International Journal of Innovative Social & Science Education Research 13(1):166-174, Jan.-Mar., 2025

© SEAHI PUBLICATIONS, 2025 www.seahipublications.org ISSN: 2360-8978

doi:10.5281/zenodo.15015583

Assessment Of Item Bias Using Differential Item Functioning (DIF) Technique In NECO Conducted Economics Examination In Taraba State

AMUCHE, Christian Igomu & ZUBAIRU Muhammad Ubandoma

Department of Educational Foundations,
Taraba State University, Jalingo, Nigeria

ABSTRACT
The study evaluates of Item Bias using Differential Item Functioning technique in NECO conducted Economics Examination 2021/2022 in Taraba State. Two research question and two research hypotheses. The ex-post facto research design, the population of this study consists of sixty-five thousand eight hundred and ninety-nine (65,899) candidates that registered and sat for Economics multiple-choice items of NECO SSCE June/July examination (2021/2022) distributed across fourteen education zones in Taraba State. Multi-stage sampling procedure with appropriate technique was used to draw 100 students to same as sample for the study. The NECO 2021 Economics multiple-choice test (NEMT) was the instrument for data collection. The students were provided with an answer sheet by the researcher. From the findings, it is observed that these items that showed DIF are due to the structure of the questions and stem, thus these could be the characteristics that affected the test takers response to getting the item correctly. Base on the findings study recommend that test experts and developer should explore the use of differential item functioning method, particularly the use of logistic regression to detect both uniform and no uniform biased items. A study of this should be conducted to provide further empirical evidence on the validity of the method in detecting biased test items. Evaluators and educational practitioners who are engaged in the development of assessment tools should use logistic regression for bias correction.
Keywords: Differential Item Functioning, NECO, Economics, examination

INTRODUCTION

Achievement tests are frequently utilized in educational and training contexts. For instance, schools often implement achievement tests to evaluate whether students are ready for a specific educational level (Abdullahi, 2014). These tests may be taken by students to assess their readiness for advancing to the next grade or passing a particular subject. Achievement tests serve to measure an individual's prior knowledge. In educational settings, these assessments aim to evaluate a person's competence, achievements, or understanding within a certain domain, a subject that has been examined by various scholars for centuries. In Nigeria, notable national examination bodies include the National Business and Technical Examination Board (NABTEB), the West African Examination Council (WAEC), the National Examination Council (NECO), and the Joint Admission Matriculation Board (JAMB). These organizations cater to candidates from diverse backgrounds throughout the nation. Examinees taking tests administered by these bodies come from different environments and personal circumstances, resulting in a variety of influences on their performance. Therefore, completely eliminating the possibility of bias in test items is challenging.

Bias can be present in items on unidimensional tests. For example, consider these two questions designed to evaluate multiplication skills in economics: (i) What is 6x7? and (ii) How much is the product of six

and seven? While question (i) can be answered solely through mathematical knowledge, question (ii) requires not only mathematical skills but also a level of reading proficiency. Item bias manifests when assessing different skills or when the responses of one group depend on an additional ability, as seen with item (ii). In this case, a test is classified as biased against a certain group if it suggests that individuals from that group perform worse than they actually would in academic or workplace settings. Academic performance assessments can also be affected by this notion of bias. For example, one demographic consistently scores lower than what would be anticipated based on their classroom performance.

The outcomes of the NECO exams may contain test items susceptible to item bias, potentially leading to misleading policy decisions as policymakers evaluate educational strategies and adapt their curricula, teaching approaches, and assessment practices in light of these external examinations. The validity of such assessments or implementations depends significantly on the impartiality of the test items. However, not all examination bodies include item bias detection in their item analyses. An analysis of item bias for the items created by WAEC and NECO is essential to determine the validity of the examination items. Could this be applicable to the test items produced by WAEC and NECO? This study examined the examinees' probability of selecting a correct answer and the types of bias they encountered to evaluate the reliability of the 2021/2022 NECO SSCE Economics multiple-choice examination.

