POTENCY OF TWO STATISTICAL APPROACHES IN DETECTION OF DIFFERENTIAL ITEM FUNCTIONING OF STANDARDIZED ECONOMICS TEST IN NIGERIA

[1]AMAECHI, C. E., [2]ONAH, F. E. & [3]NOSIKE, M. C.

[1]Federal College of Education,
Eha-amufu, Enugu State
chideraemma19@gmail.com

[2]Department of Life Science Education
(Measurement and Evaluation),
Faculty Of Education,
Imo State University, Owerri
fredonah2@yahoo.com

[3]Department of Physical Science Education
(Measurement and Evaluation),
Faculty of Education,
Imo State University, Owerri.

Abstract

The study ascertained the potency of two statistical approaches (Binary Logistic Regression and Mantel-Haenszel Chi-Square) in detection of differential item functioning of standardized Economics test in Nigeria. Two research questions were asked and two hypotheses were formulated to guide the study. The design of this study is an inferential survey. The population of the study was 4,434,979 secondary school students and the sample comprised of seven hundred and ninety eight (798) students. The sample comprised of seven hundred and ninety eight (798) students involving purposive, stratified and simple random sampling techniques. The instrument for data collection is a 50 item WAEC General Economics Paper I Multiple Choice Test. Two statistics were used to test the hypotheses, binary logistic regression and Mantel-Haenszel Chi-Square statistics at 0.05 level of significance. Result of the analysis revealed among others that; Binary Logistic Regression and Mantel-Haenszel Chi-Square statistics have equal potency in the detection of gender DIF with 13 items; but 8 items were consistent with the two statistics as having gender DIF while 5 items were not consistent. It was concluded that Binary Logistic Regression Statistics has a higher potency in detecting location DIF with 9 items than Mantel-Haenszel Chi-Square Statistics that detected 8 items; but 6 items were consistent with the two statistics as having location DIF while the other items were not consistent. Based on the findings, the researcher recommended that; when detecting location DIF, the use of Binary Logistic Regression Statistics should be encouraged since it detected more DIF in this regard.

Keywords: Economics, DIF, Bias, Binary Logistic Regression and Mantel-Haenszel Chi-Square statistics

Introduction

Economics is one of the subjects offered in senior secondary schools in Nigeria. It is one of the social science subjects that heavily utilize statistical and mathematical models to analyze real-life problems. It is a social science that studies human behaviour in an effort to allocate scarce resources efficiently and effectively in order to minimize cost. It is a subject concerned with the efficient utilization or management of limited or scarce resources for maximum satisfaction of human needs (Amaechi, 2018; Onah, Amaechi & Nosike, 2017). As such, to achieve this at the Senior Secondary School level, the assessment of students' academic achievement in Economics demands attention, thus, it ought to be as fair as possible with minimal error.

Differential Item Functioning (DIF) as defined by Angoff cited in Moyo and Nenty (2017) referred to the differences in the statistical properties of an item between groups (Male/Female examines) of equal ability. It is intended to be invariant with respect to irrelevant aspects of the test-takers, such as gender, location, school type, school ownership, ethnicity and socio-economic status. It is also expected to be altered by interventions targeted at those items. For instance, the use of calculators in arithmetic tests or the use of assistive device on mobility tests. As such, differential item functioning (DIF) occurs when examinees from different groups (Male/Female examinees) show differing probabilities of success on the item after being matched on the underlying ability that the item is intended to measure (Walker, 2011). Lending support to this, Ajeigbe & Afolabi (2014) contended that Differential item functioning (DIF) referred to a difference in the way a test item functions for comparable groups of test takers. Formally defined, an item displays DIF if subjects of equal proficiency, or equal ability level on the construct intended to be measured by a test but from separate subgroups of the population differ in their expected score on this item. It is noteworthy that when assessing an item for DIF, the groups must be matched on the measured attribute; otherwise this may result in inaccurate detection of DIF (Amaechi, 2019).

