

Detect Differential Item Function in Gender and School Ownership Using Mantel-Haenszel (Mh) in 2024 WASSCE Economics Objective Test Items

Zakari Sabi SIANO

*Department Arts and Sciences Education,
Kwara State University Malete, Nigeria.
Zakari.siano@kwasu.edu.ng*

Stephen O. BELLO

*Department of Social Science Education,
University of Ilorin Ilorin Nigeria.
bellosteve2k3@gmail.com*

&

Abdulrauf YUNUSA

*Department of Social Science Education,
University of Ilorin Ilorin Nigeria.
Abdulrauf4real@gmail.com*

Abstract

An item in a test is going to be fair to everybody if students of the same ability perform the same way in any given item. In order to determine whether an item is doing well or not in any test, it is necessary to establish that it is fair to everyone. Teachers in the classroom are more interested in developing items for their students and very few pay attention to whether the items are functioning differentially. Technical issues such as, item difficulty, options functionality, discrimination, and other aspects that make an item standardizes for assessment are not given enough attention. Assessment experts look at these technical issues but most people do not actually look into differential item functioning. The purpose of this study was to identify 2024 WASSCE Economics objective test items that may be functioning differentially among groups of examinees (Gender and school ownership) The researchers adopted descriptive research design of survey type, the target population was all public or private, senior secondary three students offering Economics in Kwara Central Senatorial district. A total of 799 students were involved. Multistage sampling procedure comprising of proportionate sampling technique was employed. Mantel-Haenszel approach was used for the data obtained. Finding shows that: some items functioned differentially with respect to gender and some items functioned differentially with respect to school ownership. It was therefore recommended that, (a) Mantel-Haenszel statistical procedure should be used for DIF analysis of any achievement test in order to realise items that are free from bias towards a particular group. (b)Examining bodies like WAEC should ensure that items to be used to examine candidates are free from any form of DIF by subjecting DIF analysis before administration.

Key Words: Item-discrimination, Item- difficulty, Options functionality and DIF

Introduction

Differential item function (DIF) analysis in any standardized test does not only ensure that item content is appropriate, but helps test experts as well as test developers to identify items that function differentially between groups of testees. Thus, differential item function is about the ability of examinees. So, for any item, every examinee having the same ability must answer the item the same way whether right or wrong irrespective of gender, school location, school type, ethnicity, language and so on. Different DIF can be used in different occasions depending on the researchers' focus.

Osadebe and Agbure (2020) submit that in any DIF analysis, two types of DIF exists which includes uniform and non-uniform. According to them, uniform DIF is said to happen when differences in correct probability are found across all ability levels for a particular item while non-uniform DIF on the other hand, occurs when there is interaction between the ability and group membership such that an item may seem difficult for those at the higher level in one group and after a point, it becomes more difficult for those at lower level in the other group. In any DIF analysis, there are various techniques or approaches that can be used to identify or analyze item that function differentially. Laro, (2016) posit that DIF techniques can be classified into parametric and nonparametric. Osadebe and Agbure (2020) submit that if the issues of DIF are not taken care of, it can pose a threat to the psychometric properties of the test, most especially in Economics among secondary school testees.

Amaechi and Onah (2019) noted that, Economics is one of the subjects offered in senior secondary schools in Nigeria and it is one of the social science subjects that heavily utilize statistical and mathematical models to analyze real-life problems. The issue of test items

functioning differentially in various subjects including Economics with respect to factors like gender, and school ownership type among others over the years, have been a source of worry and concern to researchers and relevant stakeholders in education industry, which calls for attention. Consequently, differential item function analysis is required in this subject to ensure an equitable assessment procedure, eliminate ambiguous or misleading items as well as to uphold test effectiveness and fairness. Corroborating the above statement, Ikeh et al. (2020) observed that in 2018 WAEC Economics multiple-choice items, students performed differentially with respect to school ownership.

The study reveals that out of 50 items in 2018 WAEC Economics questions, 31 items which represent 62% of the test items displayed DIF. Among the 31 items, 27 items representing 54% were identified to exhibit DIF in favour of private school students while only 4 items representing 8% differentially functioned in favour of public school students. Does this differential performance in this subject resulted as a reason of the mentioned factors?

