



DETECTING DIFFERENTIAL ITEM FUNCTIONING IN 2024 WASSCE ECONOMICS OBJECTIVE TEST ITEMS USING MANTEL-HAENSZEL (MH) STATISTICAL PROCEDURE

LAARO Yahaya Babatunde¹, SIANO, Sabi Zakari², ISMAIL Aminat Omotola³, OLATINWO Taiye Mujidat⁴

¹Department of Social Sciences Education, Faculty of Education, University of Ilorin, Ilorin

²Department of Arts and Sciences Education, Faculty of Education, Kwara State University, Malete

³Department of Social Science Education, Faculty of Education., University of Ilorin, Ilorin

⁴Department of Social Science Education, Faculty of Education. University of Ilorin, Ilorin

ARTICLE INFO

Article History

Accepted on July, 2025

Keywords: Differential Item Functioning, School Location, and Data Analysis Techniques
Introduction

ABSTRACT

Teachers in the classroom quickly set tests just to represent the syllabus or to achieve the objectives. Technical issues like difficulty, discrimination, options functionality and other things that make an item unique in a test are not given sufficient attention. Test experts look at those technical issues but not many of them look at differential item functioning. The concern of this study was to identify 2024 WASSCE Economics objective test items that are functioning differently among groups of examinees (Gender and school location). This study adopted a descriptive research design using a survey. The target population for the study was all rural and urban senior secondary three students offering Economics in the Kwara Central Senatorial District. In the study, 798 students were involved. A multistage sampling procedure comprising of proportionate sampling technique was employed. The Mantel-Haenszel approach was used for the data obtained. Findings show that some items functioned differently concerning gender, and some items functioned differently concerning school location. It was therefore recommended that (a) the Mantel-Haenszel statistical procedure should be used for differential item function (DIF) analysis of any achievement test to ensure test items that are free from bias towards a particular group. (b) Examining bodies like WAEC should ensure that items to be used to examine their candidates are free from any form of DIF by subjecting DIF analysis before administration.

Introduction

Teachers in the classroom quickly set tests just to represent the syllabus or to achieve their objectives. Technical issues like difficulty, discrimination, options functionality and other things that make an item unique in a test are not given sufficient attention. Test experts look at those technical issues, but not many of them look at differential item functioning. Odili (2014) defined differential item functioning as the phenomenon that occurs when test takers from different groups who have been matched on the same ability levels perform in test items in a different way. Bastug (2016), in his view, submits that DIF occurs when examinees who have the same ability level but from different groups have different probabilities of a correct response. From another angle, French et al (2014) explained the concept of DIF as a situation where examinees from two groups who have equal levels of the measured ability have different probabilities of endorsing the same item response.

In secondary schools, despite the direct relevance of Economics and laudable substantial investment in the study of the subject by the Nigerian government, students still perform differently in the subject when it comes to internal and external examinations. Buttressing further, Ikeh et al. (2020), for instance, noted that in 2018, WAEC Economics multiple-choice items were conducted, and students still performed differently with respect to school location. The analysis revealed that out of 50 items in the 2018 WAEC Economics questions, 31 items, which represent 62% of the test items, displayed significant location DIF at the 0.05 level of significance. Among the 31 items, 27 items representing 54% were identified to exhibit significant location DIF in favour of urban school students, while only 4 items representing 8% differentially functioned in favour of rural school students. The way items function differentially has been a source of worry to researchers, stakeholders in education, the general public and the students themselves. The reasons why Students perform differently in various subjects, including Economics, have been attributed to a number of factors like school location, gender and so on.