Furthermore, DIF examine difference in the conditional probability of the responses between two groups; therefore, the groups can be designated “reference” and “focal” groups (Amajuoyi, 2015). Camili (2006) contended that while the designation does not matter, a typical practice in the literature is to designate the reference group as the group who are suspected to have an advantage while the focal group referred to the group anticipated to be disadvantaged by the test. The absence of DIF is determined by the fact that the conditional probability distribution of the response is not dependent on group membership. This is applicable to dichotomous scored tests. Dichotomous scores are scores from a test which has two categories of outcome which may be pass and fail, yes and no and true and false, usually coded as 1 and 0 respectively (Amaechi, 2019).

For a dichotomously scored test, the probability of correctly responding to an item is the same for members of either group once matched on the same attribute. This indicates that there is no DIF because members of the reference and focal groups with the same underlying ability or attribute have the same probability of responding correctly (Amaechi, 2019). Therefore, there is no disadvantage for one group over the other. Put differently, when the focal and reference groups are at the same location on theta (θ) and there is a different probability of getting a correct response or endorsing an item. The group with higher conditional probability of correctly responding to an item is the group advantaged by the test item. This suggests that the test item displays DIF and functions differently for the groups and therefore, may be potentially biased (Walker, 2011; Osterlind & Everson, 2009).

Failure to understand and account for differences on test scores may lead to misinterpretation and consequently, actions and decisions that are invalid. DIF explicitly involves conditioning on θ. This is an indication that an examinee's score is conditional on grouping such that having information about group membership changes the probability of a correct response. Therefore, if the group differs on θ and performance depends on θ, item bias would be suggested even on absence of DIF (Walker, 2011). For this reason, it is generally agreed in literature that differences on response conditional on group membership alone is inadequate for establishing bias but performance dependent on the θ is best used for item bias decision (Amajuoyi, 2015). This suggests that DIF investigates the items in a test, one at a time, for signs of interactions with sample characteristics.

There are several methods to detect if items have DIF effects. Some of the methods include but not limited to: chi-squares techniques: Mantel-Haenszel (M-H)/Generalized Mantel-Haenszel (GMH) methods, Scheuneman Modified Chi-square ($SS\chi^2$),Cochran's Chi-square Test Method ($CT\chi^2$) and Lord's Chi-Square (χ^2) Test. Others are Standard Mean Difference (SMD) techniques, Standardization, Simultaneous Item Bias Test, Transformation Item Difficulty, Distractor Analysis Method, Expert Judges, Item Discrimination Index, Logistic Regression, Factor Analysis, Rank Order, Item Characteristic Curve (ICC), Item Response Theory-Likelihood Ratio, Comparison Method, Log Linear Model, Transformed Item Difficulty, Parameter Index, Analysis of Variance Methods and Methods of Comparing Plots of Transformed Item Difficulties, Correlations and Methods Based on Experimental Manipulations, Reference and Focal Groups Method. For this study, Binary Logistic Regression and Mantel-Haenszel Chi-Square were used. Binary Logistic Regression, detects both uniform and non uniform items that function differentially while Mantel-Haenszel Chi-Square detects only uniform DIF items. In the context of Nigerian measurement, researchers seem to use the above statistical approaches, hence, the use of it in this study.

These groups are stratified into matching ability levels and their relative performance on each item is quantified. The ability levels are usually determined by the total scores on the test. In this way, the DIF analysis for one item is as independent as possible from the DIF analysis of the other items. The presence of DIF may have serious consequences for the interpretation of test scores for both groups and individuals. According to Zumbo (2012), test items are considered biased because they contain sources of difficulty that are not relevant to the construct being measured and these extraneous sources affect test-takers' performance. Those items will bring about differential functions across students' with gender and location differences.

Gender is a set of characteristics distinguishing between males and females, particularly, in the case of man and woman which, depending on the context, may vary from sex to social role and to gender identity (Bland, 2013). Nevertheless, Chang (2013) reported that although there is a decrease in the gap in gender difference in students' performance in subjects, female representation in subjects involving calculation like sciences is still low in comparison with their male counterparts.