The performances of testees vary across school ownership and gender, which may be attributed to various factors. In order to ensure fairness in assessment practices and procedures in the high stake tests, differential item functioning should be investigated so as to avoid bias. Many researchers have worked on similar aspect of test items, some of them are, Oladele, et al (2020) studied assessment of differential item functioning in public examinations mathematics constructed-response test and argued that differential item function exist in this test questions. The authors further researched in Paradigm shifts in computer adaptive testing in Nigeria in terms of simulated evidence still noted that care must be taken to avoid items functioning differently in a given test. Annan-Brew and Cobinnah, (2020) worked on gender-related

differential item functioning of 2015 WASSCE core mathematics results in Southern Ghana using logistic regression procedure.

Similarly, Sa'ad, (2020) in a related study examined differential item functioning of 2014 NECO English language examination in north senatorial district of Kano State, in which 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.

Akinboboye, et al (2014) noted that gender related differential item functioning could be detected using Mantel-Heanszel approach. In addition, Alade, (2020) assessed the differential items functioning of 2018 WASSCE Mathematics achievement test in Lagos Sate, Guilera, et al (2009) wrote on scientific production on the Mantel-Haenszel procedure as a way of detecting differential item functioning. Jimoh, et al (2019) researched in differential item functioning of Mathematics joint mock multiple-choice test items in Kwara State, and postulated that differential item function usually affect testees' performance in different ways. Osarumwense (2022) also worked on Non-parametric method of detecting differential item functioning in senior school certificate examination (SSCE) 2019 Economics multiple choice items in Kashere. It is therefore necessary to investigate items that function differentially in 2024 WASSCE Economics multiple choice test.

Purpose of the Study

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

- a. economics objective test items of WASSCE 2024 functioning differentially between male and female testees.
- b. economics objective test items of WASSCE 2024 functioning differentially between private and public schools testees.

Research Questions

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

- a. Does differential item function occur in 2024 WASSCE economics objective test items between male and female testees?
- b. Does differential item function occur in 2024 WASSCE economics objective test items between private and public schools testees?

Methodology

This study adopted descriptive research design. The population for this study was all private and public senior secondary school students offering Economics in Kwara Central senatorial district. The researchers selected a sample of seven hundred and ninety eight (798) SS3 Economics students from selected senior secondary schools in Kwara Central senatorial district. The overall total number of schools used for this research study was twenty five (29) public and thirty two (28) private senior secondary schools located in rural and urban areas.

Multistage sampling procedure comprising of proportionate sampling technique as well as purposive sampling was employed for effective selection of sample in the study. The 2024 WASSCE Economics objective test was adopted as instrument for the study. The items were developed, moderated, validated and used by the WAEC in 2024. The instrument which comprised of two sections was titled Economics Achievement Test (EAT) and it has 50 items. Each item consists of a question and a list of possible answer lettered A-D of which only one

option would be the correct answer. Each item of the instrument was scored 1 for correct option and 0 for wrong option with a maximum of 50 and a minimum of 0. The instrument is a standardized 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 was used to describe the demographic characteristics of the examinees while Mantel-Haenszel approach was used to answer the research questions raised.

Table 1: Demographic Characteristics of the Examinees

Categories	Characteristics	Frequency	Percentage
Gender	Male	356	44.60
	Female	442	55.40
	Total	798	100.00
School Type	Private	448	56.14
	Public	350	43.86
	Total	798	100.00

The demographic characteristics reveals 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 of Research Questions

To answer the research questions raised in this study, Mantel-Haenszel approach was used to assess DIF in 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 significance at 0.05. In the classification, items could be categorized into three groups for example Negligible DIF which is known as A level, moderate DIF known as B level while large DIF items is 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 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 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 odd ratio is greater 1 while the focal group is favoured if the odd ratio is less than 1.

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

Table 2: Result of Mantel-Haenszel analysis between male and female testees

Items	MHX ²	MH P-Value	α MH	Log Odds ratio	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-Haenszel statistics in table 2 above reveals that 4 (8%) items i.e. items 7, 15, 29, and 40 functioned differentially with respect to gender (male and female) out of 50 objective test items in the 2024 WASSCE Economics test. This is because Mantel-Haenszel chi-square value was greater than the critical value of 3.84 at 0.05 significance level. The rest 46 items (92%) did not function differentially. Also, 3 items (75%) i.e. 7, 15 and 29 showed negative value in favour of reference group (female) while one item (25%) i.e. 40 showed positive value in favour of focal group (male). Educational Testing Services (ETS) categorization scheme indicated item 7 to be moderate level of DIF (level B), while the rest items 15, 29 and 40 were small level of DIF i.e. negligible DIF (level A). Item in negligible status can be retained in the test while those that exhibited moderate status can be adjusted.

