Educational Research (ISSN: 2141-5161) Vol. 4(4) pp. 367-373, April 2013
Available online@ http://www.interesjournals.org/ER

Full Length Research Paper

Differential item functioning method as an item bias indicator

***Uruemu Ogbebor and Adams Onuka**

Institute of Education, University of Ibadan, Nigeria.

Abstract

Differential item functioning is an approach that is widely used to find out items that are bias. This study investigated items that are bias using differential item functioning approach in relation to school type (private and public schools), school location (urban and rural schools) using National Examinations Council (NECO) Economics questions for 2010. The research design employed in this study was a comparative research type of design. The study sample comprised students in Delta State, Nigeria. Four hundred and forty seven (447) students were used. And the test contains 60 items which was administered to the students. Logistic regression was used to analysis the data. The research findings showed that out of sixty items in NECO economics questions 10 items were biased in relation to school type and 8 items in relation to school location. The implication of these findings is that NECO economics examinations questions have presences of differential item functioning (DIF). From the result of the findings, it was then recommended that test experts and developers should explore the use of DIF approach to detect biased items.

Key words: Differential Item Functioning, Logistic Regression, Item Bias, Latent Trait.

INTRODUCTION

The issue of educational measurement in research pointing towards enhancing the fairness of test or examination across sub groups of examinees is very essential because important decisions are made based on scores of the examinee. Test consists of a set of uniform questions or task to which a student or testees is to respond independently and the result of which can be treated in such a way as to provide a quantitative comparison of the performance in different students (Nworgu, 2011).

The term testees or examinees can be used interchangeably it implies an individual or group of individual who is tested as by a standardized or teacher-made examination. Ogbebor, (2012) opined that, tesstees or test takers of the same latent trait should respond to test item correctly irrespective of their gender, school location and school type. A fair test is one that enables all examinees to have an equal chance to demonstrate the skills and knowledge which they have acquired and which are vital to the purpose of the test. Test fairness can be viewed as any test given to a set of testee with an equal chance to demonstrate what they know. Various aspects of fairness in testing have been highlighted in literature, including fairness in regards to standardization, test consequences/score use, and item bias (Kunnan, 2000; Shohamy, 2000).

A fair test is one that affords all examinees an equal opportunity to demonstrate the skills and knowledge which they have acquired and which are relevant to the test's purpose (Roever, 2005). The existence of bias is an issue to be addressed because tests are used as a gatekeeper for educational opportunities and it is a very important issue that test items are fair for every examinee. Bias is the existence of some irrelevant elements present in items that causes differential performance for individuals of the same ability but from different ethnic, sex, type of school attended, location of schools and cultural or religious groups. An examination item is said to be bias if it functions differently for a specified subgroup of test takers. Ogbebor, (2012) states that Bias test measure characteristics that are not

*Corresponding Author E-mail: ogbeborremu@yahoo.com, Tel: +234-803-702-7063

necessary or items that are irrelevant to the test. Frequently, examination items are considered biased because they contain sources of difficulty that are not relevant to the construct being measured and these extraneous sources affects test-takers' performance (Zumbo, 1999).

Item bias and differential item functioning (DIF) has critical political, social and ethical implication for test developers, policy makers and examines. The study of item bias and DIF is critical as such, this research would helps to provide an empirical foundation for the identification and subsequent elimination of examination items that appear to be relatively more difficult for one group of test-takers than another. Further research on these issues will allow us to comprehend more fully the possible substantive interpretation that can be made by focusing on test items considered to be biased.

Differential item functioning (DIF) is a statistical technique used to assess the existence of item bias, it is a systematic error in the predictive or construct validity of an item that may be attributable to factors irrelevant to the test.