In a similar occasion, the school location (urban/rural) which a child finds himself/herself, goes a long way to determine one's academic achievement in life. Location of schools could also be a factor that affects academic performance of students in schools generally. Kpolovie, Joe & Okoto (2014) stated that school's location means urban and rural schools. Location is a particular place in relation to other areas. Osuafor & Okonkwo (2013) indicated that schools in urban areas have electricity, water supply, more teachers more learning facilities and infrastructure. To support this Juweto (2015) stated that urban areas are those with high population density, high variety and beauty while rural areas are those with low population, subsistence mode of life, monotonous and burden. Students attending rural schools face challenges of higher poverty than those attending urban schools. However, Osuafor and Okonkwo (2013) found that urban students performed better than rural students in all forms of achievement test used. They further reported that students in urban and rural location performed in a similar manner.

Researchers such as Pedrajita (2015) and Salubayba (2013) revealed that Binary logistic regression can help to detect both uniform and non-uniform DIF while Mantel-Haenszel Chi-Square Statistics can only be used to detect uniform DIF. Pedrajita (2015) worked on the use of contingency table approaches in differential item functioning analysis: a comparison. The results of the approaches were compared. The findings revealed the presence of items indicating gender-based DIF, and thus concluded that there was a high degree of correspondence between the Logistic Regression and the Mantel-Haenszel Statistic in identifying potentially biased test items. Similarly, Salubayba (2013) revealed that Mantel-Haenszel (MH) detected fewer DIF items that resulted to more DIF-free items, while IRT-1PL detected more DIF items that produced brief test instrument. Ibrahim (2017) results showed that a non-significant relationship existed between the proportions of test items that functioned differentially in the dichotomous test items when the different methods were used. The study concluded that the statistical approaches were effective in detecting DIF across dichotomous test items but complement one other in their ability to detect DIF in dichotomous test items. Looking at past studies, to the best of the researcher's knowledge, no study was particular on the presents study. The problem of this study posed as a question: what is the potency of two statistical approaches (Binary Logistic Regression and Mantel-Haenszel Chi-Square) in detection of differential item functioning of standardized Economics test in Nigeria? The answers to these questions are what the present study provided.

The general purpose of this study was to ascertain the potency of two statistical approaches (Binary Logistic Regression and Mantel-Haenszel Chi-Square) in detection of differential item functioning of standardized Economics test in Nigeria. Specifically, the study ascertained:

1. The DIF statistics (Binary Logistic Regression and Mantel-Haenszel Chi-Square) that detected more items in WAEC with respect to gender, and
2. The DIF statistics (Binary Logistic Regression and Mantel-Haenszel Chi-Square) that detected more items in WAEC with respect to school location.

Research Questions

The following research questions were posed and they guided the study:

a). Which of the DIF statistics (Binary Logistic Regression and Mantel-Haenszel Chi-Square)detected more items in WAEC Economics test with respect to gender?

b). Which of the DIF statistics (Binary Logistic Regression and Mantel-Haenszel Chi-Square) detected more items in WAEC Economics test with respect to school location?

Research Hypotheses

The following null hypotheses were formulated and tested at 0.05 level of significance:

Ho_1: The number of items that were detected with gender DIF using Binary Logistic Regression and Mantel-Haenszel Chi-Square are not significant.

Ho_2: The number of items that were detected with school location DIF using Binary Logistic Regression and Mantel-Haenszel Chi-Square are not significant.

Methodology

The design of this study is a survey which involved the inferential method. The population of the study was 4,434,979 secondary school students in the ***11,875***public secondary schools in the 36 states including Abuja the Federal Capital Territory (FCT) of Nigeria. The sample comprised of seven hundred and ninety eight (798) students involving purposive, stratified and simple random sampling techniques. Purposively, the researcher chose those students matured for WAEC, while stratified random sampling was used to draw the sample with consideration to the bio-data of the students. The instrument for data collection is 50 item WAEC General Economics Paper I Multiple Choice Test 2015. Two statistics were used to test the hypotheses, binary logistic regression and Mantel-Haenszel Chi-Square statistics at 0.05 level of significance. Any item whose Wald test/Chi-Square (χ^2) significance level is below 0.05, was termed as significantly DIF item, but any item with Wald test/Chi-Square (χ^2) significance level above 0.05, was termed non- significant DIF item.

Results

In this section, the researchers presented data analysis and results according to the earlier stated hypotheses.

