ISSN: 4217 - 4520
Volume 10, Number 3, 2024
Academic Journal of Advancement in Education
The University of Auckland, Auckland, New Zealand.

ASSESSMENT OF ITEM BIAS IN THE EQUATED SCORES OF THE 2021 NECO SENIOR SECONDARY CERTIFICATE EXAMINATION (SSCE) FOR AGRICULTURAL SCIENCE IN NORTH CENTRAL NIGERIA

[1]Amakiri, Hager Atisi Eremina Ph.D and [2]Omale Onuh
[1]Department of Educational Psychology, Guidance and Counselling, Ignatius Ajuru University of Education, Rivers State, Nigeria, [2] Department of Guidance and Counselling, Joseph Sarwuan Tarka University, Makurdi, Benue State, Nigeria

Email:amakiriatisi@yahoo.com,omaleonuh@gmail.com

ABSTRACT

This study investigated the comparability of scores from the 2021 NECO Senior Secondary Certificate Examination (SSCE) for Agricultural Science in North Central Nigeria, specifically focusing on item bias and score equating between internal and external test forms. The research highlights the potential issues in using different test forms, which can introduce variations in difficulty levels, leading to unfair comparisons and impacting decisions such as university admissions and scholarships. A non-equivalent anchor test (NEAT) design was used to compare two test forms with common items, adjusting for differences in difficulty. The population consisted of 97,413 Senior Secondary students in the North Central region, with 89,680 internal and 7,733 external candidates. A sample of 2,682 students was selected using a multi-stage sampling method. Data were collected through a self-structured proforma, "NECO Agricultural Science OMR Score Retrieval Proforma," which included information on test items, student scores, and gender. The data were analysed using Item Response Theory (IRT), with the mean equating method and Exploratory Structural Equation Modeling (ESEM) to address the research questions. The study found significant differences in the mean ability estimates when scaled through mean equating, suggesting that raw and equated scores were not identical. Additionally, 17 items in the internal test and 10 items in the external test exhibited bias based on school type, potentially due to variations in teaching methods, teacher experience, and available resources. The study recommended among others: refining equating methods, ensuring test item fairness, and improving teaching resources to address biases and improve score comparability.

Keywords: ***Score equating, item bias, Agricultural Science, NECO SSCE, Internal and External tests, school type bias, item bias detection,***

INTRODUCTION

Test scores are critical as they significantly impact a student's future, influencing decisions related to university admissions, scholarships, and certification. In Nigeria, major examination bodies, such as National Examination Council (NECO), West African Examination Council (WAEC), National Business and Technical Examination Board (NEBTEB), and Joint Admission and Matriculation Board (JAMB), administer high-stakes tests that can affect students' futures. These bodies administer tests multiple times a year, often requiring different test forms for each session due to security concerns. WAEC, NECO, and NABTEB have both internal and external testing sessions, while JAMB administers several parallel forms annually.

Despite using similar content and formats, different test forms can vary in difficulty, which can lead to unfair comparisons between examinees. If two examinees take different test forms with varying difficulty levels, the one who takes the more difficult test may be unfairly disadvantaged due to potentially lower scores. However, these bodies often use raw scores to determine passing grades and university admissions without adjusting for differences in test difficulty. This practice can disadvantage those who take harder tests, even if their performance is comparable.

To address this issue, it is essential that test forms are designed to be comparable in terms of content, difficulty, and reliability. This ensures fairness in the comparison of examinees' scores. Equating is the process used to make scores from different test forms comparable by adjusting for differences in difficulty. (Uysala & Kilmen 2016). It links scores from different forms so they can be used interchangeably. By using statistical methods like equating, scores can be adjusted to ensure they reflect similar meanings, even if the forms have different difficulty levels. This process is crucial for maintaining fairness and ensuring that test results provide an accurate measure of students' abilities across different test forms. The study aims to investigate potential item bias in the Senior Secondary Certificate Examination for Agricultural Science in NECO in North Central Nigeria, highlighting the importance of score comparability in educational assessments.

Test bias is often examined at the item level, with differential item functioning (DIF) analyses being part of the framework for probing item bias (Ibrahim, 2018). The presence of item bias according to Olabode and Adeleke (2017) will confound the calibration of the invariant item and testees' ability (ability estimate) parameters which are the hallmark of the Item Response Theory (IRT) framework. By the use of the IRT framework in the analysis of test items, psychometricians have found that scores in a test may be invariant from what the test is meant for. Ideally, the probability of answering an item correctly should only depend on an examinee's ability level on the construct being measured by the item, along with any relevant item parameter such as difficulty or discrimination. If the probability also depends on a construct-irrelevant subgroup factor, such as race, ethnicity, socioeconomic status, or gender, then DIF may be present. Which may in turn affect the groups taking the test.