It is therefore necessary to investigate items that function differently in the 2024 WASSCE Economics objective test to ascertain the level of potentially biased items. Different research studies have also been conducted by researchers on DIF, for instance, Sa'ad (2020) in another study examined the Differential Item Functioning of the 2014 NECO English Language Examination in the North Senatorial District of Kano State, as three variables were considered, which include gender, school type and school location. In another study, Omorogiuwa and Iro-Aghedo (2016) researched on determination of Differential Item Functioning by gender in the

2015 NABTEB Mathematics Multiple-Choice Examinations in Nigeria. Also, Amuche and Fan (2014) investigated differential item functioning about school type and school location using the 2012 NECO Biology in Taraba State. The Mantel-Haenszel method was equally adopted to carry out their analysis. Ibrahim and Abba (2019) carried out DIF analysis on the English language using the 2010 NECO examination.

Based on the research activities conducted by various researchers as mentioned above, different subjects like English Language, Biology, and Mathematics were the subjects used, and none of these studies used the 2024 WASSCE Economics. Different locale was equally used for their studies, but not Kwara Central, Nigeria. Also, with respect to DIF size, most of the studies conducted did not actually tell whether items flagging DIF exhibited a negligible, moderate or large size of DIF to the best knowledge of the researcher. This creates room for a gap to be covered in this study.

Purpose of the Study

The concern of this study was to identify 2024 WASSCE Economics objective test items that are functioning differently among groups of examinees. Specifically, the study intends to assess:

- a. Economics objective test items of WASSCE 2024 function differently between male and female testees.
- b. Economics objective test items of WASSCE 2024 function differently between urban and rural testees.
- c. Economics objective test items of WASSCE 2024 function differently between private and public school testees.

Research Questions

The following research questions were derived to pilot this research study.

- a. Does differential item functioning occur in the 2024 WASSCE economics objective test items between male and female testees?
- b. Does differential item functioning occur in the 2024 WASSCE economics objective test items between urban and rural testees?

Methodology

This study adopted a descriptive research design. The population for this study was all rural and urban senior secondary school students offering Economics in the Kwara Central

senatorial district. The researcher selected a sample of seven hundred and ninety-eight (798) SS3 Economics students from selected senior secondary schools in the Kwara Central senatorial district. The overall total number of schools used for this research study was twenty-five (25) rural and thirty-two (32) urban senior secondary schools located in rural and urban areas.

A multistage sampling procedure comprising of proportionate sampling technique as well as purposive sampling was employed for effective selection of the sample in the study. The 2024 WASSCE Economics objective test was adopted as an instrument for the study. The items were developed, moderated, validated and used by the WAEC in 2024. The instrument, which consisted of two sections, was titled the Economics Achievement Test (EAT), and it has 50 items. Each item consists of a question and a list of possible answers, lettered A-D, of which only one option is the correct answer. Each item of the instrument was scored 1 for the correct option and 0 for the wrong option, with a maximum of 50 and a minimum of 0. The instrument is a standardised test which is presumed to be accurate, reliable and valid.

The data gathered from the administration of the instrument were scored dichotomously. Frequency and percentage were used to describe the demographic characteristics of the examinees, while the Mantel-Haenszel approach was used to answer the research questions raised.

Table 2: Demographic Characteristics of the Examinees

Categories	Characteristics		Frequency	Percentage
Gender	Male		356	44.60
	Female		442	55.40
	Total		798	100.00
School Location	Urban	524	65.70	
	Rural	274	34.30	
	Total	798	100.00	

The demographic characteristics reveal that three hundred and thirty-five (44.0%) of the students were males, while the rest (55.40%) were females. Five hundred and twenty-four (65.70%) of the students were from urban schools, while the rest two hundred and seventy-four (34.30%) were from rural schools. Four hundred and forty-eight (56.14%) of the students were from private senior secondary schools, while three hundred and fifty (43.86%) of the examinees labeled as were from public senior secondary schools.