Hypothesis One: *The number of items that were detected with gender DIF using Binary Logistic Regression and Mantel-Haenszel Chi-Square are not significant.*

Table 1a: Gender DIF Using Binary Logistic Regression Statistics

Variables in the Equation								
Items	**B**	**S.E.**	**Wald**	**df**	**Sig.**	**Exp(B)**	**95% C.I.for EXP(B)**	
							Lower	**Upper**
1	-.027	.200	.018	1	.892	.973	.658	1.439
2	-.158	.200	.621	1	.431	.854	.577	1.264
3	-.199	.193	1.056	1	.304	.820	.561	1.198
4	.107	.205	.274	1	.600	1.113	.745	1.662
5	.080	.206	.149	1	.699	1.083	.723	1.621
6	-.106	.195	.294	1	.588	.900	.614	1.319
7	.002	.202	.000	1	.993	1.002	.674	1.489
8	.024	.200	.014	1	.904	1.024	.692	1.517
9	.511	.215	5.625	1	**.018**	1.667	1.093	2.542
10	-.129	.200	.418	1	.518	.879	.594	1.300
11	.322	.199	2.618	1	.106	1.381	.934	2.040
12	-.153	.203	.570	1	.450	.858	.576	1.277
13	-1.596	.196	66.391	1	**.000**	.203	.138	.298
14	.336	.200	2.835	1	.092	1.400	.946	2.070
15	-1.294	.192	45.193	1	**.000**	.274	.188	.400
16	-.113	.206	.300	1	.584	.893	.597	1.337
17	.002	.196	.000	1	.993	1.002	.682	1.471
18	.302	.194	2.437	1	.118	1.353	.926	1.978
19	-.093	.196	.225	1	.635	.911	.621	1.337
20	-.898	.182	24.385	1	**.000**	.407	.285	.582
21	-.058	.198	.087	1	.768	.943	.640	1.391
22	.419	.176	5.653	1	**.017**	1.521	1.076	2.148
23	.170	.206	.683	1	.409	1.185	.792	1.774
24	1.473	.185	63.622	1	**.000**	4.362	3.037	6.265
25	-.067	.193	.120	1	.729	.935	.640	1.366
26	.082	.208	.154	1	.695	1.085	.722	1.632
27	.064	.179	.127	1	.722	1.066	.751	1.513
28	.151	.195	.594	1	.441	1.162	.793	1.705
29	-.032	.198	.027	1	.871	.968	.657	1.427
30	.476	.214	4.972	1	**.026**	1.610	1.059	2.447
31	-.166	.206	.651	1	.420	.847	.566	1.268
32	.012	.203	.003	1	.953	1.012	.680	1.506
33	.030	.198	.023	1	.879	1.031	.699	1.519
34	.380	.216	3.092	1	.079	1.462	.957	2.233
35	-.218	.198	1.209	1	.272	.804	.546	1.186
36	-.087	.195	.198	1	.656	.917	.626	1.343
37	.639	.208	9.449	1	**.002**	1.894	1.261	2.847
38	.205	.218	.887	1	.346	1.228	.801	1.882
39	.596	.206	8.364	1	**.004**	1.816	1.212	2.720
40	-.128	.200	.409	1	.522	.880	.595	1.302
41	.208	.213	.954	1	.329	1.231	.811	1.867
42	-.764	.175	19.149	1	**.000**	.466	.331	.656
43	-.152	.193	.622	1	.430	.859	.589	1.253
44	.554	.171	10.493	1	**.001**	1.740	1.245	2.433
45	-.174	.196	.790	1	.374	.840	.572	1.233
46	-1.038	.177	34.558	1	**.000**	.354	.250	.500
47	.745	.208	12.856	1	**.000**	2.106	1.402	3.164
48	-.124	.201	.379	1	.538	.884	.596	1.310
49	.173	.196	.773	1	.379	1.188	.809	1.746
50	.214	.196	1.193	1	.275	1.239	.843	1.820

In Table 1(a), the number of WAEC items in Economics that function differentially with respect to gender were detected using Binary Logistic Regression Statistics. With Wald test significance level below 0.05, the result indicated that, 13 items that is, item numbers 9, 13, 15, 20, 22, 24, 30, 37, 39, 42, 44, 46 and 47 in the Economics test were identified as significantly exhibiting differential item functioning between male and female students at .05 level of significance while the remaining 37 items did not significantly exhibit differential item functioning between male and female students. This reveals that Binary Logistic Regression Statistics detected and flagged 13 items as having DIF with respect to gender. Having done this, researcher proceeded to test the hypothesis using Mantel-Haenszel Chi-Square Statistics and the output is illustrated in table 1b.