If two groups have different underlying multidimensional ability distributions, the test items are capable of measuring these multiple dimensions, and the dimensions are collapsed into a single dimension (score), item bias may occur. Item bias occurs when a group of interests differs in their performance on the skills being measured by the valid items. This is an unavoidable outcome of most tests and should not be viewed in a pejorative manner. Biased items can lead to biased measurement of ability because the measurement is affected by so-called nuisance factors (Shealy & Stout 2007). Also, Hathcoat, (2013) states that test bias is a fundamentally important issue in a test as pervasive and systematic sources of error can lead to erroneous inferences regarding the interpretation and use of test scores. Bias can also result in systematic errors that distort the inferences made in any selection and classification. If an item is not biased, it has to be fair for every group of test-takers. Given today's social and political climate, Petersen (2008) recommended that all testing programs with high-stakes outcomes should conduct equating studies for gender and major racial/ ethnic subgroups, especially because results are likely to be comparable across these subgroups. Examination bodies also need to conduct studies for major subgroups that could differ in ways related to the ability being measured and/or that comprise a varying proportion of the testing population at different administrations.

Test scores play a crucial role in evaluating students' skills and abilities, especially in high-stakes subjects like Agricultural Science. For these scores to be fair and accurate, the tests used must be well-developed, standardized, and aligned with quality attributes. Without proper test development and the application of appropriate procedures to transform and compare scores, the educational system may face challenges in making fair decisions regarding student performance. These issues may arise, particularly in the context of Agricultural Science, where lack of standardization and score transformation could impact students' outcomes.

Agricultural Science, which involves the cultivation of soil, crop production, and livestock rearing, is facing challenges in Nigerian secondary schools, including low student performance. Factors contributing to this include a lack of standardized assessment, moderation, and score transformation. This is reflected in reports from the National Policy on Education and the chief examiner's analysis of student performance in 2018 and 2019. Data shows a notable performance

difference between Internal and External candidates, with Internal candidates performing slightly better, yet the performance disparity between the two groups remains significant.
This variation in performance, along with differences in population size and mean scores, may lead to inequitable comparisons between the two groups. Since decisions about students' academic progress are often based on these scores, the lack of comparability between Internal and External tests can result in unfair advantages for one group over another. Consequently, this study aims to assess item bias in the equated scores of the 2021 NECO Senior Secondary Certificate Examination (SSCE) for Agricultural Science, specifically focusing on Internal and External multiple-choice test items in North Central Nigeria, to ensure fair and equitable assessment across student subgroups.

Statement of the Problem

In Nigeria, high-stakes exams administered by bodies such as NECO, WAEC, NEBTEB, and JAMB play a crucial role in determining students' futures. These exams are often given in multiple forms across different sessions for security and to track educational trends. However, using different test forms introduces challenges related to score comparability. The difficulty levels of these forms may vary, making it problematic to fairly compare the performance of students who take different versions of the test. This issue becomes especially significant when test results influence major decisions such as passing grades, university admissions, scholarships, and certifications.
The potential discrepancy in difficulty between test forms can lead to unfair comparisons and disadvantage some students, potentially impacting their educational opportunities. This undermines the fairness of the assessment process, as it may not accurately reflect students' true abilities. In addition, test bias, such as Differential Item Functioning (DIF), can further complicate score comparability. DIF occurs when certain test items favor specific subgroups of students based on characteristics like gender, socioeconomic status, or ethnicity, which can lead to unfair conclusions about a student's abilities. For the NECO Senior Secondary Certificate Examination (SSCE) in Agricultural Science, ensuring that both internal and external test forms are comparable is vital to make fair and informed decisions regarding students' academic progression. Without proper equating and addressing item bias, decisions about university admissions, scholarships, and other academic opportunities may be flawed. Therefore, this study aims to investigate potential item bias and perform score equating for the 2021 NECO SSCE Agricultural Science Internal and External multiple-choice test items in North Central Nigeria, ensuring fairer assessments for all students.

Aim and Objectives of the study:

The aim of the study is to investigate the comparability of scores from the 2021 NECO Senior Secondary Certificate Examination (SSCE) for Agricultural Science in North Central, Nigeria. Specifically, the objective of the study are to:

1. Equate scores of 2021 Internal and External NECO SSCE Agricultural Science in North Central Nigeria
2. detect bias items in the resultant equated scores of 2021 Internal and External NECO SSCE Agricultural Science based on school type in North Central Nigeria

Research Questions:

The following research questions were used to guide the study.

