Ekele, B., Obinne, A. D. E., Adulojo, M. O., & Omale, O. (2024). Language-based differential item functioning in 2021-2022 NECO chemistry exams in North Central Nigeria. *Journal of Research in Science and Mathematics Education (J-RSME), 3*(3), 142-153.

Original scientific paper https://doi.org/10.56855/jrsme.v3i3.1135

Received: 25 July 2024.
Revised: 22 September 2024.
Accepted: 16 November 2024.

Language-Based Differential Item Functioning in 2021-2022 NECO Chemistry Exams in North Central Nigeria

B Ekele[1], A D E Obinne[2], M O Adulojo[3], Omale Onuh[4*]
[1,2,3,4] Joseph Sarwuan Tarka University, Makurdi, Nigeria

Abstract

Purpose: This research examines the presence of Differential Item Functioning (DIF) in the NECO SSCE Chemistry Multiple Choice Test administered in Nigeria's North Central region between 2021 and 2022. DIF refers to differences in test item performance across various student groups. **Methodology:** Using Classical Test Theory (CTT) and Item Response Theory (IRT) frameworks, the study employed an ex-post-facto design to analyze existing data from 3,012 chemistry students. Statistical methods such as factor analysis, local independence testing, and ability estimation were used for data analysis. **Findings:** The findings revealed minimal DIF related to language. The study highlights the need to address biases and inequities in test items and provides recommendations for enhancing fairness, validity, and reliability in assessments. **Significance:** These include revising test items, collaborating with subject experts, offering training on inclusive assessment practices, applying robust sampling techniques, conducting continuous monitoring, and developing targeted interventions to support students affected by DIF.

Keywords: Chemistry, DIF, Language, NECO, North central.

Introduction

Chemistry serves as the cornerstone for technological progress, offering individuals a powerful and practical tool for achieving scientific and technological objectives (Ndifon, Umoinyang & Friday, 2016; Kochkarova & Turgunov, 2023). Beyond its global significance, a solid foundation in Chemistry equips individuals with lifelong learning skills to solve everyday problems effectively. Furthermore, Chemistry education is designed to empower learners with the skills, abilities, and competencies necessary to contribute meaningfully to societal advancement (Okorie & Ugwuanyi, 2019; Ananda et al., 2023).

In Nigeria, Chemistry is a core subject taught at the Senior Secondary School level, reflecting its critical role in national development. Emphasizing Chemistry education at this stage is rooted in the belief that a thorough assessment of students' performance in the subject can significantly drive techno-

* Corresponding author: *Onuh Omale, omaleonuh@gmail.com*

logical advancement. Accurate measurement of students' achievements in Chemistry is thus imperative, serving as a key indicator of their preparedness to contribute to national progress.

Despite the recognized importance of Chemistry education, student performance in the Senior School Certificate Examination (SSCE) has been unsatisfactory. According to the National Examinations Council (NECO) Chief Examiners' Reports from 2018 to 2022, only 55.51% of students on average achieved a minimum credit pass (C6) in Chemistry. Furthermore, performance varied significantly across states, genders, locations, and school types. For example, in 2022, states like Kwara (95.18%) and Kogi (94.98%) recorded high pass rates, while others in the North Central region fell below the 50% benchmark.

These disparities may stem from various factors, including students' lack of interest in Chemistry, ineffective teaching methods (Ochieng, Hemed, & Sebtuu, 2019), inadequate teacher qualifications (Okorie & Akubilo, 2013), and a general disinterest in Chemistry practical assessments (Ezeudu, 2013). Another potential contributor to these variations is Differential Item Functioning (DIF), which occurs when test items exhibit bias due to factors such as language, location, religion, or gender.

While previous research has extensively examined some factors influencing student performance, the role of DIF is gaining attention as a significant contributor. This study seeks to investigate DIF in NECO Chemistry examinations within the culturally diverse North Central region of Nigeria. By identifying and addressing potential sources of bias, the findings aim to enhance the fairness and accuracy of Chemistry assessments and ultimately support improved student outcomes.

Differential Item Functioning (DIF) encompasses a set of statistical techniques used to identify items that behave differently for distinct groups of students. According to Hambleton, Swaminathan, and Rogers, as cited in Madu (2012), an item exhibits DIF when individuals with the same level of ability but from different groups do not have an equal likelihood of answering the item correctly. Ideally, an item's probability of being answered correctly should solely depend on a student's ability level in the specific domain measured by the item, in addition to any pertinent item characteristics like difficulty or discrimination. Differential Item Functioning, also known as "measurement bias," arises when individuals from various groups (typically categorized by factors like gender, social class, location, school type, ethnicity, or religion), possessing identical latent traits (skills or abilities), exhibit differing probabilities of providing specific responses on a test or questionnaire. An item doesn't demonstrate Differential Item Functioning solely because people from different groups do not have varying probabilities of providing particular responses; it specifically demonstrates DIF when individuals from different groups, despite having the same underlying true ability, exhibit distinct probabilities of providing specific responses. When a test-taker's performance on an item task aligns with their abilities, they typically provide a correct response. However, at times, test items may unintentionally introduce demands that differ from what the test developers intended, leading to varying interpretations or meanings among members of different groups or sub-groups within the larger population. Such items are said to exhibit Differential Item Functioning.

