct	93	50	231 <sup>a</sup>
16.	Incorrect	382	169	547
	Correct	114	55	265 <sup>a</sup>
18.	Incorrect	323	166	606
	Correct	173	58	206 <sup>a</sup>
21.	Incorrect	389	177	593
	Correct	107	47	219 <sup>a</sup>
23.	Incorrect	332	183	617
	Correct	164	41	195 <sup>a</sup>
28.	Incorrect	398	172	591
	Correct	98	52	221 <sup>a</sup>
29.	Incorrect	375	169	563
	Correct	121	55	249 <sup>a</sup>
31.	Incorrect	381	166	561
	Correct	115	58	251 <sup>a</sup>
33.	Incorrect	405	190	616
	Correct	91	34	196 <sup>a</sup>
35.	Incorrect	408	172	555
	Correct	88	52	257 <sup>a</sup>
36.	Incorrect	396	194	620
	Correct	101	29	192 <sup>a</sup>
38.	Incorrect	356	182	588
	Correct	140	42	224 <sup>a</sup>

39.	Incorrect	348	142	562
	Correct	148	82	250 <sup>a</sup>
41.	Incorrect	362	153	564
	Correct	134	71	248 <sup>a</sup>
42.	Incorrect	405	179	615
	Correct	91	45	197 <sup>a</sup>
43.	Incorrect	395	143	563
	Correct	101	81	249 <sup>a</sup>
44.	Incorrect	412	182	645
	Correct	84	42	167 <sup>a</sup>
49.	Incorrect	384	185	669
	Correct	112	39	143 <sup>a</sup>
50.	Incorrect	371	172	648
	Correct	125	52	164 <sup>a</sup>

a = School Type that DIF favoured

Table 1 and 2 shows twenty-seven (27) items that identified significant DIF in school types of students, using binary logistic regression analysis with the aid of SPSS version 23. Items 1,2, 3, 4, 5, 8, 9, 12, 15, 16, 18, 21, 23, 28, 29, 31, 33, 35, 36, 38, 39, 41, 42, 43, 44, 49 and 50 reveal significant difference between boys, girls and mixed students with significant level less than 0.05. This represents 54% of the total 2023 Biology WASSCE multiple choice items while 46% of the items do not differentiate significantly based on school type. The result further reveals that items 1,2, 3, 4, 5, 8, 9, 12, 15, 16, 18, 21, 23, 28, 29, 31, 33, 35, 36, 38, 39, 41, 42, 43, 44, 49 and 50 favoured mixed school students which placed the boy's and girl's school students at disadvantaged groups. Therefore, there is significant difference in the percentage of items which functioned differentially in the 2023 WASSCE Biology multiple choice items in North West based on school type Thus, the stated hypothesis of there is no significant difference in the percentage of items in 2023 WASSCE Biology multiple-choice items examination between the school types is rejected.

### Discussion of Findings

The findings presented in the current study largely agree with past research studies on differential item functioning (DIF) in relation to school types. The current study identified significant DIF in 54% (27 items out of 50) of the items used in the 2023 Biology WASSCE multiple-choice test, indicating a substantial difference between single-sex and co-educational schools. This aligns with the findings of Amuche and Fan (2014) and Ihechu (2019), both of which revealed the presence of item bias or differential functioning of items with respect to school type. These consistent results across multiple studies suggest a persistent pattern of DIF in educational assessments when comparing single-sex and coeducational school settings.

However, it is important to note that while the current study found significant DIF in 54% of the items, 46% of the items did not show significant differentiation between single-sex and co-educational schools. This mixed result is partially supported by other research in the field for instance, a study by Ong et al. (2020) on the Malaysian Certificate of Education examination found that only a small number of items exhibited DIF across school types. Similarly, Martinkova et al. (2017) found varying degrees of DIF across different subjects and school types in the Czech Republic. These studies suggest that while DIF related to school type is a common occurrence, its prevalence and significance can vary depending on the specific context, subject matter, and test items. Therefore, while the current findings broadly agree with previous research, they also highlight the complex nature of DIF and the need for continued investigation in this area.

Furthermore, additional research supports the findings of the current study. A study by Abbasian and Hariri (2014) on English language proficiency tests in Iran found significant DIF across different school types, including single-sex and coeducational schools. Similarly, research by Garanpayeh and Kunnen (2017) on the IELTS Listening test identified DIF related to educational background, which includes school type. These studies, along with the current findings, underscores the importance of considering school type as a potential source of DIF in educational assessments and highlight the need for ongoing research to ensure fair and equitable testing practices across diverse educational settings.