1. How equitable are 2021 NECO Agricultural Science examination scores in North Central Nigeria?

2. How biased are the items in the resultant equated score of the 2021 NECO SSCE Agricultural Science Internal and External examination based on school type in North Central Nigeria?

Literature Review

Item Response Theory (IRT)

The Item Response Theory (IRT) was pioneered by Fredrick, Rasch, and Larzarfeld in the 1950s and 1960s. Item Response Theory is a general statistical theory about examinees, item and test performance, and how performance relates to the abilities that are measured by the items in the test. It is a theory of testing based on the relationship between individual performances on a test item and the test takers' levels of performance on an overall measure of the ability that the item was designed to measure. n their 2018 study, Atsua, Uzoeshi, and Oludi compared the scores of JSS III students who took the Basic Education Certificate Examination (BECE) in 2015 and 2016 in Civil Education, using classical test theory for equating. The findings showed that the test data adhered to unidimensionality assumptions and that the 3-PL model provided the best fit for the test data. The ability estimates of students were similar across both tests.

In 2017, LaFlair, May, and Arvizu compared seven equating methods (including Mean, Linear, and Circle-Arc methods) in small-scale language testing. Their results indicated that the Circle-Arc method produced the least equating errors, with ability estimates being consistent across the Mean and Circle-Arc methods.

Asiret and Sunbul's 2016 study evaluated equating methods under varying conditions such as sample size and test difficulty, finding that the Circle-Arc and Mean methods had fewer errors, particularly for small samples. Uysal and Kilmen (2016) compared IRT equating methods for mixed-format tests, revealing that the Stocking-Lord method produced fewer errors, especially for positively skewed ability distributions and longer tests. They also found that groups with similar ability distributions had fewer errors than those with differing distributions. Goodness and Asiegbu (2015) explored biased test items in the 2009 Rivers State JSSCE Business Studies test using Item Response Theory, revealing school-type bias in the test items. Uruemu and Onuka (2013) used Differential Item Functioning (DIF) to identify biased items in NECO Economics questions. They found 10 biased items related to school type and 8 related to school location, recommending the use of DIF to detect biased items in future exams. Kilmen and Demirtasli (2012) compared IRT-based equating methods across different sample sizes and ability distributions. They found that the Stocking-Lord method produced fewer errors, while Mean-Mean and Mean-Sigma methods had more errors. Significant differences were found in the ability estimates between groups with varying sample sizes and ability distributions. Pedrajita (2009) examined item bias using logistic regression between public and private school students and male and female examinees. His study revealed class and gender biases in the Chemistry Achievement Test, highlighting similar findings to the current study in terms of bias detection

METHODOLOGY

The study used a non-equivalent anchor test (NEAT) design to compare two test forms with common items, adjusting for differences. The population consisted of 97,413 Senior Secondary students in the North Central region of Nigeria who took Agricultural Science in the 2021 academic session, with 89,680 internal and 7,733 external NECO SSCE candidates. A sample of 2,682 students was selected using a multi-stage sampling method. Proportionate stratified random sampling was first used to select students from different states, and systematic sampling was then employed to select students' answer scripts at regular intervals. Data were collected using a self-structured proforma titled "NECO Agricultural Science OMR Score Retrieval Proforma" (NASOSRP),

which included information on test items, student scores, and gender. Correct responses were scored 1, and incorrect responses scored 0, with a total possible score of 60. The proforma was validated for face and content validity. The data were analysed using Item Response Theory (IRT), with research questions addressed using the mean equating method and Exploratory Structural Equation Modelling (ESEM). Items were considered biased if the p-value was less than 0.05.

RESULTS

Research Question One:

How equitable are 2021 NECO Agricultural Science Examinations scores in North Central Nigeria?

Table 1: Presents the Agricultural Science 2021 Examination scores (I and E) and their equivalent values using mean equating

Observed Score (X)	Mean ES eqEI	Mean ES eqIE
0	1.4	-1.4
1	2.4	-0.4
2	3.4	0.6
3	4.4	1.6
4	5.4	2.6
5	6.4	3.6
6	7.4	4.6
7	8.4	5.6
8	9.4	6.6
9	10.4	7.6
10	11.4	8.6
11	12.4	9.6
12	13.4	10.6
13	14.4	11.6
14	15.4	12.6
15	16.4	13.6
16	17.4	14.6
17	18.4	15.6
18	19.4	16.6
19	20.4	17.6
20	21.4	18.6
21	22.4	19.6
22	23.4	20.6
23	24.4	21.6
24	25.4	22.6
25	26.4	23.6
26	27.4	24.6
27	28.4	25.6
28	29.4	26.6
29	30.4	27.6
30	31.4	28.6
31	32.4	29.6
32	33.4	30.6