Differential Item Functioning (DIF) occurs when the characteristics of an item, such as its difficulty level (b), discrimination level (a), or lower asymptotes (c), as determined by Item Response Theory (IRT), vary among different groups of test takers. Essentially, when certain test items behave differently for a particular subgroup within the overall group being examined, it implies that these items are relatively more challenging for one group compared to another. The researcher's objective is to investigate the presence of DIF within the culturally diverse context of North Central Nigeria.

In North Central Nigeria, a rich tapestry of cultural diversity exists, comprising over twenty (20) distinct ethnic groups, each with its unique customs and traditions that shape the lives of its people. This cultural diversity significantly impacts the academic performance of students within each state of the region. The influence of these cultural factors on students' academic achievements is multifaceted and encompasses various dimensions. As revealed by Adeyemo and Babajide (2012), these factors encompass tribal affiliations, traditional belief systems, cultural practices, occupational backgrounds, peer-related dynamics, as well as parental, social, and economic statuses, all of which exert a substantial influence on students' educational achievements. Furthermore, these influential factors span a wide spectrum, encompassing social elements (also referred to as environmental, acquired, external, or phenotypic factors), which encompass aspects like family backgrounds, school-related dynamics, teacher-related variables, and broader societal influences. Additionally, cultural factors play a pivotal role and encompass a range of elements, including tribal beliefs, moral values, time-honored traditions, languages, and legal frameworks. In this study, particular emphasis is placed on investigating the dominant language, recognizing its significance as a cultural factor that can significantly impact educational outcomes.

Recognizing the intricate relationship between culture and education is essential to ensure that every child has an equitable opportunity to excel in their educational journey. Aduloju and Obademi (2019) underscore the profound interconnection between culture and education, highlighting that culture not only provides the substance of what is to be learned but also serves as the custodian and conveyor of cultural values through education. Culture exerts a pervasive influence across various facets of our lives, including the realm of education. As aptly pointed out by Kauchak and Eggen (2011), students carry with them a legacy of learning that is indelibly shaped by the cultures prevailing within their homes and communities. These cultural influences permeate a student's attitudes and values, either acting as facilitators or impediments to the educational process. Within the purview of this study, multifaceted dimensions of culture that may impact students' achievements in schools are considered. These encompass religion, gender, school type, school location, and language. Each of these cultural factors plays a pivotal role in shaping the educational experiences and outcomes of students.

Numerous researchers have delved into the realm of Differential Item Functioning (DIF), with a particular focus on the role played by gender and location in educational assessments. For instance, Omale (2018) conducted a study assessing item bias in GST 115 among university students. The findings revealed that the test items did demonstrate bias against gender. Similarly, studies by Abedalaziz (2011a, 2011b) and Karami and Nodoushan (2011) underscored the presence of biased test items that disadvantaged students based on their gender. Furthermore, Adedoyin and Mokobi's (2013) investigation into test item bias in Mathematics examinations in Botswana unveiled the existence of gender-related bias in certain items. These collective findings emphasized the need to consider the gender of test takers when developing test items, as these items may function differentially for male and female students. Additionally, the location of students' schools has emerged as a significant factor influencing item bias. Mokobi and Adedoyin (2014) identified that out of the 24 items analyzed, five items exhibited bias related to the location of students' schools, distinguishing between those attending rural schools and those attending urban schools. Notably, the Item Characteristic Curves for these five identified test items displayed dissimilar patterns for students from rural and urban school backgrounds, indicating location-based bias. In essence, these findings emphasize that both gender and school location are critical considerations when designing test items to accurately measure the latent abilities of diverse student groups. Given the ample

body of research findings, it is imperative to conduct a thorough assessment of Differential Item Functioning (DIF) within the context of NECO Chemistry examinations in the culturally diverse North Central region of Nigeria. This region comprises six (6) states and FCT, each characterized by distinct cultural attributes, languages, predominant religions, and school types.