Answering Research Questions

To answer the research questions raised in this study, the Mantel-Haenszel approach was used to assess DIF in the 2024 WASSCE Economics multiple-choice test items. Also, the DIF size was determined by the researcher based on Michaelides' (2008) classification. An item shows DIF if the p-value of MH is significant at 0.05. In her classification, items could be categorised into three groups for example Negligible DIF, which is known as A level, moderate DIF, known as B level, and large DIF items are known as C level. If the absolute value of MH D-DIF is less than or equal to 1, the item is negligible, which simply means no DIF in the test. If the Absolute value of MH D-DIF is between 1 and 1.5, then the item is moderate, and it can be readjusted. Lastly, if the absolute value of the MH D-DIF value is greater than 1.5, then that item is large and it can be discarded. It is imperative to equally add that an item favoured the reference group if the odds ratio is greater 1, while the focal group is favoured if the odds ratio is less than 1.

Research Question 1: Does differential item functioning occur in the 2024 WASSCE Economics objective test items between male and female testees

Table 3: Result of Mantel-Heanszel analysis between male and female testees

Items	MHX ²	MH P-Value	α MH Odds ratio	Log	MH D-DIF	DIF Level	Favouring Group
1	2.199	0.138	1.258	0.229	-0.538		
2	0.717	0.397	1.152	0.141	-0.331		
3	0.967	0.325	0.844	-0.170	0.399		
4	0.368	0.544	0.904	-0.101	0.237		
5	0.621	0.431	0.881	-0.127	0.298		
6	0.000	0.987	0.992	-0.008	0.019		
7	8.592	0.003	1.542	0.433	-1.018	B	FEMALE

8	0.122	0.727	1.079	0.076	-0.179		
9	0.307	0.580	1.095	0.091	-0.214		
10	0.069	0.792	0.952	-0.049	0.115		
11	0.660	0.416	1.143	0.134	-0.315		
12	0.100	0.752	0.946	-0.056	0.132		
13	0.089	0.765	1.055	0.053	-0.125		
14	0.031	0.583	1.111	0.105	-0.247		
15	5.591	0.018	1.488	0.398	-0.935	A	FEMALE
16	1.401	0.236	1.252	0.225	-0.523		
17	0.230	0.632	1.103	0.098	-0.230		
18	2.812	0.094	1.284	0.250	-0.588		
19	1.121	0.290	1.237	0.213	-0.501		
20	3.459	0.063	0.757	0.278	0.653		
21	0.456	0.500	0.894	-0.112	0.263		
22	1.853	0.173	0.811	-0.209	0.491		
23	1.879	0.170	1.257	0.228	0.536		
24	0.134	0.714	0.925	-0.078	0.183		
25	0.414	0.520	1.114	0.108	-0.254		
26	2.828	0.093	0.778	- 0.251	0.589		
27	0.000	0.983	0.993	-0.007	0.016		
28	0.004	0.949	1.003	0.003	0.007		
29	4.797	0.029	1.436	0.362	-0.851	A	FEMALE
30	0.053	0.819	1.045	0.044	-0.103		
31	0.065	0.799	1.053	0.051	-0.119		
32	0.095	0.758	1.055	0.054	-0.127		
33	0.864	0.353	1.186	0.170	-0.399		

34	0.430	0.512	1.130	0.122	-0.287		
35	0.138	0.710	1.071	0.069	-0.162		
36	0.995	0.318	1.187	0.172	-0.404		
37	2.118	0.146	0.801	-0.222	0.522		
38	0.495	0.482	0.878	-0.130	0.305		
39	0.017	0.896	0.971	-0.029	0.068		
40	5.320	0.021	0.704	-0.350	0.823	A	MALE
41	0.239	0.625	0.921	-0.082	0.193		
42	0.944	0.331	1.165	0.153	-0.359		
43	0.014	0.905	1.029	0.029	-0.086		
44	1.396	0.237	1.211	0.192	-0.451		
45	0.039	0.844	1.043	0.042	-0.099		
46	0.030	0.862	0.961	-0.040	0.094		
47	0.362	0.548	1.108	0.103	-0.240		
48	0.009	0.923	1.027	0.027	-0.063		
49	0.000	0.988	0.987	-0.013	-0.031		
50	0.130	0.718	1.064	0.062	-0.146		