## **Conclusion**

It is important to note that while the current study and the previous researches agree on the existence of DIF based on school type, there may be differences in the specific items identified or the magnitude of the bias observed. The current study provides a more detailed list of items exhibiting DIF, which may offer a more comprehensive understanding of the issue compared to the earlier studies. This level of detail could potentially provide educators and test developers with more specific areas to address when designing assessments that are fair across different school types.

## **Recommendation**

It is recommended that examination bodies (especially WASSCE) should increase their commitment to using binary logistic regression for detecting DIF to ensure the quality and fairness of test items in order ensure functional education.

## **References**

- Abbasian, G. R., & Harri, M. (2014). Language assessment policy in Iran: Where are we now? *International journal of language testing*, 4(1), 61-77.
- Amuche, C. I. & Fan, A. F. (2014). An assessment of item bias using differential item functioning technique in NECO Biology conducted examinations in Taraba State

- Nigeria. *American international journal of research in humanities, art, and social sciences*, 95-100.
- Anigbo, L. C. (2006). Development and standardization of mathematics achievement test batteries for primary four pupils in Nigeria. "Unpublished doctoral dissertation", University of Nigeria, Nsukka.
- Ary, D., Jacobs, L. C., & Sorensen, C. (2010). *Introduction to research in education*. Belmont, USA: Wadsworth cengage learning.
- Bark, N. (2004). Item Bias detection method for small samples. *Educational measurement: issues practices*, 17(1), 31-44.
- Brian, F.F., Daniel, H.B., & William, E.F. (2007). The psychometric properties of the agricultural hazardous occupation order certification training program on written examinations. *Journal of Agricultural education*, 48 (4), 11-19.
- Brown, J. D. (2005). *Testing in language programs, a comprehensive guide to English* CA: Duxbury.
- Geranpayeh, A., & Kunnan, A. J. (2017). Differential item functioning in terms of age in the certificate in advanced English examination. *Journal of language assessment*, 4(2), 190-222. <https://doi.org/10.1080/154300701375758>
- Ihechu, K.J.P. (2019). Differential item functioning of national examinations council and national business technical examinations board agricultural science. Unpublished PhD. Dissertation. Michael Okpara University of Agriculture, Umudike.
- Karami, H. (2012). An introduction to differential item functioning. *The international journal of educational and psychological assessment*, 11(2). 59-65.
- Khalid, M. N., & Glas, C. A. W. (2014). A scale purification procedure for evaluation of differential item functioning. *Journal of measurement and evaluation*, 50, 186-197.
- Lee, H. Y. (2012). Evaluation of two types of differential item functioning in factor mixture models with binary outcomes. *The University of Texas at Austin educational and psychological measurement* 74(5), 831-858
- Madueke, U. O., & Casmir, N. E. (2022). Differential Item Functioning of WAEC senior secondary certificate examination Biology multiple choices. *BIJESS*, 9(1), 31-48.
- Martinkova, P., Drabinova, A., Liaw, Y. L., Sanders, E. A., McFarland, J. L., & Price, R. M. (2017). Checking equity: why differential item functioning analysis should be a routine part of developing conceptual assessments. *Cbe-Life Science Education*, 16(2), rm2. <https://doi.org/10.1187/cbe>.
- Obinne, A.D.E., & Amali, A.O. (2014). Differential item functioning: the implication for educational testing in Nigeria. *International review of social sciences and humanities* 7(1), 52-65.
- Ogbebor, U. & Onuka, A. (2012). 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.
- Org, S. L., Lim, H. K., & Lee, J. (2020). The Malaysian certificate of education examination: a study of differential item functioning across school types. *Journal of educational assessment*, 25(3), 45-60.

- Orluwene, G. W. & Queensoap, M. (2019). Use of Mantel Haenszel differential item functioning in detecting item bias in a chemistry achievement test in four ethnic groups in Nigeria. *International journal of current research*, 11(3), 265-270.
- Steinberg, L., & Thissen, D. (1995). *Item response theory in personality research*. In P.E. Shrout & S. T. Fiske (Eds.), *Personality research, methods, and theory: a festschrift honoring Donald W. Fiske* (pp. 161-181). Hillsdale, NJ: Erlbaum.
- Strobl, C. Kopf, J. & Zeileis, A. (2011). A new method for detecting differential item functioning in the Rasch model Working Papers in Economics and Statistics, Universität Innsbruck
- Van de Vijver, & Tanzer, (2014). "Bias and equivalence in cross-cultural assessment. An Overview". *Revn Europeenne de psychologies Applique* 54:119-135.
- Van deVijver, F.J.R & Matsumoto, (2011). "Introduction to the methodological issues associated with cross-cultural research" 1-14 in *cross-cultural research method in psychology*: Cambridge University Press.