Statement of Problem

The National Examinations Council (NECO) conducts the Senior Secondary Certificate Examination (SSCE) in Nigeria, including the Chemistry as a subject, which serves as a critical determinant of students' academic progress and future opportunities in science and technology. However, despite the standardization process undergone by the examination body, the researcher, being a staff of NECO has observed disparities in how individual test scores differs across cultures, particularly within North Central Nigeria. As observed by the researcher, the indigenes of North Central Nigeria speak over 20 languages, and since Chemistry is a branch of science that deal with chemical structure and compounds, the pronunciation of the concepts in Chemistry by teachers who teach these concepts may differ across the state, also the level of experience of the teachers may also differ, this may lead to some items been of advantage or disadvantage to some group. It is within this purview that the researcher seeks to detect whether there is or there are evidences of DIF in NECO Chemistry of 2021 and 2022 in a Multi-Cultural North Central, Nigeria.

Objective of the Study

1. Assess the ability estimate of students of 2021 and 2022 NECO multiple choice Chemistry test items
2. Determine DIF of 2021 and 2022 NECO Chemistry multiple choice items base on language (State)

Research Questions

The following research questions were raised to guide this study

1. What are the ability estimates of students in 2021 and 2022 NECO Chemistry multiple choice test items?
2. How do the items of 2021 and 2022 NECO multiple choice Chemistry function differentially based on language (state)?

Method

The study employed an ex-post-facto research design, which examines pre-existing data to identify relationships between variables. It was conducted in the North Central geopolitical zone of Nigeria, chosen due to its significance in national assessments. The study's population comprised 227,084 Senior Secondary students who wrote the Chemistry NECO SSCE in the 2021 and 2022 academic sessions within the region. From this population, a sample size of 3012 students was drawn using a multi-stage sampling procedure. This process involved proportionate stratified random sampling to ensure representation across subgroups and systematic sampling to select individuals within each stratum.

Data collection utilized the Chemistry Student Scores Retrieval Proforma (CSSRP), designed to organize students' scores across 62 columns, representing the number of test questions, along with columns for serial number and student sex. The proforma included 1506 rows, corresponding to the sampled students.

To collect the data, a cover letter from the researcher's department was addressed to the National Headquarters of NECO, requesting the release of Chemistry NECO SSCE scores for the years under study. The data were analyzed using Jmetrik software, which provided ability estimates to address the research questions. Research Questions One and Two were analyzed using a combination of statistical and practical significance procedures embedded in Jmetrik. The Cochran-Mantel-Haenszel (CMH) statistic was used to test for statistical significance in detecting Differential Item Functioning (DIF), providing insights into potential biases across subgroups.

Results and Discussions

Research question one: What are the ability estimates of students in 2021 and 2022 NECO Chemistry multiple choice test items?

Table 1. Mean Ability Estimates of 2021 and 2022 Students Scores in NECO Chemistry

2021		2022	
No.	**Ability**	**No.**	**Ability**
1	0.11	1	0.33
2	0.43	2	0.33
3	0.27	3	0.37
4	0.06	4	-0.28
5	0.14	5	0.27
6	0.25	6	-0.66
7	0.43	7	0.27
8	0.23	8	1.01
9	0.01	9	0.14
10	0.28	10	-1.74
11	-0.03	11	-0.62
12	0.70	12	-0.05
13	-0.04	13	-0.92
+ +	+ +		+ +
1494	-120	1494	-0.69
1495	-0.57	1495	-1.27
1496	-1.13	1496	-0.96
1497	-0.14	1497	0.70
1498	-0.27	1498	-1.07
1499	-0.46	1499	0.61
1500	-0.07	1500	-0.65

1501	-0.81	1501	0.12
1502	-0.38	1502	0.73
1503	-0.52	1503	-0.42
1504	-0.70	1504	-0.45
1505	-0.50	1505	-0.20
1506	-0.36	1506	0.41
Mean =	**1.01**		**1.03**
SD =	**0.86**		**0.93**

Key:++ Abridged ability estimate of students' scores 2021 and 2022

Reveals the mean ability estimates of students in NECO 2021 to be 1.01 with SD 0.86 while that of NECO 2022 is 1.03 and SD 0.93 respectively. This indicated that both tests had the mean ability within the acceptable range, the students were of equal ability and their scores could be compared

Research Question two: How do the items of 2021 and 2022 NECO multiple choice Chemistry function differentially based on language (state)?