33	34.4	31.6
34	35.4	32.6
35	36.4	33.6
36	37.4	34.6
37	38.4	35.6
38	39.4	36.6
39	40.4	37.6
40	41.4	38.6
41	42.4	39.6
42	43.4	40.6
43	44.4	41.6
44	45.4	42.6
45	46.4	43.6
46	47.4	44.6
47	48.4	45.6
48	49.4	46.6
49	50.4	47.6
50	51.4	48.6
51	52.49	49.58
52	53.21	50.71
53	54.49	51.58
54	55.41	52.58
55	56..06	53.71
56	57.71	54.58
57	58.19	54.06
58	59.71	56.58
59	60.23	57.58
60	61.32	58.84
MEAN	**31.48**	**29.08**
SD	**17.31**	**17.43**

KEY: I = Internal E= External;

Table 1 compares the equated scores for the INTERNAL and EXTERNAL tests using the mean equating method. The "X" column lists 61 observed scores (ranging from 0 to 60) for both tests. The "eqEi" column shows equated EXTERNAL scores for each INTERNAL score, while the "eqIe" column shows equated INTERNAL scores for each EXTERNAL score. For example, a score of 0 on INTERNAL is equivalent to 1.4 on EXTERNAL, while a score of 60 on INTERNAL is equivalent to 61.3 on EXTERNAL. The results indicate that the INTERNAL test is about two points more difficult than the EXTERNAL test, meaning the two tests are not equivalent, which contradicts the requirement for test equating. Additionally, when the EXTERNAL test is placed on the INTERNAL scale, the average scores differ (M = 29.08, SD = 17.43 for EXTERNAL vs. M = 31.48, SD = 17.31 for INTERNAL), suggesting that the two tests do not measure the same construct, Agricultural Science ability.

Research Questions Two: How biased are the items in the resultant equated score of the 2021 NECO SSCE Agricultural Science Internal and External examination based on school type in North Central Nigeria?

Table 2: Model result of the ESEM of 2021 NOCEO SSCE Agricultural Science Internal with School type as covariate

Item	Covariate	Estimate	S.E.	Est./S.E.	P-value	Comment
1.	SCHOOL TYPE	-0.385	0.118	-3.261	0.001	**DIF**
2.	SCHOOL TYPE	0.248	0.067	3.724	0.000	**DIF**
3.	SCHOOL TYPE	0.053	0.081	0.652	0.514	**No DIF**
4.	SCHOOL TYPE	0.037	0.068	0.551	0.581	**No DIF**
5.	SCHOOL TYPE	0.285	0.078	3.657	0.000	**DIF**
6.	SCHOOL TYPE	-0.113	0.178	-0.636	0.525	**No DIF**
7.	SCHOOL TYPE	0.281	0.071	3.972	0.000	**DIF**
8.	SCHOOL TYPE	0.068	0.072	0.944	0.345	**No DIF**
9.	SCHOOL TYPE	-0.114	0.064	-1.768	0.077	**No DIF**
10.	SCHOOL TYPE	0.256	0.119	2.150	0.032	**DIF**
11.	SCHOOL TYPE	0.027	0.084	0.320	0.749	**No DIF**
12.	SCHOOL TYPE	0.161	0.082	1.977	0.048	**DIF**
13.	SCHOOL TYPE	0.043	0.065	0.651	0.515	**No DIF**
14.	SCHOOL TYPE	0.086	0.113	0.764	0.445	**No DIF**
15.	SCHOOL TYPE	-0.279	0.127	-2.206	0.027	**DIF**
16.	SCHOOL TYPE	0.026	0.086	0.297	0.766	**No DIF**
17.	SCHOOL TYPE	0.067	0.065	1.038	0.299	**No DIF**
18.	SCHOOL TYPE	-0.061	0.071	-0.864	0.387	**No DIF**
19.	SCHOOL TYPE	0.070	0.072	0.971	0.332	**No DIF**
20.	SCHOOL TYPE	0.038	0.112	0.342	0.732	**No DIF**
21.	SCHOOL	-0.198	0.081	-2.454	0.014	**DIF**