Purpose of the Study

The purpose of the study was to examine the item bias of the examination items constructed by NECO among different subgroups of senior secondary school three students in Taraba State Nigeria. The study compared the fairness of the Economics items constructed by the examination body. Specifically, the study is poised to determine:

i. Whether items set by NECO Economics multiple-choice examination in 2021/2022 June/July functions differently regarding gender.
ii. Whether items set by NECO Economics multiple-choice examination in 2021/2022 June/July functions differently regarding school location.

Research Questions

Based on the objectives of this research study, the study will answer the following research questions:

i. which items on NECO 2021/2022 June/July SSCE multiple-choice Economics examination functions differentially to gender?
ii. which items on NECO 2021/2022 June/July SSCE multiple-choice Economics examination functions differentially to school location?
iii. which items on NECO 2021/2022 June/July SSCE multiple-choice Economics examination function differentially to school ownership?

Differential Item Functioning (DIF)

The term "DIF statistics" describes a growing and evolving collection of statistical techniques that are useful for detecting a regular variation in test and item performance by population subgroups. There is no one method for studying DIF, which is a psychometric nomenclature that is defined as a collection of empirical, statistically-based techniques targeted at a particular but important aspect of item investigation. Moreover, DIF does not suggest a causal relationship or specify the direction of differences that are found.

According to the latent trait or, more generally, the attribute that the items or test are measuring, "differential item functioning" (DIF) refers to how items behave differently in groups, usually demographic groups (Camilli, 2014). DIF is used to determine how items perform in various subgroups (Schnipke, 2000). It occurs when test-takers from various populations with similar skill levels have varying chances of correctly answering a question, claim Roussos and Stout (2000). Ostertind and Everson (2014) offer the following example to help readers gain a comprehensive understanding of DIF measurement bias. In this case, Y—a reaction to a particular test item—is determined by the latent construct of interest being measured. Theta (0) is the name of the latent construct of interest when Y is an indication of (®). Y can be arranged according to the probability distribution of Y on (0) using the expression f(Y(0)). When the same latent trait's preferences tend to act differently in a test item, this is

referred to as differential item functioning (DIF). DIF occurs when test-takers from different subgroups with the same overall test score or ability level systematically differ in their likelihood of completing a particular test item or items, according to Dogan, Guerrero, and Tatsuoka (2015). If the sources of extraneous error are not examined, these subgroups are likely to perform differently. The language of the test item was identified by Johnston and Selepeng (2001) as a potential source of extraneous error. However, Differential Item Functioning (DIF) in bias analysis is expressed in two forms: Uniform and Non-uniform (DIF).

Forms of DIF

The most basic kind of DIF is uniform DIF, in which the degree of conditional dependency is largely constant throughout the spectrum of latent traits (9). Across all ability levels, one group continuously benefits from the item of interest (Walker 2015). When a subgroup of examinees with varying ability levels consistently respond to a specific item or subset of items more consistently than the other group, it is thought that uniform DIF has taken place. As a result, that specific subgroup is considered to have an advantage over the other group and to be more capable than the less favored group. In bias analysis, the less advantaged group is referred to as the "focal" or group of focus, whereas the advantaged group is called the "reference" group (Walker 2015; Huang and Han, 2012).

On the other hand, when ability and group membership interact, non-uniform DIF happens. Higher level members of one group may find an item challenging. It gets harder for Su and Wang, who are at a lower level in the other group, after a certain point. (2015) explained the differences between uniform and non-uniform Dl F based on group membership and ability. They claimed that uniform DIF shows no correlation between group membership and examinees' or respondents' ability or degree of agreement. All of the members of one group outperform all of the members of the other group on an item when it has uniform DIF.