Table 2. Analysis of DIF of NECO Chemistry of 2021 based on language (State)

Item	χ^2	p-value	Valid N	E.S. (95% C.I.)		Class	Remarks	Decision
1	1.84	0.17	1461	1.34 (0.88,	2.05)	A	Not Sig.	No DIF
2	0.31	0.58	1493	1.09 (0.82,	1.45)	A	Not Sig.	No DIF
3	0.02	0.89	1502	1.02 (0.78,	1.33)	A	Not Sig.	No DIF
4	1.95	0.16	1474	1.24 (0.92,	1.67)	A	Not Sig.	No DIF
5	1.90	0.17	1502	1.25 (0.91,	1.71)	A	Not Sig.	No DIF
6	1.61	0.21	1498	0.85 (0.67,	1.09)	A	Not Sig.	No DIF
7	0.00	0.94	1493	1.01 (0.78,	1.31)	A	Not Sig.	No DIF
8	1.65	0.20	1388	0.69 (0.39,	1.22)	A	Not Sig.	No DIF
9	1.37	0.24	1494	1.19 (0.89,	1.61)	A	Not Sig.	No DIF
10	1.07	0.30	1467	0.82 (0.55,	1.20)	A	Not Sig.	No DIF
11	1.85	0.17	1365	1.59 (0.81,	3.09)	A	Not Sig.	No DIF
12	0.21	0.65	1504	0.95 (0.76,	1.19)	A	Not Sig.	No DIF
13	3.50	0.06	1492	1.28 (0.99,	1.66)	A	Not Sig.	No DIF
14	0.12	0.73	1495	1.05 (0.81,	1.35)	A	Not Sig.	No DIF
15	0.10	0.75	1473	0.94 (0.66,	1.35)	A	Not Sig.	No DIF
16	2.02	0.16	1493	0.78 (0.55,	1.10)	A	Not Sig.	No DIF
17	0.04	0.84	1496	1.02 (0.83,	1.26)	A	Not Sig.	No DIF
18	2.91	0.09	1496	1.23 (0.97,	1.57)	A	Not Sig.	No DIF
19	0.19	0.66	1472	0.92 (0.64,	1.33)	A	Not Sig.	No DIF
20	2.70	0.10	1502	1.28 (0.95,	1.71)	A	Not Sig.	No DIF
21	**0.03**	**0.02**	**1504**	**0.98 (0.80,**	**1.21)**	**A**	**Sig**	**DIF**
22	1.38	0.24	1504	0.85 (0.65,	1.11)	A	Not Sig.	No DIF

23	0.02	0.89	1485	0.99 (0.79,	1.22)	A	Not Sig.	No DIF
24	0.21	0.65	1495	1.07 (0.81,	1.42)	A	Not Sig.	No DIF
25	1.34	0.25	1495	1.16 (0.90,	1.49)	A	Not Sig.	No DIF
26	0.07	0.78	1434	1.05 (0.72,	1.54)	A	Not Sig.	No DIF
27	0.01	0.91	1495	0.99 (0.78,	1.25)	A	Not Sig.	No DIF
28	0.03	0.87	1502	1.02 (0.77,	1.36)	A	Not Sig.	No DIF
29	0.95	0.33	1494	0.90 (0.73,	1.11)	A	Not Sig.	No DIF
30	0.10	0.75	1502	0.96 (0.76,	1.22)	A	Not Sig.	No DIF
31	3.09	0.08	1489	0.79 (0.61,	1.03)	A	Not Sig.	No DIF
32	1.36	0.24	1487	0.87 (0.68,	1.10)	A	Not Sig.	No DIF
33	0.25	0.62	1487	0.90 (0.60,	1.36)	A	Not Sig.	No DIF
34	0.19	0.66	1487	1.05 (0.83,	1.34)	A	Not Sig.	No DIF
35	0.38	0.54	1485	1.12 (0.78,	1.61)	A	Not Sig.	No DIF
36	0.44	0.51	1498	0.93 (0.75,	1.15)	A	Not Sig.	No DIF
37	0.37	0.54	1498	0.91 (0.66,	1.25)	A	Not Sig.	No DIF
38	**1.06**	**0.03**	**1494**	**1.17 (0.87,**	**1.59)**	**A**	**Sig**	**DIF**
39	0.02	0.88	1502	1.02 (0.81,	1.28)	A	Not Sig.	No DIF
40	0.25	0.62	1500	0.95 (0.77,	1.17)	A	Not Sig.	No DIF
41	0.80	0.37	1487	1.13 (0.86,	1.48)	A	Not Sig.	No DIF
42	2.85	0.09	1502	0.84 (0.68,	1.03)	A	Not Sig.	No DIF
43	0.61	0.43	1481	1.17 (0.79,	1.71)	A	Not Sig.	No DIF
44	0.04	0.84	1500	1.02 (0.83,	1.26)	A	Not Sig.	No DIF
45	0.01	0.92	1494	0.99 (0.80,	1.22)	A	Not Sig.	No DIF
46	0.19	0.66	1494	0.94 (0.73,	1.22)	A	Not Sig.	No DIF
47	0.03	0.86	1481	1.02 (0.79,	1.32)	A	Not Sig.	No DIF
48	0.00	0.98	1465	1.00 (0.75,	1.34)	A	Not Sig.	No DIF
9	1.09	0.30	1485	0.85 (0.63,	1.15)	A	Not Sig.	No DIF
50	0.08	0.78	1474	0.95 (0.68,	1.34)	A	Not Sig.	No DIF
51	**3.88**	**0.04**	**1494**	**1.36 (1.00,**	**1.84)**	**A**	**Sig.**	**DIF**
52	0.05	0.81	1502	1.03 (0.82,	1.30)	A	Not Sig.	No DIF
53	0.34	0.56	1500	0.94 (0.76,	1.16)	A	Not Sig.	No DIF
54	1.79	0.18	1487	0.83 (0.63,	1.09)	A	Not Sig.	No DIF
55	0.18	0.67	1502	1.05 (0.85,	1.29)	A	Not Sig.	No DIF
56	0.72	0.40	1481	0.85 (0.58,	1.24)	A	Not Sig.	No DIF
57	0.17	0.68	1500	1.04 (0.85,	1.29)	A	Not Sig.	No DIF
58	0.16	0.69	1494	0.96 (0.78,	1.18)	A	Not Sig.	No DIF
59	0.11	0.74	1494	1.04 (0.81,	1.35)	A	Not Sig.	No DIF
60	0.40	0.53	1481	0.92 (0.71,	1.19)	A	Not Sig.	No DIF