Research Question 3: Does differential item functioning occur in 2024 WASSCE Economics objective test items between private and public schools testees?

Table 3: Result of Mantel-Heanszel analysis between private and public schools testees

Items	MHX ²	MH	α MH	Log	MH D-DIF	DIF Level	Favouring
		P-Value		Odds ratio			Group
1	60.925	0.000	3.246	1.177	-2.766	C	PRIVATE
2	2.167	0.141	1.271	0.240	-0.564		
3	31.116	0.000	2.487	0.911	-2.141	C	PRIVATE
4	13.748	0.000	1.775	0.574	-1.349	B	PRIVATE
5	17.809	0.000	1.888	0.636	-1.492	B	PRIVATE
6	14.010	0.000	1.737	0.552	-1.297	B	PRIVATE
7	4.107	0.043	1.356	0.304	-0.714	A	PRIVATE
8	5.396	0.020	0.657	-0.420	0.987	A	PUBLIC
9	0.059	0.808	1.047	0.046	-0.108		
10	4.400	0.036	1.375	0.318	-0.747	A	PRIVATE
11	1.248	0.264	0.835	-0.180	0.423		
12	20.412	0.000	1.951	0.668	-1.569	C	PRIVATE
13	18.081	0.000	1.805	0.623	-1.464	B	PRIVATE
14	4.205	0.040	1.438	0.363	-0.853	A	PRIVATE
15	11.809	0.001	0.564	-0.573	1.347	B	PUBLIC
16	12.769	0.000	1.976	0.681	-1.600	C	PRIVATE
17	0.113	0.737	1.076	0.074	-0.174		
18	64.918	0.000	3.336	1.205	-2.832	C	PRIVATE
19	0.941	0.332	1.221	0.200	-0.470		
20	18.948	0.000	1.906	0.642	-1.509	C	PRIVATE
21	7.605	0.000	1.537	0.430	-1.011	C	PRIVATE
22	1.366	0.242	0.835	-0.181	0.425		
23	2.202	0.138	0.782	-0.246	0.578		
24	0.020	0.888	1.040	0.039	-0.092		
25	33.934	0.000	2.439	0.892	-2.096	C	PRIVATE
26	22.186	0.000	2.000	0.693	-1.629	C	PRIVATE
27	30.578	0.000	2.248	0.810	-1.903	C	PRIVATE

28	19.788	0.000	2.196	0.786	-1.847	C	PRIVATE
29	0.139	0.710	0.930	-0.072	0.169		
30	11.120	0.001	1.649	0.500	-1.175	B	PRIVATE
31	0.006	0.936	1.025	0.024	-0.056		
32	4.463	0.035	1.367	0.312	-0.773	A	PRIVATE
33	1.425	0.233	1.243	0.218	-0.512		
34	1.199	0.274	1.218	0.198	-0.465		
35	3.247	0.072	1.339	0.292	-0.686		
36	5.848	0.006	1.502	0.402	-0.945	A	PRIVATE
37	12.760	0.000	1.711	0.537	-1.262	B	PRIVATE
38	0.483	0.487	0.880	-0.537	0.128		
39	3.789	0.052	1.337	0.290	-0.682		
40	22.551	0.000	2.060	0.723	-1.699	C	PRIVATE
41	2.554	0.110	1.277	0.245	-0.576		
42	0.196	0.658	1.079	0.076	-0.177		
43	0.020	0.889	0.969	-0.032	0.075		
44	3.177	0.075	0.753	-0.283	0.665		
45	1.412	0.235	1.215	0.194	0.446		
46	0.010	0.922	0.973	-0.028	1.048		
47	1.919	0.166	1.251	0.224	-0.526		
48	0.293	0.588	0.910	-0.095	0.223		
49	0.078	0.781	1.054	-0.052	0.122		
50	10.411	0.001	1.607	0.474	-1.114	B	PRIVATE

The results of Mantel-Haenszel statistics in table above reveals that 26 (52%) items i.e. items 1, 3, 4, 5, 6, 7, 8, 10, 12, 13, 14, 15, 16, 18, 20, 21, 25, 26, 27, 28, 30, 32, 36, 37, 40, and 50 functioned differentially with respect to school type (private and public) out of 50 objective test items in the 2022 WASSCE Economics test. This is because Mantel-Haenszel chi-square value was greater than the critical value of 3.84 at 0.05 significance level. The rest 24 items