	TYPE					
22.	SCHOOL TYPE	0.212	0.101	2.089	0.037	**DIF**
23.	SCHOOL TYPE	0.138	0.074	1.858	0.063	**No DIF**
24.	SCHOOL TYPE	0.044	0.066	0.669	0.504	**No DIF**
25.	SCHOOL TYPE	-0.165	0.086	-1.919	0.055	**No DIF**
26.	SCHOOL TYPE	0.102	0.065	1.555	0.120	**No DIF**
27.	SCHOOL TYPE	-0.039	0.117	-0.336	0.737	**No DIF**
28.	SCHOOL TYPE	0.257	0.069	3.745	0.000	**DIF**
29.	SCHOOL TYPE	0.034	0.148	0.229	0.818	**No DIF**
30.	SCHOOL TYPE	0.035	0.092	0.377	0.706	**No DIF**
31.	SCHOOL TYPE	-0.060	0.085	-0.705	0.481	**No DIF**
32.	SCHOOL TYPE	0.091	0.098	0.935	0.350	**No DIF**
33.	SCHOOL TYPE	0.030	0.154	0.195	0.845	**No DIF**
34.	SCHOOL TYPE	-0.178	0.101	-1.767	0.077	**No DIF**
35.	SCHOOL TYPE	-0.305	0.065	-4.705	0.000	**DIF**
36.	SCHOOL TYPE	0.004	0.023	0.172	0.864	**No DIF**
37.	SCHOOL TYPE	-0.027	0.078	-0.340	0.734	**No DIF**
38.	SCHOOL TYPE	-0.078	0.062	-1.254	0.210	**No DIF**
39.	SCHOOL TYPE	0.093	0.062	1.491	0.136	**No DIF**
40.	SCHOOL TYPE	0.116	0.088	1.312	0.190	**No DIF**
41.	SCHOOL TYPE	0.021	0.060	0.341	0.733	**No DIF**
42.	SCHOOL TYPE	0.087	0.091	0.954	0.340	**No DIF**
43.	SCHOOL TYPE	-0.078	0.064	-1.218	0.223	**No DIF**
44.	SCHOOL TYPE	0.006	0.065	0.097	0.923	**No DIF**
45.	SCHOOL TYPE	0.197	0.108	1.828	0.068	**No DIF**
46.	SCHOOL TYPE	0.323	0.062	5.237	0.000	**DIF**

47.	SCHOOL TYPE	-0.095	0.077	-1.230	0.219	**No DIF**
48.	SCHOOL TYPE	0.010	0.062	0.159	0.874	**No DIF**
49.	SCHOOL TYPE	0.017	0.062	0.273	0.785	**No DIF**
50.	SCHOOL TYPE	0.422	0.084	5.046	0.000	**DIF**
51.	SCHOOL TYPE	0.300	0.062	4.848	0.000	**DIF**
52.	SCHOOL TYPE	-0.023	0.072	-0.326	0.744	**No DIF**
53.	SCHOOL TYPE	-0.104	0.064	-1.628	0.103	**No DIF**
54.	SCHOOL TYPE	-0.104	0.063	-1.651	0.099	**No DIF**
55.	SCHOOL TYPE	0.217	0.072	2.989	0.003	**DIF**
56.	SCHOOL TYPE	-0.204	0.076	-2.695	0.007	**DIF**
57.	SCHOOL TYPE	-0.073	0.077	-0.953	0.340	**No DIF**
58.	SCHOOL TYPE	-0.134	0.069	-1.936	0.053	**No DIF**
59.	SCHOOL TYPE	0.008	0.063	0.128	0.898	**No DIF**
60.	SCHOOL TYPE	-0.223	0.071	-3.143	0.002	**DIF**

Key: P-value <0.05 = DIF, P-value>0.05 =No DIF

Table 2 shows the Item Biased Assessment of the 60 item Agricultural Science internal examination with School types as covariate to the model to examine their direct effects on the factors and selected indicators. School Type has significant direct effect on (17) seventeen of the items of the 2021 Agricultural Science internal (Item 1, 2, 5, 7, 10, 12, 15, 21, 22, 28, 35, 46, 50, 51, 55, 56, and 60 with P-value less than 0.05) were functioning differentially with respect to School type. Hence, 17 of the items of the test were biased against students based on their School Type affiliation.