Key: item = item name, χ^2 **=** CMH chi-square statistic**, p-value =** p-value for the CMH statistic**, Valid N =** valid sample size**, E.S. (95% C.I.) =** effect size and 95% confidence interval for the effect size in parentheses**, Class =** Educational Testing Service (ETS) DIF classification.

Table 2 revealed that a total of 57 items have no statistically significant chi-square statistic at the 0.05 level and the effect sizes are all negligible resulting in a classification of **"A"** for all the multiple-choice items of NECO Chemistry 2021 when evaluating DIF for Language (State).

Table 3. Analysis of DIF of NECO Chemistry of 2022 with respect to language (State)

Item	χ^2	p-value	Valid N	E.S. (95% C.I.)	Class	Remarks	Decision
1	0.01	0.91	1461	1.02 (0.68, 1.56)	A	Not Sig.	No DIF
2	0.02	0.88	1493	1.02 (0.77, 1.36)	A	Not Sig.	No DIF
3	0.39	0.53	1502	0.92 0.71, 1.20)	A	Not Sig.	No DIF
4	1.35	0.24	1474	1.20 (0.89, 1.61)	A	Not Sig.	No DIF
5	0.06	0.81	1502	1.04 (0.76, 1.43)	A	Not Sig.	No DIF
6	1.62	0.20	1498	0.86 (0.67, 1.09)	A	Not Sig.	No DIF
7	0.04	0.84	1493	1.03 (0.79, 1.34)	A	Not Sig.	No DIF
8	0.54	0.46	1388	0.81 (0.46, 1.43)	A	Not Sig.	No DIF
9	**6.61**	**0.01**	**1494**	**1.48 (1.10, 2.00)**	**A**	**Sig.**	**DIF**
10	0.98	0.32	1467	0.82 (0.56, 1.21)	A	Not Sig.	No DIF
11	2.80	0.09	1365	0.56 (0.29, 1.11)	A	Not Sig.	No DIF
12	0.35	0.56	1504	0.93 (0.75, 1.17)	A	Not Sig.	No DIF
13	3.25	0.07	1492	1.27 (0.98, 1.64)	A	Not Sig.	No DIF
14	3.16	0.08	1495	1.26 (0.98, 1.63)	A	Not Sig.	No DIF
15	3.26	0.07	1473	0.72 (0.50, 1.03)	A	Not Sig.	No DIF
16	0.04	0.84	1493	0.96 (0.68, 1.36)	A	Not Sig.	No DIF
17	2.57	0.11	1496	0.84 (0.68, 1.04)	A	Not Sig.	No DIF
18	0.33	0.56	1496	0.93 (0.73, 1.18)	A	Not Sig.	No DIF
19	0.88	0.35	1472	0.84 (0.58, 1.21)	A	Not Sig.	No DIF
20	0.62	0.43	1502	1.12 (0.84, 1.50)	A	Not Sig.	No DIF
21	0.63	0.43	1504	1.09 (0.88, 1.34)	A	Not Sig.	No DIF
22	0.27	0.60	1504	1.07 (0.82, 1.40)	A	Not Sig.	No DIF
23	2.46	0.12	1485	0.84 (0.68, 1.04)	A	Not Sig.	No DIF
24	2.76	0.10	1495	1.27 (0.96, 1.68)	A	Not Sig.	No DIF
25	0.45	0.50	1495	1.09 (0.85, 1.40)	A	Not Sig.	No DIF
26	0.43	0.51	1434	0.88 (0.60, 1.29)	A	Not Sig.	No DIF
27	0.35	0.55	1495	1.07 (0.85, 1.36)	A	Not Sig.	No DIF
28	1.14	0.28	1502	0.86 (0.64, 1.14)	A	Not Sig.	No DIF
29	0.01	0.92	1494	1.01 (0.82, 1.25)	A	Not Sig.	No DIF
30	0.04	0.84	1502	0.98 (0.77, 1.24)	A	Not Sig.	No DIF
31	0.35	0.55	1489	0.93 (0.71, 1.20)	A	Not Sig.	No DIF
32	0.18	0.67	1487	1.05 (0.83, 1.33)	A	Not Sig.	No DIF
33	0.69	0.40	1487	0.84 (0.55, 1.27)	A	Not Sig.	No DIF
34	**3.72**	**0.03**	**1487**	**1.26 (1.00, 1.60)**	**A**	**Sig.**	**DIF**
35	3.00	0.08	1485	0.72 (0.50, 1.04)	A	Not Sig.	No DIF
36	0.03	0.86	1498	0.98 (0.79, 1.22)	A	Not Sig.	No DIF