The results of Mantel-Heanszel statistics in Table 3 above reveal that 4 (8%) items, i.e. items 7, 15, 29, and 40, functioned differently concerning gender (male and female) out of 50 objective test items in the 2024 WASSCE Economics test. This is because the Mantel-Hean's chi-square value was greater than the critical value of 3.84 at a 0.05 significance level. The rest 46 items (92%) did not function differently. Also, 3 items (75%), i.e. 7, 15 and 29 showed negative value in favour of the reference group (female) while one item (25%), i.e. 40 showed positive value in favour of the focal group (male). Educational Testing Services (ETS) categorisation scheme indicated item 7 to be a moderate level of DIF (level B), while the rest items 15, 29 and 40 were a small level of DIF, i.e. negligible DIF (level A). Items in negligible status can be retained in the test, while those that exhibited moderate status can be adjusted.

Research Question 2: Does differential item functioning occur in the 2024 WASSCE Economics objective test items between urban and rural testees?

Table 4: Result of Mantel-Hean's analysis between urban and rural testees

Items	MHX ²	MH	α MH	Log	MH D-DIF	DIF Level	Favouring
		P-Value		Odds ratio			Group
1	0.814	0.367	0.860	-0.151	0.355		
2	0.565	0.452	1.144	0.135	-0.317		
3	8.450	0.004	1.631	0.489	-1.149	B	URBAN
4	3.305	0.053	1.349	0.300	-0.705		
5	3.759	0.053	1.362	0.309	-0.726		
6	5.364	0.021	1.435	0.361	-0.848	A	URBAN
7	0.933	0.334	0.855	-0.157	0.369		
8	5.809	0.016	0.641	-0.444	1.043	B	RURAL
9	11.938	0.001	0.588	-0.531	1.248	B	RURAL
10	0.586	0.444	1.137	0.129	-0.303		
11	0.017	0.897	0.968	-0.033	0.078		
12	0.926	0.336	1.169	0.156	0.367		
13	0.200	0.655	1.082	0.078	0.731		
14	3.000	0.083	0.733	-0.311	0.731		
15	5.281	0.022	0.672	-0.398	0.935	A	RURAL
16	1.796	0.180	0.770	-0.261	0.613		
17	0.006	0.936	0.970	-0.031	0.073		
18	9.457	0.002	1.609	0.476	-1.119	B	URBAN
19	0.097	0.756	1.083	0.080	-0.188		
20	4.227	0.040	1.381	0.323	-0.759	A	URBAN

21	0.400	0.527	0.895	-0.111	0.261		
22	0.923	0.337	0.854	-0.157	0.369		
23	0.000	0.987	1.017	0.016	- 0.038		
24	1.709	0.191	1.296	0.259	-0.609		
25	4.018	0.045	1.382	0.323	-0.759	A	URBAN
26	0.404	0.525	1.112	0.106	-0.245		
27	7.706	0.006	1.532	0.427	-1.003	A	URBAN
28	1.404	0.236	0.805	-0.217	0.509		
29	1.370	0.242	0.813	-0.208	0.509		
30	0.074	0.785	1.055	0.053	-0.125		
31	1.181	0.277	0.821	-0.187	0.439		
32	0.000	0.937	1.023	0.023	-0.054		
33	2.173	0.140	1.327	0.283	-0.665		
34	5.011	0.025	0.674	-0.394	0.956	A	RURAL
35	5.071	0.024	0.691	-0.369	0.867	A	RURAL
36	5.207	0.022	0.679	-0.387	0.909	A	RURAL
37	2.914	0.088	0.764	-0.269	0.632		
38	0.793	0.373	0.847	-0.166	0.390		
39	2.717	0.099	1.295	0.258	-0.606		
40	0.202	0.653	1.084	0.081	-0.190		
41	0.040	0.842	0.959	-0.042	0.099		
42	0.046	0.830	0.956	-0.044	0.103		
43	2.923	0.087	0.759	-0.276	0.649		
44	0.624	0.430	0.871	-0.138	0.324		
45	6.200	0.013	0.668	-0.404	0.949	A	RURAL
46	0.006	0.938	0.999	-0.001	0.002		