Table 3: Model result of the ESEM of 2021 NOCE Agricultural Science SSCE External with School type as a covariate

Item	Covariate	Estimate	S.E.	Est./S.E.	P-value	Comment
1.	SCHOOL TYPE	0.324	0.139	-2.770	0.206	**No DIF**
2.	SCHOOL TYPE	0.251	0.067	3.763	0.000	**DIF**
3.	SCHOOL TYPE	0.054	0.089	0.607	0.431	**No DIF**
4.	SCHOOL TYPE	0.035	0.068	0.506	0.613	**No DIF**
5.	SCHOOL TYPE	0.282	0.082	3.461	0.001	**DIF**

6.	SCHOOL TYPE	0.110	-0.494	0.621	0.312	**No DIF**
7.	SCHOOL TYPE	0.281	0.072	3.899	0.345	**No DIF**
8.	SCHOOL TYPE	0.067	0.076	0.884	0.377	**No DIF**
9.	SCHOOL TYPE	0.117	-1.764	0.078	0.083	**No DIF**
10.	SCHOOL TYPE	0.256	0.137	1.864	0.052	**DIF**
11.	SCHOOL TYPE	0.026	0.085	0.304	0.761	**No DIF**
12.	SCHOOL TYPE	0.162	0.091	1.794	0.073	**No DIF**
13.	SCHOOL TYPE	0.042	0.067	0.627	0.531	**No DIF**
14.	SCHOOL TYPE	0.087	0.138	0.633	0.527	**No DIF**
15.	SCHOOL TYPE	-0.279	-2.124	0.034	0.543	**No DIF**
16.	SCHOOL TYPE	0.027	0.089	0.300	0.764	**No DIF**
17.	SCHOOL TYPE	0.069	0.066	1.042	0.297	**No DIF**
18.	SCHOOL TYPE	-0.063	-0.853	0.394	0.090	**No DIF**
19.	SCHOOL TYPE	0.068	0.074	0.921	0.357	**No DIF**
20.	SCHOOL TYPE	0.042	0.143	0.293	0.769	**No DIF**
21.	SCHOOL TYPE	-0.200	-2.470	0.014	0.003	**DIF**
22.	SCHOOL TYPE	0.213	0.118	1.810	0.070	**No DIF**
23.	SCHOOL TYPE	0.139	0.077	1.819	0.069	**No DIF**
24.	SCHOOL TYPE	0.044	0.068	0.641	0.522	**No DIF**
25.	SCHOOL TYPE	-0.165	-1.829	0.067	0.076	**No DIF**
26.	SCHOOL TYPE	0.102	0.067	1.523	0.128	**No DIF**
27.	SCHOOL TYPE	-0.041	-0.288	0.773	0.000	**DIF**
28.	SCHOOL TYPE	0.258	0.073	3.559	0.412	**No DIF**
29.	SCHOOL TYPE	0.034	0.191	0.177	0.859	**No DIF**
30.	SCHOOL TYPE	0.034	0.094	0.367	0.714	**No DIF**
31.	SCHOOL	-0.057	-0.641	0.522	0.564	**No DIF**

	TYPE					
32.	SCHOOL TYPE	0.092	0.114	0.801	0.423	**No DIF**
33.	SCHOOL TYPE	0.032	0.202	0.161	0.872	**No DIF**
34.	SCHOOL TYPE	-0.176	0.115	-1.540	0.124	**No DIF**
35.	SCHOOL TYPE	0.004	0.024	0.177	0.859	**No DIF**
36.	SCHOOL TYPE	-0.028	0.079	-0.348	0.728	**No DIF**
37.	SCHOOL TYPE	-0.077	0.062	-1.246	0.213	**No DIF**
38.	SCHOOL TYPE	0.093	0.062	1.491	0.136	**No DIF**
39.	SCHOOL TYPE	0.115	0.089	1.290	0.012	**DIF**
40.	SCHOOL TYPE	0.021	0.067	0.321	0.748	**No DIF**
41.	SCHOOL TYPE	0.087	0.092	0.940	0.347	**No DIF**
42.	SCHOOL TYPE	-0.077	0.064	-1.203	0.229	**No DIF**
43.	SCHOOL TYPE	0.007	0.065	0.101	0.020	**DIF**
44.	SCHOOL TYPE	0.196	0.109	1.790	0.073	**No DIF**
45.	SCHOOL TYPE	0.323	0.062	5.248	0.000	**No DIF**
46.	SCHOOL TYPE	-0.096	0.078	-1.229	0.219	**No DIF**
47.	SCHOOL TYPE	0.010	0.062	0.153	0.879	**No DIF**
48.	SCHOOL TYPE	0.018	0.063	0.285	0.776	**No DIF**
49.	SCHOOL TYPE	0.423	0.084	5.017	0.000	**DIF**
50.	SCHOOL TYPE	0.300	0.062	4.835	0.000	**DIF**
51.	SCHOOL TYPE	-0.024	0.072	-0.333	0.739	**No DIF**
52.	SCHOOL TYPE	-0.104	0.064	-1.623	0.105	**No DIF**
53.	SCHOOL TYPE	-0.104	0.063	-1.647	0.100	**No DIF**
54.	SCHOOL TYPE	0.216	0.072	2.983	0.232	**No DIF**
55.	SCHOOL TYPE	-0.204	0.076	-2.688	0.084	**No DIF**
56.	SCHOOL TYPE	-0.073	0.077	-0.942	0.346	**No DIF**