37	0.40	0.53	1498	1.11 (0.81,	1.52)	A	Not Sig.	No DIF
38	**3.88**	**0.04**	**1494**	**1.36 (1.00,**	**1.84)**	**A**	**Sig.**	**DIF**
39	0.05	0.81	1502	1.03 (0.82,	1.30)	A	Not Sig.	No DIF
40	0.34	0.56	1500	0.94 (0.76,	1.16)	A	Not Sig.	No DIF
41	1.79	0.18	1487	0.83 (0.63,	1.09)	A	Not Sig.	No DIF
42	0.18	0.67	1502	1.05 (0.85,	1.29)	A	Not Sig.	No DIF
43	0.72	0.40	1481	0.85 (0.58,	1.24)	A	Not Sig.	No DIF
44	0.17	0.68	1500	1.04 (0.85,	1.29)	A	Not Sig.	No DIF
45	0.16	0.69	1494	0.96 (0.78,	1.18)	A	Not Sig.	No DIF
46	0.11	0.74	1494	1.04 (0.81,	1.35)	A	Not Sig.	No DIF
47	0.40	0.53	1481	0.92 (0.71,	1.19)	A	Not Sig.	No DIF
48	0.76	0.38	1465	0.88 (0.66,	1.17)	A	Not Sig.	No DIF
49	0.45	0.50	1485	1.11 (0.82,	1.50)	A	Not Sig.	No DIF
50	**3.00**	**0.01**	**1474**	**1.00 (0.71,**	**1.41)**	**A**	**Sig.**	**DIF**
51	0.04	0.84	1496	1.02 (0.83,	1.26)	A	Not Sig.	No DIF
52	2.91	0.09	1496	1.23 (0.97,	1.57)	A	Not Sig.	No DIF
53	0.19	0.66	1472	0.92 (0.64,	1.33)	A	Not Sig.	No DIF
54	2.70	0.10	1502	1.28 (0.95,	1.71)	A	Not Sig.	No DIF
55	0.03	0.86	1504	0.98 (0.80,	1.21)	A	Not Sig.	No DIF
56	1.38	0.24	1504	0.85 (0.65,	1.11)	A	Not Sig.	No DIF
57	0.02	0.89	1485	0.99 (0.79,	1.22)	A	Not Sig.	No DIF
58	0.21	0.65	1495	1.07 (0.81,	1.42)	A	Not Sig.	No DIF
59	1.34	0.25	1495	1.16 (0.90,	1.49)	A	Not Sig.	No DIF
60	0.07	0.78	1434	1.05 (0.72,	1.54)	A	Not Sig.	No DIF

(Sources: field work 2023)

Key: item = item name, χ^2 **=** CMH chi-square statistic**, p-value =** p-value for the CMH statistic**, Valid N =** valid sample size**, E.S. (95% C.I.) =** effect size and 95% confidence interval for the effect size in parentheses**, Class =** Educational Testing Service (ETS) DIF classification

Table 3 shows that items 9, 34, 38 and 50 with the chi-square statistic of 6.61, 3.72, 3.88 and 3.00 respectively and $p < 0.05$ indicated that the items function differentially based on language of the respondents. The common odds ratios for these items are 1.48, 1.26, 1.36 and 1.86, which indicated that the items were easier for reference group examinees. However, all other items have no statistically significant chi-square statistic at the 0.05 level and the effect sizes are all negligible resulting in a classification of **"A"** for all the multiple-choice items of NECO Chemistry 2022.