RESEARCH METHODS

This study used the ex-post facto research design to achieve its objectives. The population for this study includes 65,899 candidates who registered for and participated in the multiple-choice Economics sections of the NECO SSCE June/July examination for the 2021/2022 academic year across 14 educational zones in Taraba State. A suitable method and a multi-stage sampling approach were utilized to select a sample of 100 students for the study. Data was gathered using the NECO 2021 Economics multiple-choice test (NEMT). Three experts from Taraba State University Jalingo's Faculty of Education were given copies of the NEMT in order to assess its validity. Logistic regression was used to analyze the data. It involved the following steps:

Result and Findings

Research Question One

Which items on NECO 2021/2022 June/July SSCE multiple-choice Economics examination functions differentially to gender?

Table 1

Logistic Regression to Detect Gender Bias

Item	B	S.E	Sig	Exp (B)	Lower	Upper
1	.157	.225	.483	1.170	.754	1.817
2	.243	.238	.308	1.275	.799	2.035
3	.095	.190	.616	1.100	.758	1.597
4	-.076	.190	.691	.927	.639	1.346
5	-.235	.231	.309	.791	.503	1.243
6	.311	.211	.142	1.364	.902	2.065
7	-.177	.190	.454	. .837	.577	1.216

8	-.339	.191	.**004***	.712	.490	1.035
9	.417	.195	.343	1.517	1.035	2.224
10	.92	.197	.639	1.097	.746	1.613
11	.242	.218	.268	1.273	.831	1.952
12	-.227	.190	.**033***	.797	.549	1.158
13	.663	.201	.531	1.941	1.310	2.876
14	1.039	.361	.**004***	2.826	1.393	5.733
15	.249	.202	.219	1.283	.863	1.908
16	-.959	.266	.**000***.	.383	.227	.646
17	-.023	.191	.905	.977	.672	1.422
18	-.319	.191	.094	.727	.500	1.056
19	.241	.199	.226	. 1.272	.861	.1.879
20	.317	.193	.101	1.373	.941	2.004
21	.163	.247	.509	1.177	.725	1.911
22	.164	.354	.**001***	.897	.6785	1.976
23	-.543	.307	.077	.581	.318	1.061
24	.218	.261	.402	1.244	.747	2.073
25	-.494	.325	.129	.610	.323	1.155
26	-.131	.202	.507	.877	.590	1.304
27	.083	.196	.672	1.087	.740	1.596
28	-458	.266	.085	.632	.375	1.066
29	-.111	.271	.682	.895	.527	1.522
30	.046	.190	.808	1.047	.721	1.521
31	.299	.197	.129	1.349	.916	1.985
32	.122	.256	.635	1.129	.683	1.866
33	.166	.191	.386	1.181	.811	1.718
34	-.141	.216	.513	.868	.568	1.326
35	.204	.198	.290	1.233	.836	1.817
36	.242	.223	.278	1.273	.823	1.971
37	-.140	.201	.486	.869	.587	1.289
38	.374	.287	.192	1.454	.829	2.550
39	.257	.201	.202	1.293	.871	1.918
40	-.326	.198	.100	.722	.489	1.064
41	.086	.191	.653	1.89	.750	1.583
42	.136	.278	.626	1.145	.664	1.974
43	-1.488	.459	.**001***	.226	.092	.555

44	.460	.218	**.034***	1.585	1.035	2.427
45	.065	.215	.761	1.068	.700	1.628
46	.461	.201	**.021***	1.586	1.070	2.450
47	-.209	.280	.455	.811	.469	1.404
48	.263	.207	.203	1.301	.867	1.953
49	.414	.191	**.031***	1.513	1.039	2.202
50	-.506	.228	**.027***	.603	.386	.943
51	.103	.272	.705	1.109	.650	1.891
52	.106	.245	.666	1.112	.688	1.796
53	.134	.216	.536	1.143	.749	1.744
54	-.071	.193	.711	.931	.638	1.358
55	-.161	.207	.437	.851	.567	1.278
56	.255	.248	.305	1.290	.793	2.099
57	.168	.211	.425	1.183	.783	1.788
58	-.014	.246	**.007***	.986	.609	1.595
59	.564	.207	.677	1.758	1.171	2.639
60	-.060	.195	.760	.942	.643	1.381

Using logistic regression, Table 1 displays the items according to gender (male and female). Eleven of the sixty items in the NECO Economics questions contained DIF. Items 8, 12, 14, 16, 22, 43, 44, 46, 49, 50, and 58 are among them.