Table 1b: Gender DIF Using Mantel-Haenszel Chi-Square Statistics

Tests of Conditional Independence			
Items	**Mantel-Haenszel Chi-Squared**	**df**	**Asymp. Sig. (2-sided)**
1	.025	1	.876
2	1.114	1	.291
3	.046	1	.829
4	.020	1	.889
5	1.275	1	.259
6	.996	1	.318
7	.648	1	.421
8	.777	1	.378
9	.009	1	.924
10	1.490	1	.222
11	.046	1	.831
12	.954	1	.329
13	91.421	1	**.000**
14	.794	1	.373
15	72.677	1	**.000**
16	4.150	1	**.042**
17	.150	1	.698
18	3.938	1	**.047**
19	.465	1	.495
20	31.418	1	**.000**
21	.980	1	.322
22	.169	1	.681
23	.170	1	.680
24	53.157	1	**.000**
25	.009	1	.923
26	.044	1	.834
27	6.361	1	**.012**
28	.006	1	.940
29	.057	1	.812
30	.000	1	.992
31	2.445	1	.118
32	.000	1	.983
33	18.353	1	**.000**
34	.810	1	.368
35	44.955	1	**.000**
36	.897	1	.344
37	2.084	1	.149
38	2.582	1	.108
39	2.003	1	.157
40	.945	1	.331
41	1.201	1	.273
42	32.226	1	**.000**
43	.533	1	.466
44	10.026	1	**.002**
45	.169	1	.681
46	32.191	1	**.000**
47	5.864	1	**.015**
48	.046	1	.829
49	.493	1	.483
50	.001	1	.976

In Table 1(b), the number of WAEC items in Economics that function differentially with respect to gender were detected using Mantel-Haenszel Chi-Square Statistics. With Chi-Square Test of Independence significance level below 0.05, the result also indicated that, 13 items that is, item numbers 13, 15, 16, 18, 20, 24, 27, 33, 35, 42, 44, 46 and 47 in the Economics test were identified as significantly exhibiting differential item functioning between male and female students at .05 level of significance while the remaining 37 items did not significantly exhibit differential item functioning between male and female students. This reveals that Mantel-Haenszel Chi-Square Statistics detected and flagged 13 items as having DIF with respect to gender.

Comparing the two statistics for detecting DIF, it was observed that both statistics (Binary Logistic Regression Statistics & Mantel-Haenszel Chi-Square Statistics) detected the equal number of items as having gender DIF, but the difference is that 8 items were consistent with the two statistics as having gender DIF while 5 items were not consistent. The consistent items are 13, 15, 20, 24, 42, 44, 46 and 47. Hence, both statistics contributed equally in the detection of DIF with respect to gender in 2015 WAEC test.

Hypothesis Two: *The number of items that were detected with school location DIF using Binary Logistic Regression and Mantel-Haenszel Chi-Square are not significant.*