The mean ability estimates of students in NECO 2021 were 1.01 with SD 0.86 while that of NECO 2022 was 1.03 and SD 0.93 respectively. This was buttressed by the hypothesis that indicated the mean ability does not differ significantly. This result is in agreement with that of Adewale (2016) who reported a similar finding from his study on equating a two-year Basic Education Certificate. This finding also corroborates that of Atsua et. al. (2018) who in their separate studies reported that the mean ability estimates of students were not significantly different. The result of the present study may be ascribed to the fact that the groups used in the study were not randomly equivalent and the tests used for the mean

ability estimation were parallel.

In addition, only three (3) items were functioning differently in 2021 and four items were functioning differently in 2022 based on language. This finding aligns with the findings of Queensoap and Orluwene (2017) whose findings revealed that 46 items exhibited biases between the focal group (Ijaw) and reference groups (Hausa, Igbo and Yoruba). The findings also aligned with the findings of Ajeigbe and Afolabi (2014), which show there was the occurrence of DIF items in both Mathematics and English Language multiple-choice items of the OSQE for 2008. Fourteen items representing 28% of the 50 items in the mathematics examination exhibited DIF and 10 items, representing 20% of the 50 items in the English Language examination exhibited DIF.

Conclusion

Based on the findings, it was concluded that through the application of Differential Item Functioning (DIF) analysis, the researcher identified nuances in item behaviour across various demographic groups, shedding light on potential bias and inequity within the examination process. Based on the findings of this study, the following recommendations were made. Examination Bodies should conduct a thorough review of test items to identify and address any biases or inconsistencies that may contribute to Differential Item Functioning (DIF). Examination bodies and government should collaborate with subject matter experts to ensure that test items accurately assess the intended knowledge and skills without favouring specific demographic groups.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Abedalaziz, N. (2011a). A Gender-related differential item functioning of mathematics test items. International *Journal of Educational and Psychological Assessment. 6*(1), 95-100.

Abedalaziz, N. (2011b). Detecting DIF using item characteristics curve approaches. *The International Journal of Educational and Psychology Assessment, 8* (2), 1-15.

Adedoyin, O. O., & Mokobi, T. (2013). Using IRT psychometric analysis in examining the quality of junior certificate mathematics multiple choice examination test items**.** *International Journal of Asian Social Science,3*(4), 992-1011.

Adeyemo, S. A., & Babajide, V. T. (2012), The influence of social and economic disadvantage on students' academic achievement in senior secondary schools physics. *International Journal of Educational Research and Technology. 3 (2): 3-10.*

Aduloju, M. O., & Obademi. O. (2014). *Essentials of Sociology of Education.* Makurdi: Cekard Publishers.

Aduloju, M. O., Obinne, A. D. E & Omale, O. (2018). Detection of Gender Bias in General Study Examination of University of Agriculture Makurdi Using Differential Item Functioning Techniques. *African Journal of Theory and Practice of Educational Assessment* (AJTPEA), *5*, 87-100

Ajeigbe, T. O., & Afolabi, E. R. I. (2014). Assessing unidimensionality and differential item functioning in qualifying examination for senior secondary school students, Osun State, Nigeria. *World Journal of Education,4*(4), 30-37.

Anaekwe, M. C. (2014). Basic Research Methods and Statistics in Education and Social Sciences (3rd edition). Sofie Publicity and Printry Limited, Onitsha, Nigeria.

Ananda, L.R., Rahmawati, Y., & Khairi, F. (2023). Critical thinking skills of chemistry students by integrating design thinking with STEAM-PjBL. *Journal of Technology and Science Education*, *13*(1), 352-367. https://doi.org/10.3926/jotse.1938

Atsua, T.G, Uzoeshi, V.I & Oludi, P. (2018). Equating 2015 and 2016 Basic Education Certificate Examination on Civic Education using Classical Test Theory and Item Response Theory in Oyo state, Nigeria. *Journal of Pristine*, *14*(1), 2250-9593.

Aye, M. A., & Htet, L.O. (2010). An Application of Linear Test Equating Method in Scoring. *Yangon Institute of Education Research Journal, 2*(1), 1 – 15.

Ayodele, C. S. (2013). Transformation of Continuous Assessment Scores among Schools in Nigeria. *European Scientific Journal, 8*(26) 171 – 180.

Eleje, L. I., & Esomonu, N. P. (2018). Test of Achievement in Quantitative Economics for Secondary Schools: Construction and Validation Using Item Response Theory, *Asian Journal of Education and Training 4(1)*, 18 – 28.

Emaikwu, S. O. (2016). Fundamentals of Test, Measurement and Evaluation with Psychometric Theories. 3rdedition, Makurdi. SAP Ltd (selfers Academic Press). 78 – 85.

Erguven, M. (2014). Two Approaches to Psychometric Processes: Classical Test Theory and Item Response Theory. *Journal of Education, 4*(7). 23 – 30.