Research Question Two

Which items on NECO 2021/2022 June/July SSCE multiple-choice Economics examination functions differentially to school location (rural and urban)?

Table 2: Logistic Regression of Sixty NECO Item for School Location

Item	B	S.E	Sig	Exp (B)	Lower	Upper
1	.017	.243	.965	1.017	.550	1.760
2	-.236	.246	**.000***	.788	.512	1.28
3	1.220	.200	.340	3.388	2.290	5.012
4	-.403	.191	0.35	.669	.460	.973
5	.194	.233	.406	1.214	.769	1.917
6	-.217	.209	**.002***	.805	.535	1.212
7	-.840	.194	.254	.432	.295	.632
8	-.339	.191	.075	.712	.490	1.035
9	-.618	.199	**.002***	.539	.365	.796
10	-.370	.197	.059	.690	.469	1.015
11	-.506	.217	**.019***	.603	.394	.921
12	.098	.190	.604	1.103	.760	1.602

13	.107	.199	.591	1.113	.754	1.643
14	-.254	.315	.419	.776	.419	1.437
15	-.116	.201	.562	.860	.600	1.320
16	-.432	.249	.084	.650	.398	1.059
17	-.611	.193	**.002***	.543	.372	.793
18	.370	.191	.053	1.447	.995	2.105
19	.122	.198	.538	1.130	.766	1.667
20	-.017	.193	.928	.983	.674	1.434
21	-.332	.252	.188	.717	.438	1.176
22	.424	.234	.070	1.528	.966	2.417
23	.080	.293	.785	1.083	.610	1.925
24	-.333	.266	.212	.717	.425	1.209
25	-.087	.314	.781	.916	.495	1.695
26	.276	.202	.171	1.318	.888	1.957
27	.199	.196	.311	1.220	.830	1.792
28	.020	.258	.938	1.020	.615	1.693
29	.324	.269	.228	1.382	.816	2.441
30	-.316	.191	.097	.729	.502	1.059
31	.029	.191	.883	1.029	.701	1.511
32	-.143	.259	.580	.867	.522	1.439
33	-.054	.191	**.012***	.948	.651	1.379
34	-.329	.218	.131	.720	.470	1.103
35	.209	.198	.290	1.233	.836	1.817
36	.093	.223	.678	1.097	.706	1.698
37	.181	.200	.366	1.198	.810	1.773
38	.540	.289	.062	1.715	.975	3.020
39	-.068	.202	.737	.934	.629	1.388
40	.333	.197	.091	1.395	.948	2.050
41	-.133	.191	.487	.876	.602	1.273
42	.213	.278	.447	1.237	.718	2.132
43	-.644	.377	.088	.525	.251	1.100
44	-.153	.218	.483	.858	.559	1.316
45	-.405	.219	.065	.667	.434	1.025
46	-.100	.201	.619	.905	.611	1.341
47	-1.069	.314	**.001***	.343	.185	.636
48	-.080	.208	.701	.923	.614	1.387

49	-.463	.192	.061	.629	.432	.918
50	-.251	.224	.263	.778	.502	1.207
51	-.195	.276	.476	.822	.479	1.411
52	-.575	.255	**.024***	.563	.341	.929
53	-.053	.216	**.002***	.948	.621	1.449
54	.598	.194	.365	1.819	1.244	2.661
55	.223	.206	.280	1.249	.834	1.871
56	-.054	.250	.827	.947	.581	1.544
57	.124	.211	.558	1.132	.749	1.710
58	-.258	.248	.299	.773	.475	1.257
59	.265	.206	.198	1.304	.870	1.953
60	- .2 51	.196	.201	.778	.530	1.143

The items in Table 2 are categorized by school location (rural versus urban) using the logistic regression method. DIF appeared in nine of the sixty items in the NECO Economics questions. Items 2, 6, 9, 11, 17, 33, 47, 52, and 53 are among them.