Table 2a: Location DIF Using Binary Logistic Regression Statistics

Variables in the Equation								
Items	**B**	**S.E.**	**Wald**	**Df**	**Sig.**	**Exp(B)**	**95% C.I.for EXP(B)**	
							Lower	**Upper**
1	.230	.234	.967	1	.326	1.259	.795	1.993
2	.140	.238	.348	1	.555	1.151	.722	1.834
3	.407	.236	2.973	1	.085	1.503	.946	2.387
4	-.250	.247	1.028	1	.311	.778	.480	1.263
5	.273	.239	1.312	1	.252	1.314	.823	2.098
6	-.032	.224	.020	1	.886	.968	.624	1.503
7	-.057	.229	.063	1	.802	.944	.603	1.478
8	-.057	.232	.060	1	.806	.945	.600	1.488
9	.093	.252	.135	1	.713	1.097	.669	1.798
10	-.074	.226	.108	1	.742	.928	.596	1.446
11	.113	.234	.234	1	.628	1.120	.708	1.773
12	.480	.243	3.895	1	**.048**	1.616	1.003	2.604
13	-.235	.213	1.212	1	.271	.791	.520	1.201
14	.324	.234	1.922	1	.166	1.383	.875	2.186
15	-.761	.209	13.259	1	**.000**	.467	.310	.704
16	.181	.237	.584	1	.445	1.199	.753	1.908
17	.058	.230	.064	1	.800	1.060	.675	1.664
18	-.197	.217	.827	1	.363	.821	.537	1.256
19	.226	.234	.931	1	.335	1.254	.792	1.985
20	-.570	.202	7.925	1	**.005**	.566	.380	.841
21	.224	.238	.885	1	.347	1.251	.784	1.996
22	-.421	.204	4.268	1	**.039**	.656	.440	.979
23	-.095	.249	.147	1	.701	.909	.558	1.480
24	-.967	.210	21.308	1	**.000**	.380	.252	.573
25	.242	.235	1.062	1	.303	1.274	.804	2.020
26	-.170	.250	.463	1	.496	.844	.517	1.377
27	.361	.205	3.086	1	.079	1.435	.959	2.146
28	-.092	.230	.160	1	.690	.912	.581	1.432
29	-.155	.230	.454	1	.501	.856	.545	1.345
30	-.120	.250	.231	1	.631	.887	.543	1.447
31	-.020	.248	.006	1	.937	.981	.603	1.593
32	.133	.241	.306	1	.580	1.143	.712	1.833
33	-.012	.234	.003	1	.958	.988	.624	1.563
34	.061	.253	.058	1	.810	1.063	.647	1.746
35	.458	.231	3.931	1	**.047**	1.581	1.005	2.487
36	-.016	.229	.005	1	.943	.984	.628	1.540
37	.064	.243	.070	1	.792	1.066	.662	1.718
38	-.057	.257	.050	1	.824	.944	.570	1.563
39	.147	.243	.366	1	.545	1.159	.719	1.867
40	.108	.238	.206	1	.650	1.114	.699	1.776
41	.008	.251	.001	1	.975	1.008	.616	1.650
42	1.171	.216	29.349	1	**.000**	3.226	2.112	4.929
43	-.001	.227	.000	1	.998	.999	.641	1.558
44	-.702	.203	11.937	1	**.001**	.496	.333	.738
45	.333	.236	1.994	1	.158	1.395	.879	2.215
46	-1.108	.218	25.777	1	**.000**	.330	.215	.506
47	-.167	.243	.472	1	.492	.846	.525	1.363
48	.392	.242	2.627	1	.105	1.480	.921	2.378
49	.325	.232	1.958	1	.162	1.384	.878	2.183
50	.036	.233	.025	1	.876	1.037	.657	1.637

In Table 2(a), the number of WAEC items in Economics that function differentially with respect to location were detected using Binary Logistic Regression Statistics. With Wald test significance level below 0.05, the result indicated that, 9 items that is, item numbers 12, 15, 20, 22, 24, 35, 42 and 46 in the Economics test were identified as significantly exhibiting differential item functioning between students from urban and rural schools at .05 level of significance while the remaining 41 items did not significantly exhibit differential item functioning between students from urban and rural schools. This reveals that Binary Logistic Regression Statistics detected and flagged 9 items as having DIF with respect to location.

Table 2b: Location DIF Using Mantel-Haenszel Chi-Square Statistics

Tests of Conditional Independence			
Items	**Mantel-Haenszel Chi-Squared**	**Df**	**Asymp. Sig. (2-sided)**
1	.082	1	.775
2	.026	1	.872
3	2.494	1	.114
4	1.083	1	.298
5	.057	1	.811
6	.118	1	.732
7	2.536	1	.111
8	.770	1	.380
9	2.045	1	.153
10	.999	1	.318
11	.742	1	.389
12	.003	1	.956
13	**20.780**	**1**	**.000**
14	.102	1	.750
15	**51.021**	**1**	**.000**
16	.002	1	.968
17	.304	1	.581
18	.400	1	.527
19	.015	1	.902
20	**44.587**	**1**	**.000**
21	.004	1	.948
22	.556	1	.456
23	.617	1	.432
24	**27.556**	**1**	**.000**
25	.586	1	.444
26	1.927	1	.165
27	**20.163**	**1**	**.000**
28	.074	1	.786
29	.008	1	.927
30	1.124	1	.289
31	.238	1	.626
32	1.369	1	.242
33	.221	1	.639
34	2.094	1	.148
35	**5.197**	**1**	**.023**
36	.001	1	.974
37	1.587	1	.208
38	1.610	1	.204
39	.135	1	.713
40	.742	1	.389
41	1.209	1	.272
42	**4.033**	**1**	**.045**
43	.004	1	.953
44	2.067	1	.151
45	.548	1	.459
46	**27.207**	**1**	**.000**
47	1.104	1	.293
48	.001	1	.970
49	.439	1	.508
50	.156	1	.693