57.	SCHOOL TYPE	-0.134	0.070	-1.930	0.054	**No DIF**
58.	SCHOOL TYPE	0.008	0.063	0.132	0.895	**No DIF**
59.	SCHOOL TYPE	-0.223	0.071	-3.147	0.002	**DIF**
60.	SCHOOL TYPE	0.068	0.066	1.029	0.304	**No DIF**

Key: P-value <0.05 = DIF, P-value>0.05 =No DIF

Table 3 shows the Item Biased Assessment of the 60 item Agricultural Science external examination with School types as covariate to the model to examine their direct effects on the factors and selected indicators. School Type has significant direct effect on (10) ten of the items of the 2021 Agricultural Science external (Item 2, 5, 10, 21, 27, 39, 43, 49, 50, and 59 with P-value less than 0.05) were functioning differentially with respect to School type. Hence, 10 of the items of the test were biased against students based on their School Type affiliation.

Discussion of Findings

The study found significant differences in the mean ability estimates of students' scores when scaled through mean equating using INTERNAL and EXTERNAL tests, indicating that raw and equated scores were not the same. This contrasts with findings from Asiret and Sunbul (2016) and Atsua, Uzoeshi, and Oludi (2018), who reported similar results, and disagrees with Laflair et al. (2017), who found similar results across two test forms using mean and circle arc equating methods. The discrepancy in this study may be due to test administration to groups with similar ability distributions, which aligns with Uysal and Kilmen (2016), who found fewer equating errors in groups with similar ability distributions. Additionally, the result aligns with Kilmen and Demirtasli (2012), who found significant differences in mean ability estimates in different ability groups under various equating methods.

The study also identified biases in the test items: 17 items were biased in the INTERNAL Agricultural Science test of 2021, while 10 items in the EXTERNAL test were biased against students based on school type. This finding is consistent with Pedrajita (2009) and Goodness and Asiegbu (2015), who found class bias in Chemistry Achievement Tests, and Uruemu and Onuka (2013) and Enunwah, Akwa, and Okon (2014), who reported bias in NECO Economics questions based on school ownership types. The bias could be linked to differences in teaching methods, teacher experience, and available school facilities.

CONCLUSION

The study concludes that significant differences were found between the raw scores and the equated scores of students, indicating that the two test for equating (Internal and External) produced different results. The study also identified bias in certain test items, with some items biased based on school type affiliation. These biases may stem from differences in teaching methods, experience, and available resources across schools

RECOMMENDATIONS

Based on the findings of the study, it was recommended that:

1. It is recommended to further explore and refine the equating methods used, especially when tests are administered to groups with similar ability distributions.
2. To minimize bias, test developers should carefully review and evaluate test items to ensure that they are free from bias related to school type or other demographic factors.

3. Schools should focus on improving teaching methods and providing adequate resources and facilities. This will help bridge the gap in performance discrepancies, especially for schools with different types of ownership or resources, and may contribute to reducing biases in test outcomes.
4. Regular assessments should be conducted to check for biases in test items, particularly for large-scale exams. This will allow test developers and educational authorities to detect and correct potential biases before they impact the results.

REFERENCES

Adewale, J.G. (2016). Equating two years of BECE Results in Basic Science and Technology in Oyo State, Nigeria. International Centre for Educational Evaluation (ICEE) Institute of Education, University of Ibadan-Nigeria.

Aduloju, M.O. & Okoh, O. (2021*)* Analysis of Item Difficulty and Discrimination Indices as Functions of changes in Mathematics Objective tests of the Benue State Examination Board Basic Education Certificate (BECE) From 2015 to 2017 *Nigerian Journal of Educational Research and Evaluation Vol 20 PP 163-181.*

Asiret, S. & Sunbul, O. S. (2016). Investigating Test Equating Methods in Small Samples Through Various Factors. *Educational Sciences: Theory and Practice*. 16, 647 – 668.

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.

Federal Republic of Nigeria (2014). *National Policy on Education*. (Revised) Lagos. NERDC Press.