DISCUSSION

The findings indicate that items highlighted by logistic regression concerning gender (male and female) exhibit differential item functioning (DIF). Among the sixty items in the NECO Economics questions, eleven items were found to include DIF. These items are 8, 12, 14, 16, 22, 43, 44, 46, 49, 50, and 58. The analysis also suggests that the structure of the questions and stems contributed to the presence of DIF; thus, these factors may have affected the test-takers' ability to correctly respond to the items. Nworgu (2011) notes that contemporary research has indicated that tests utilized for national and regional assessments vary in their functioning depending on the subgroups to whom they are administered. This suggests that being part of a particular group, rather than actual ability, significantly influences a student's scores on such assessments. Adedoyin (2010), in his investigation into gender bias in public examinations, found that five out of sixteen test items met the statistical criteria of the 3PL item response theory and exhibited gender bias.

DIF items were identified through logistic regression analysis focusing on subgroups, such as students from public and private institutions. It was revealed that ten out of the sixty items on the NECO economics exam paper showed DIF; these items include 3, 8, 10, 22, 29, 31, and 54. From the sixty items in the NECO economics examination question paper, logistic regression identified eight items (2, 6, 9, 11, 17, 33, 47, 52, and 53) that displayed DIF when comparing urban and rural school students. Research conducted by Felder, Mohr, Dietz, and Ward (1994) demonstrated that urban students perform better than their rural peers. This conclusion is supported by findings from Tremblay, Ross, and Berthelot (2001), Kolcic (2006), and Considine and Zappala (2002). However, the conclusions of this study differ from those of Lee and McIntire (2001), who found no significant performance difference between students in rural and urban settings. This implies that certain items designed to assess students' competencies may be biased towards urban examinees and unfair to rural students. The data aligns with the work of Amuche and Fan (2014), who identified biased elements in the 2002 NECO Mathematics test items that favored urban examinees while disadvantaging those from rural backgrounds.

This indicates that certain assessments meant to evaluate students' abilities are prejudiced, favoring those from urban institutions while disadvantaging their rural counterparts. The results imply that the format and content of the questions contribute to the observed Differential Item Functioning (DIF); consequently,

these aspects may have influenced the test-takers' ability to answer the questions correctly. The results of this research align with those of Pedrajita (2009), who employed logistic regression to identify bias in Chemistry Achievement test questions. Among the 22 items administered to the students, the investigation revealed that 11 showed a preference for public schools while the other 11 favored private schools. This suggests that both the type of school the examinees attend and their economic capabilities play a role in their performance on the test items. Chukwudi (2019) corroborated this by finding that 38 out of 60 multiple-choice questions on the 2017 BECE Mathematics exam exhibited bias associated with school ownership, consistent with these findings. Specifically, 31 items were biased against public schools, while seven items were biased against private institutions.

CONCLUSION

Based on the forgoing findings the following conclusions were made. There were presences of gender, school ownership and school location bias in NECO June/July 2021 Economics questions.

RECOMMENDATIONS

Based on the findings and conclusions, the following suggestions are proposed:

1. To recognize both uniformly and non-uniformly biased items, test developers and specialists should explore the use of the differential item functioning method, particularly logistic regression.
2. To provide further empirical evidence for the effectiveness of this method in detecting biased test items, a study should be conducted.
3. For bias correction, educators and evaluators involved in creating assessment tools should employ logistic regression.
4. To develop a valid and reliable gender-fair test, measurement professionals should apply logistic regression, which may involve revising or replacing biased items in a school-type assessment.