In Table 2(b), the number of WAEC items in Economics that function differentially with respect to location were detected using Mantel-Haenszel Chi-Square Statistics. With Chi-Square Test of Independence significance level below 0.05, the result indicated that, 8 items that is, item numbers 13, 15, 20, 24, 27, 35, 42 and 46 in the Economics test were identified as significantly exhibiting differential item functioning between students from urban and rural schools at .05 level of significance while the remaining 42 items did not significantly exhibit differential item functioning between male and female students. This reveals that Mantel-Haenszel Chi-Square Statistics detected and flagged 8 items as having DIF with respect to location.

Comparing the two statistics for detecting DIF, it was observed that Binary Logistic Regression Statistics detected 9 items while Mantel-Haenszel Chi-Square Statistics detected 8 items as having location DIF, hence, 6 items were consistent with the two statistics as having location DIF while the other items were not consistent. The consistent items are 15, 20, 24, 35, 42 and 46.

Discussion of Findings

It was revealed in this study that both statistics (Binary Logistic Regression Statistics & Mantel-Haenszel Chi-Square Statistics) detected equal number of 13 items as having gender DIF, but the difference is that 8 items were consistent with the two statistics as having gender DIF while 5 items were not consistent. The consistent items are 13, 15, 20, 24, 42, 44, 46 and 47. Hence, both statistics contributed equally in the detection of DIF with respect to gender in 2015 WAEC test, indicating that they have equal potency in the detection of gender DIF. Although the statistics are used to detect DIF in tests, Binary logistic regression is seen to consider the total test scores of the examinees while Mantel-Haenszel Chi-Square Statistics do not consider this. Also to be noted is that Binary logistic regression can help to detect both uniform and non-uniform DIF while Mantel-Haenszel Chi-Square Statistics can only be used to detect uniform DIF. This finding is in line with the findings of Pedrajita (2015) who revealed the presence of items indicating gender-based DIF, and thus concluded that there was a high degree of correspondence between the Logistic Regression and the Mantel-Haenszel Statistic in identifying potentially biased test items.

Further result of this study shows that Binary Logistic Regression Statistics detected 9 items while Mantel-Haenszel Chi-Square Statistics detected 8 items as having location DIF, hence, 6 items were consistent with the two statistics as having location DIF while the other items were not consistent. The consistent items are 15, 20, 24, 35, 42 and 46. This finding shows that Binary Logistic Regression Statistics has a higher potency in detecting location DIF than Mantel-Haenszel Chi-Square Statistics. This present finding is similar to Salubayba (2013) finding who revealed that Mantel-Haenszel (MH) detected fewer DIF items that resulted to more DIF-free items. On the hand, IRT-1PL detected more DIF items that produced brief test instrument. In line with the above finding, Ibrahim (2017) results showed that a non-significant relationship existed between the proportions of test items that functioned differentially in the dichotomous test items when the different methods were used. The study concluded that GMH, SIBTEST and LDFA were effective in detecting DIF across dichotomous test items but complement each other in their ability to detect DIF in dichotomous test items. The use of contingent related approaches could be linked to the result of this finding.

Conclusion

The study ascertained the potency of two statistical approaches (Binary Logistic Regression and Mantel-Haenszel Chi-Square) in detection of differential item functioning of standardized Economics test in Nigeria. Based on the findings accruing from this study, it was concluded that Binary Logistic Regression and Mantel-Haenszel Chi-Square statistics have equal potency in the detection of gender DIF with 13 items; but 8 items were consistent with the two statistics as having gender DIF while 5 items were not consistent. It was also concluded that Binary Logistic Regression Statistics has a higher potency in detecting location DIF with 9 items than Mantel-Haenszel Chi-Square Statistics that detected 8 items; but 6 items were consistent with the two statistics as having location DIF while the other items were not consistent.