47	1.451	0.228	0.817	-0.203	0.447
48	0.068	0.794	0.947	-0.055	0.129
49	1.605	0.205	0.814	-0.206	0.484
50	2.369	0.124	1.274	0.242	0.569

The results of Mantel-Heanszel statistics in table 4 above reveals that 13 (26%) items i.e. items 3, 6, 8, 9, 15, 18, 20, 25, 27, 34, 35, 36, and 45 functioned differentially concerning school location (urban and rural) out of 50 objective test items in the 2024 WAEC Economics test. This is because Mantel-Haenszel's chi-square value was greater than the critical value of 3.84 at a 0.05 significance level. The rest 37 items (74%) did not function differently. Also, 6 items (46.2%), i.e. 3, 6, 18, 20, 25 and 27 showed negative value in favour of reference group (urban) while 7 items (53.8%) i.e. 8, 9, 15, 34, 35, 36 and 45 showed positive value in favour of focal group (rural). Educational Testing Services (ETS) categorisation scheme indicated items 3, 8, 9, and 18 to be a moderate level of DIF (level B), while the rest items 6, 15, 20, 25, 27, 34, 35, 36 and 45 were a small level of DIF, i.e. negligible DIF (level A).

Summary of Findings

Findings obtained from the study were summarised as follows;

1. 4 (8%) items, i.e. items 7, 15, 29, and 40, functioned differently concerning gender (male and female) out of 50 objective test items in the 2024 WASSCE Economics test. Items 7, 15 and 29 showed negative values in favour of the reference group (female), while one item, 40, showed positive values in favour of the focal group (male).
2. 13 (26%) items i.e. items 3, 6, 8, 9, 15, 18, 20, 25, 27, 34, 35, 36, and 45 functioned differentially concerning school location (urban and rural) out of 50 objective test items in the 2024 WASSCE Economics test. 6 items i.e. 3, 6, 18, 20, 25 and 27 showed negative value in favour of reference group (urban) while 7 items i.e. 8, 9, 15, 34, 35, 36 and 45 showed positive value in favour of focal group (rural).

Discussion of the Findings

Differential Item Functioning analysis helps to know items that function differentially with respect to groups like gender, school type and school location. The result of the first research question revealed that 4 items functioned differentially, i.e. items 7, 15 and 29 showed negative

values in favour of the reference group (female) while one item, 40, showed a positive value in favour of the focal group (male). The outcome of this present study agreed with the findings of Jimoh et al. (2020), who in their study observed that 16 items exhibited DIF with respect to gender group using the Mantel-Heanszel Approach in the 2018 Joint Mock Mathematics Multiple-choice. The research findings also agree with the findings of Laro (2016), who noticed gender DIF in his study. The finding equally corroborated with Kalaycioglu's (2022) findings who observed that there were eleven (11) out of 234 items identified as showing B-level gender-related DIF.

3. Also, in the second research question, the researcher attempted to find out if any items were functioning differentially between urban and rural testees. After the analysis, it was discovered that 13 items functioned differentially i.e. 3, 6, 18, 20, 25 and 27 showed negative values in favour of reference group (urban) while 7 items i.e. 8, 9, 15, 34, 35, 36 and 45 showed positive value in favour of focal group (rural). The result of this study is in agreement with Amuche and Fan (2018), who noticed eight items that function differentially in NECO Biology. The finding is also in agreement with Ogbebor and Onuka's (2013) findings, noting school location DIF. The outcome of this study contradicts the findings of Alade et al. (2020), who observed that all fifty items in WASSCE 2018 exhibited school location DIF. Whereas only thirteen items functioned differently in this research study.