(48%) did not function differentially. Also, 24 items (92.3%) i.e. 1, 3, 4, 5, 6, 7, 10, 12,13, 14, 16, 18, 20, 21, 25, 26, 27, 28, 30, 32, 36, 37, 40 and 50 showed negative value in favour of reference group (private) while 2 items (7.7%) i.e. 8 and 15 showed positive value in favour of focal group (public). Educational Testing Services (ETS) categorization scheme indicated item 1, 3, 12, 16, 18, 20, 21, 25, 26, 27, 28, and 40 be large level of DIF (level C), item4, 5, 6, 13, 15, 30, 37, and 50 showed moderate level of DIF (level B) while the remaining items i.e. 7, 8, 10, 14, 32 and 36 exhibited negligible level of DIF (level A).

Conclusion

Based on the results generated in this study, it was concluded that testees with respect to gender exhibited the same ability level as only three items showed negligible level of DIF while only one item showed moderate level out of fifty items. The outcome of this study equally showed that thirteen items exhibited DIF with four items displaying moderate effect size while nine items exhibited small effect size. Also, with respect to school type, twenty six items showed differential item functioning using Mantel-Haenszel approach, of which twelve of the items showed large effect size, eight items in moderate level while the remaining six items exhibited small effect size. It was therefore concluded that the items that exhibited moderate size level be reframed and the ones that exhibited large level be discarded in order 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. Mantel-Haenszel approach should be used for DIF analysis of any achievement test in order for test items to be free from bias towards a particular group.
2. 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.
3. Capacity building programmes should be organized for test item developers, teachers and examiners on how to effectively construct a good, quality and fair item.

References

- Aituariagbon, K. E., & Osarumwense, H. J. (2022). Non-parametric method of detecting differential item functioning in senior school certificate examination (SSCE) 2019 Economics Multiple Choice Items.) *Kashere Journal of Education*, 3(1), 146-158
- Akinboboye, J. T., Akande, J. A., Jimoh, M. I., & Adewumi, D. A. (2014). Gender related differential item functioning using Mantel-Haenszel approach. *Nigeria Journal of Educational Productivity and Contemporary Issues*. 7(2), 287-291
- Alade, O. M., Aletan, S., & Sokenu, B. S. (2020). Assessing the differential functioning of 2018 WASSCE Mathematics achievement test in Lagos State, Nigeria *African Journal of Behavioral and Scale Development Research*2(2), 8-20.
- Amaechi, C. E., & Onah, F. E. (2019). Detection of uniform and non-uniform gender differential item functioning in economics multiple-choice standardized test in Nigeria. *Journal of Nigerian Academy of Education*. 5(2), 224-233.
- Annan-Brew, R., & Cobinnah, A. (2020). Gender-related differential item functioning of 2015 WASSCE core mathematics results in Southern Ghana using logistic regression procedure. *European Scientific Journal, ESJ*, 16(16), 188-198.
- 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.
- Guilera, G., Gomez, J., & Hidalgo, M. D. (2009). Scientific production on the Mantel-Haenszel procedure as a way of detecting differential item functioning. *Psicothema*, 21(3), 492-498.
- 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/1031838/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>
- Laro, Y. B. (2016). *Analysis of gender and location differential item functioning in 2013 Kwara State joint mock mathematics*. An M. Ed. Thesis presented to the department of Social Sciences Education, Faculty of Education, University of Ilorin, Ilorin, Nigeria.
- Mantel, N., & Haenszel, W. (1959). Statistical aspect of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22, 719-748.
- Oladele, B. K., Adegoke, B. A., & LongJohn, D. A. (2020). Assessment of differential item functioning in public examinations mathematics constructed-response test. *Journal of Positive Psychology and Counseling*, 4, 221-223. <https://www.ppacjournals.org>
- Oladele, J. I., Ayanwale, M. A., & Owolabi, H. O. (2020). Paradigm shifts in computer adaptive testing in Nigeria in terms of simulated evidences. *Journal of Social Science* 63(1-3): 9-20 doi:10.31901/24566608.2020/63.1-3.2264 <https://dx.doi.org/10.7828/ijher.v11i1.890>
- Osadebe, P., & Agbure, B. (2020). Assessment of differential item functioning in social sciences multiple choice questions in basic education certificate examination. *European Journal of Educational Studies*, 4(9), 236-257. <https://doi.org/1046827/EJES.v010.1783>
- 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), 17-30