Federal Republic of Nigeria, (2013). National policy on education (6th Edi.) Lagos: Federal Ministry of Education/NERDC

Goodness. O & Asiegbu. C.N 2015 Detecting Item Bias in River State JSSCE Business Study using Item Response Theory (IRT) Approach. *European Journal of Education & Development Psychology Vol 4(1) 1-11.*

Hathcoat, J. D. (2013). Validity Semantics in Educational and Psychological Assessment *Practical Assessment, Research, and Evaluation, 18(9):1-14.*

Hambleton, R., & Rodgers, J. (2006). Item bias review. Practical Assessment, Research, and Evaluation, 4(6). Retrieved November 18, 2022, fromhttp://PAREonline.net/getvn.asp?v=4&n=6.

Huggins, A. C., & Penfield, R. D. (2012). An NCME instructional module on population invariance in linking and equating. *Educational Measurement: Issues and Practice*, *31*(1), 27-40.

He, Y. (2011). Evaluating Equating Properties for Mixed-Format Test. Ph.D. Thesis. Department of Educational Measurement and Statistics, University of Iowa, Iowa City. USA. P. 9 – 39.

Hingorjo, M.R & Jaleel F. (2006). Analysis of One-Best Multiple Choice Questions. The Difficulty Index, Discrimination Index, and Distractor Efficiency. J.Pak. Med. Assoc, 62; 142-147.

Holland, P.W. (2007). A Framework and History for Score Linking. In Dorans, N.J. Pommerich, M & Holland, P.W. (Eds). *Linking and Aligning Scores and Scales*. New York NY; Springer. P. 5 – 30.

Kilmen, S. & Demirtasli, N. (2020). Comparison of Test Equating Methods Based on Item Response Theory According to the Sample Size and Ability Distribution. *Procedia – Social and Behavioural Sciences*, 46; 130 – 134.

LaFlair, G.T, May, L.D. & Arvizu, N.G. (2017). Equating in Small-Scale Language Testing-Programmes. *Language Testing 34*(1). 127 – 144.

Metibemu, M. A & Jonathan U. (2018) Evaluating the appropriateness of using tetrachoric correlation coefficient among items in the assessment of item local independence assumption of multiple-choice test Nigerian *Journal of Educational Research and Evaluation Vol 17 (1) 133-144.*

Mokobi, T. & Adedoyin O. O. (2014). Identifying location-based 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.

NERDC (2013). The Revised 9-Basic Education Curriculum at a Glance. Lagos: NERDC Press. www.nerdcnigeria.gov.ng.

Oku, K., & Iweka, F. (2018). Development, Standardization, and Application of Chemistry Achievement Test using One-Parameter Logistic Model (1-PLM) of Item Response Theory (IRT). *American Journal of Educational Research, 6*(3), 1-58. Retrieved from. http://pubs.sciepublicon education/6/3/11/index.html#cor on 05/09/18.

Olabode J.O & Adeleke J.O (2017) An Investigation Into Differential Item Functioning of 2013 Mathematics Objectives Items conducted by WAEC. Psychological Testing and Innovative Testing Strategies(eds). pp 64-77.

Oluseyi, A.E. (2018). Equating the state Unified and West African School Certificate Mathematics Examination Items. *American International Journal of Contemporary Research*, 8*(4)*, 100 – 106.

Pedrajita, J. Q., & Talisayon, V. M. (2009). Identifying biased test items by differential item functioning analysis using contingency table approaches *A comparative study*, 67, 21-43.

Pedrajita, J.Q. (2009). Using Logistic regression to detect biased test items. *The International Journal of Educational and Psychological Assessment* 2, 54-73.

Peterson, N.S. (2008). A Discussion of Population Invariance in Equating. *Applied Psychological Measurement, 32*, P. 98 – 101.

Shealy, R., & Stout, W. [F.] (2007). A model-based standardization approach that separates true bias/DIF from group ability differences and detects test bias/DIF as well as item bias/DIF. Psychometrika, 58, 159-194.

Uruemu, O., & Adams O. (2013) Differential item functioning method as an item bias indicator. *Educational Research Journal*, 4(4), 367-373.

Uysal, I & Kilmen, K (2016). Comparison of Item Response Theory Test Equating Methods for Mixed Format Tests. *International online journal of Education Sciences 8* (2), 1 – 11. Available at; www.iojes.net.

Wolf, R. (2013). Assessing the Impact of Characteristics of the Test, Common-Items, and Examinees on the Preservation of Equity Properties in Mixed Format Test Equating. A Ph.D. Thesis. Department of Educational Psychology, University of Pittsburgh. Pennsylvania. USA.