Conclusion

Following the results generated from this research study, it was concluded that testees with respect to gender exhibited the same ability level, as only three items showed a negligible level of DIF, while only one item showed a moderate level out of fifty items. The outcome of this study equally showed that thirteen items exhibited DIF, with four items displaying a moderate effect size, while nine items exhibited a small effect size. Also, concerning school type, twenty-six items showed differential item functioning using the Mantel-Haenszel approach, of which twelve of the items showed a large effect size, eight items showed in moderate level, while the remaining six items exhibited a small effect size. It was therefore concluded that the items that exhibited a moderate size level should be reframed, and the ones that exhibited a large level should be discarded for fairness to be ensured in the conduct of subsequent examinations.

Recommendations

The researcher provided the following recommendations following the findings derived from this study:

1. The Mantel-Haenszel approach should be used for DIF analysis of any achievement test for test items to be free from bias towards a particular group.
2. Test practitioners should endeavour to perform DIF analysis in a pilot study before administration of test(s) so that items that function differentially for different test-taking groups can be identified for possible replacement.
3. Examining bodies like WAEC should ensure that items to be used to examine their candidates are free from any form of DIF by subjecting DIF analysis before administration.

References

- Abedalaziz, N. (2010). Detecting gender related using logistic regression and Mantel-Haenszel approach *Procedia Social and Behavioural Sciences*, 7(C), 406-413
<https://doi.org/10.1016/j.sbspro.2010.10.055>
- Amuche, C. I., & Fan, A. F. (2014). An assessment of item bias using differential item Functioning techniques in NECO biology conducted examinations in Taraba State. *American International Journal of Research in Humanities, Arts and Social Sciences*, 6(1). 95-100.
- Bastung, O.Y.O. (2016). A comparison of four differential item functioning procedures in the presence of Multidimensionality. *Academic Journal of Educational Research and Reviews* 11(13), 1251-1261.
- French, B. F., Hand, B., Nam, I. H., & Vazquez, J. V. (2014). Psychological test and assessment modelling. 5(3), 275-286
- Ibrahim, S., & Abba, A. (2010). Identification of school type differential items functioning of the 2010 Senior School Certificate (NECO) English language examination among students of Dawakin Kudu Education Zone-Kano state, Nigeria. *International Journal of Advanced Educational Research*, 4(3), 31-35 www.educationjournal.org
- Ikeh, F. E., Ugwu, F. C., Mfon, T. E., Omosowon, V. O., Iketaku, R. I., Opa, F. A., Eze, B. A., Kalu, I. A., Ikwueze, C., & Ani, M. I (2020). Analysis of differential item functioning in Economics multiple-choice items administered by West African Examinations Council using logistic regression procedure. *Journal of Critical Reviews* 8(1), 980-986. <http://doi.org/10.31838/jcr.08.01.109>
- Jimoh, M. I., Daramola, D. S., Oladele, J. I., Sheu, A. L., & Ogunjimi, M. O (2019). Differential item functioning of Mathematics joint mock multiple-choice test items in Kwara

State. *Journal of Multidisciplinary Studies*, 9(1), 67-92. Doi: <http://dx.doi.org/10.7828/jmds.v9i1.1451>

Laaro, Y. B. (2016). Analysis of gender and location differential item functioning in 2013 Kwara State joint mock mathematics.

Michaelides, M. P. (2008). Practical assessment research and evaluation: An illustration of a Mantel-Haenszel procedure to flag misbehaving common items in test equating. *Practical Assessment, Research, and Evaluation. A Peer Electronic Journal*, 13 (7), 1-13. <https://doi.org/10.7275/n04d-8767>

Odili, J. N. (2014). Effect of Language manipulation on differential item functioning of test items in biology in a multicultural setting. *Journal of the Association for Educational Association in Africa*, 4 269-286.

Ogbebor, U., & Onuka, A. (2013). A differential item functioning of economics question papers of National Examinations Council in Delta State, Nigeria. *Nigerian Journal of Educational Research and Education*. 12(1), 45-60.

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

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