Essen, C. B., Idaka E. I., & Metibemu, M. A. (2017). Item level Diagnostics and Model. Data fit in Item Response Theory (IRT) Using Bilog-MG3 and IRT PRO 3.0 Programmes. *Global Journal of Educational Research, 16,* 87-94.

Essen, C. B. (2015). Differential item functioning of 2014 unified tertiary matriculation examination mathematics of candidates in Akwa-Ibom State, Nigeria. Unpublished Ph.D thesis. University of Calabar, Nigeria.

Ezeudu, F. O. (2013). Effects of gender and location on students' achievement in Chemistry in Senior Secondary Schools in Nsukka L.G.A. of Enugu State. *Research on Humanities and Social Sciences, 3, (15) 50-53.*

Federal Republic of Nigeria. (2004). *National Policy on Education*. (Revised) Lagos. NERDC Press.

Hagell, P. (2014). Testing Rating Scale Unidimensionality using the Principal Component *Journal of Statistics, 4(3)* Pp. 456-465. http://dx.doi.org/10.4236/ojs.2014,46044

Holmes, F. W., & French, B. F. (2019). *Educational and psychological measurement.* New York: Routledge.

Ibrahim, A. (2018). Differential item functioning: The state of the art. *Jigawa Journal of Multidisciplinary Studies (JJMS), 1*(1), 37-50.

Isaac, O. (2019). The effect of multimedia instructional strategy on secondary school students 'achievement and retention in Biology in F.C.T. Nigeria. *Unpublished M.Ed thesis. Federal University of Agriculture Makurdi Benue State.*

Karami, H., & Nodoushan, M. A. (2011). Differential item functioning DIF: Current problems and future directions. *International Journal of Language Studies, 5*(3), 133-142.

Kauchak, D., & Eggen P (2011). Introduction of teaching*: Becoming a professional. London: Pearson*.

Kochkarova, R. R., & Turgunov, E. (2023). Improving the methodology of teaching chemistry lessons at school with the help of different games. *American Journal of Applied Science and Technology*, *3*(10), 15–19. https://doi.org/10.37547/ajast/Volume03Issue10-04

Linn, R. L., & Harnisch, D. L. (1981). Interactions between item content and group membership on achievement test items. *Journal of Educational Measurement, 18,* 109–118.

Liu, Y., & Maydeu-Olivares, A. (2014). Identifying the source of misfit in item response theory models. *Multivariate Behavioural Research, 49*, 354-371. http://dx.doi.org/10./080/00273171.2014.910744

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale, New Jersey: Erlbaum.

Madu, B. C. (2012). Analysis of gender-related differential item functioning in mathematics multiple choice items adminis-

tered by West African Examination Council (WAEC). *Journal of Education and Practice,3*(8), 71-78.

Mokobi, T., & Adedoyin, O.O. (2014). Location biased items in the 2010 Botswana Junior Certificate Examination Mathematics paper one using the item response characteristics curves. *International Review of Social Sciences and Humanities,7*(2), 63-82.

Ndifon, O. B., Umoinyang. E. I., & Friday, O. I. (2010). Differential item functioning of 2010 Junior Secondary School Certificate Mathematics Examination in Southern Education Zone of Cross River State, Nigeria. *Waser,* 1-30.

Ochieng, O. A., Hemed M. S., & Sebtuu, M. N. (2019*)*. A Study of Performance in Chemistry among lower Secondary Government Schools in Zanzibar. *International Journal of Educational Research, 7(2)*, 221-236.

Okorie, E. U., & Akubuilo, F. (2013). Towards improving quality of education in Chemistry: An investigation into Chemistry teachers' knowledge of Chemistry curriculum. *International Journal of Emerging Science and Engineering. (IJESE), 1, 30 - 34.*

Okorie, E. U., & Uguwanyi, A. A. (2019). An investigation into the extent of use of practical Activities in Teaching Chemistry in Nigerian Schools. *Journal of CUDIMAC (J- CUDIMAC), 6*(1), 37-44.

Omale, O. (2018). Detection of Item Bias in General Studies Examination in Federal University of Agriculture Makurdi Using Deferential Item Functioning Technique. *Unpublished M.ed thesis. Federal University of Agriculture Makurdi Benue State.*

Queensoap, M., & Orluwene, G. W. (2017). Examining differential item functioning in a Chemistry achievement test for students in Nigeria. *International Journal of Education and Evaluation*, *3*(7), 49-57.

Umoinyang, I. E. (2011). *The challenging of removing consistent errors in achievement test using differential item functioning (DIF) detection methods.* Paper Presented at the 30th Conference of the Association of Educational Assessment in Africa, Nairobi, Kenya. *Educational and Behavioral Statistics, 25*, 225–247.