REFERENCES

Adebule, O. S. (2019). Reliability and levels of difficulty of objective test of ten senior secondary schools in five local government areas of Akure, Ondo State. *Educational Research and Review, 4(11), 585-587.*

Adebule, S. O. (2015). A comparative analysis of difficult and discriminating indices of objective test items in a Mathematics achievement test. An unpublished M.Ed Thesis, Ondo State University, Ado Ekiti

Agu, N. N., Onyekuba, C., &Anyichie, A. C. (2021). Measuring teachers' competencies in constructing classroom-based tests in Nigerian secondary schools: Need for a test construction skill inventory.

Aiken, L. R. Jnr. (2017): Intelligence variables and mathematics achievement. *Journal of Research in Educational Psychology. 9 (20), 200 - 215.*

Ajai, J. T. &Amuchie, C. I. (2015). *Educational research methods and statistics*. Academica House Publishers Nig. Ltd Jos, Plateau State.

Bell, M. (2019). Defined academic performance. Online at http:/www.chow.com/about474075 define academic performance.htm/.retrieved on 08/06/2021.

Bichi, A. A., Hafiz, H., & Bello, S. A. (2016). Evaluation of Northwest University, Kano Post-UTME test items using Item Response Theory. *International Journal of Evaluation and Research in Education (IJERE. 5(4), 261-270.*

Camilli, A. (2016) *An Introduction to Psychological Assessment and*Psychometricshttp://dx.doi.org/10.4135/9781446221556

Cohen, R. J. (2019). Psychological testing and assessment: An introduction to tests and measurement. (7th ed.). New York: McGraw-Hill.

Cohen, L., Manion, L. and Morrison, K. (2013) Research Methods in Education (6th ed.) London: Routledge.

Iwuji, V. R. (2017). Measurement and evaluation for effective teaching and learning. Owerri, Joe Mankpa's prints.

Linn, R. L., & Gronlund, N. E. (2015). *Measurement and assessing in teaching* (7th ed.). Prentice-Hall Inc.

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

Lord, F.M., &Norvick, M.R. (1968). *Statistical theories of mental test scores.* Menlo Park,Calif.: Addison-Wesley.

Maikudi A.K, Taiwo A.M, UweloD. &Kukwi I .J(2018) examine is study developed and validated an Economics teacher-made test for assessing students' achievement inthe North Central States of Nigeria

Obarasua, &Amadioha (2022)Analyzing the Reliability and Content Validity of Basic Education Certificate Examination Mathematics Objective Test Items.

Olatunji, S. O., & Onofeghara, N. (2018). Psychological testing: foreign and indigenous psychological test. Lagos: San solid Publishers

*Orluwene, G. W., & Ig*we, B. N. (2016). Psychometric properties of WASSCE questions in Civic Education (2015) using Rivers State students. *International Journal of Psychology and Counselling, 8 –15(1), 92 – 115.*

Schultes, M.-T., Kollmayer, M., Mejeh, M., & Spiel, C. (2018). Attitudes toward evaluation: An exploratory study of students' and stakeholders' social representations.

UNESCO (2020), Multistage Sampling retrieved from*http://uis.unesco.org/englossary-term/multi-stage-sampling on 28/04/2020*

Wilmut, J., &Yakasai, I. M. (2016). A brief review of the assessment of student achievement in Kaduna, Kano and Kwara states of Nigeria.

Yang, F. M., & Kao, S. T. (2014). Item response theory for measurement validity. Shanghai Archives of Psychiatry, 26(3), 171-177. doi: 10.3969/j.issn.1002-0829.2014.03.010

Yu, C. H. (2013). A simple guide to the item response theory (IRT) and Rasch modeling. Retrieved from: http://www.creative-wisdom.com

Walker, C. (2015). What's the DIF? Why DIF analyses are an important part of the instrument in development and validation. *Journal of Educational Assessment, 29, 364-376*