Recommendations

The following recommendations are made based on the findings and discussions of the study;

1. Examining bodies, Evaluators, and educational practitioners involved in the development of assessment Economics instruments should use Binary Logistic Regression Statistics or Mantel-Haenszel Chi-Square Statistics for data analysis in detecting gender DIF items.
2. When detecting location DIF, the use of Binary Logistic Regression Statistics should be encouraged since it detected more DIF in this regard.

References

Ajeigbe, T. O. & Afolabi, E. R. I. (2014). Assessing unidimensionality and differential item functioning in qualifying examination for Senior Secondary School Students, Osun State, Nigeria. World Journal of Education, 4(4), 30-37.

Amaechi, C. E. (2018). Effects of scaffolding and guided inquiry instructional strategies on students' academic achievement and attitude towards Economics in Imo State. An Unpublished M.Ed. Dissertation, Imo State University, Owerri.

Amaechi, C. E. (2019). Detection of differential item functioning in economics multiple choice standardized test in Nigeria. An Unpublished Ph.D Dissertation, Michaal Okpara University of Agriculture, Umudike

Amajuoyi, I. J. (2015). Verification of differential item functioning (DIF) status of West African Senior School Certificate Examination (WASSCE) in Chemistry. Journal of Educational Foundations, 5(1), 165-175.

Bland, J. (2013). About gender: differences. http://www.gender.org.uk/about/oo diffs.htm.

Camilli, G. (2006). Test fairness. In R. L. Brennan (Ed.), 4Ih ed. (pp. 220-256). Westport, CT: American Council on Education.

Chang, Y. (2013). Gender Differences in Science Achievement, Science Self Concept and Science Values. [On line] available @Yuan Christian University @tea.ntue.edutw: Pp1-6.

Ibrahim, A. (2017). Empirical comparison of three methods for detecting differential item functioning in dichotomous test items. Journal of Teaching and Teacher Education, 5(1), 1-18.

Juweto, G. A. (2015). Effects of jigsaw co-operative teaching/learning strategy and school location on students' achievement and attitude towards Biology in Secondary School in Delta State. International Journal of Education and Research, 3(8), 31-40.

Kpolovie, P. I, Joe, A. I. & Okoto, T. (2014). Achievement prediction: Role of interest towards school. International Journal of sciences and Education, 1(1), 22-29.
Moyo, S. E. & Nenty, H. J. (2017). Dimensionality analysis of students' performance in 2013 BGCSE agricultural examination: Implications for differential item functioning. Research Journal of Educational Sciences,5(1),1-10.
Onah, F. E.,Amaechi, C. E.,& Nosike, M. C. (2017). Safety-health practices as correlate of Economics students' learning outcome in secondary schools in Imo State. The Educational Psychologist, 11(1), 211-218.
Osterlind, S. J., & Everson, H. T. (2009). Differential item functioning. Thousand Oaks, CA: Sage Publishing.
Osuafor, A. & Okonkwo, I. (2013). Influence of family background on academic achievement of secondary schools Economics students in Anambra State. International Multidisciplinary Journal, 7(3), 156-169.
Pedrajita, J. Q. (2015). Using contingency table approaches in differential item functioning analysis: A comparison. Education Journal, 4(4), 139-148.
Salubayba, T. M. (2013). Differential item functioning detection in reading comprehension test using Mantel-Haenszel, item response theory, and logical data analysis. The International Journal of Social Sciences, 14(1), 76-82.
Walker, C. (2011). What's differential item functioning? Why differential item functioning analyses are an important part of instrument development. Journal of Psycho-Educational Assessment, 29(1), 364-376.
Zumbo, D. B. (2012). A handbook on the theory and methods of differential item functioning (DIF): logistic regression modeling as a unitary framework for binary and likert-type item scores. Retrieved on 06/06/2012 from:
http://edu.ubc.ca/faculty/zumbo/dif/handbook.pdf.