Camilli, (1993), states that DIF analyses specify whether individuals of equal ability have the same probability of getting a given item correct. The modern approach for detecting item bias is by providing evidence of DIF. According to Roever, (2005), locating items on which a group of examinees perform significantly better than another group is logically the first step in detecting item bias. If an item on which a particular subgroup performs significantly better than another subgroups, it is said to have functioned differentially with respect to the two groups, Ogbebor, (2012) states that DIF occurs when a test item measures an ability which is unfamiliar to the subject matter, such that students' scores on the item is now sustained by abilities which are unfamiliar to the subject matter.

Studies have shown that differences in test of student achievement and low test scores in some subject areas such as mathematics and economics could be attributed to social and cultural influences that create sex role stereotypes that reduce female interest and achievement in traditionally male-dominated subjects (Williams et al., 1992; Hirschfeild et al.,1995). Studies have also shown that there are significant differences in the academic performance of students from rural and urban areas. Obe, (1984), observed that there is a significant difference in the performances of students from rural and urban schools in their academic performance; he therefore concluded that children from urban schools were superior to their rural counterparts. Owoyeye, (2002) also found out that there was a significant difference between academic performance of students in rural and urban area in public examinations. However, Ajayi and Ogunyemi, (1990) and Gana, (1997) in their different studies on the relationship between academic performance and school location revealed that there was no significant difference of students in urban and rural schools. While Ajayi, (1999) also found out that there was no significant difference between students academic achievement of rural and urban secondary school student.

A lot of research works have been conducted in this area of item bias. Pedrajita, (2009), in a study "using Logistic regression to detect test items in Chemistry Achievement", the result from the study revealed that there are gender bias and class bias in Chemistry Achievement test. Nworgu, (2011), revealed that current research evidence has implicated test used in national and regional examination as functioning differently with respect to different subgroups. This means that students' scores in such examinations are determined largely by the group to which an examinee belongs and not by ability. Gierl's, (1999), a study on DIF in Alberta examined 30 education Social Studies Diploma students, the study evaluated the effects of DIF between male and female, the results indicated that the majority of multiple choice items did not display DIF using the three-tiered ratings. Thus, 65 of 70 item displayed negligible effects, 5 items with moderate DIF, three favoured male and two favoured female, this indicate that the test contained items that functioned differently for male and female.

Item bias is of a particular concern on test of economics achievement, here differences in performance between, private and public urban and rural is commonly found. Therefore, this study finds out biased item in Economic national examinations questions using differential item approach.

METHODS

Research Design

This study is a comparative research

Population

The target population comprises of all students in SSS 3 in Delta State. Nigeria who enrolled for the senior school certificate examination in 2012 .Both male and female were involved in this research

Sampling Technique and Sample

Multi-stage sampling was employed. First purposive sampling techniques was used to select Ethiope East Local Government Area. Secondly Ethiope East Local Government Area was cluster into urban and rural areas. Purposive sampling was employed to select three (3) private schools and three (3) public schools from the urban area, and four (4) private school and three (3) public schools from the rural areas. The total schools

used for this study was 13 secondary schools in Ethiope East Local Government Area in Delta State Nigeria. An intact class was used in each of the school sampled. The total number of participant in this study was four hundred and forty seven (447) SSS 3 students who enrolled for the senior school certificate examination in 2012

Instrument

The response to each item of the NECO Economics examinations for all the students in the schools selected was used. The NECO Examinations is a standardized examination taken nationwide in Nigeria.

Data Analysis

Logistic regression was used to analyze the data. It involved the following steps

RESULTS

School Type

From table 1 shows the items in relation to school type (private and public), identified by logistic regression method using SPSS version 18.

Out of sixty items in NECO economics questions DIF was present in ten items. These items are item 9, 13, 14, 16, 43, 44, 46, 49, 50, and item 59.

School Location

From table 2 shows the items in relation to school type (private and public), identified by logistic regression method using SPSS version 18.

Out of sixty items in NECO economics questions DIF was present in eight items. These items are item 3, 7, 9, 11, 17, 47, 52 and 54.

DISCUSSION

Logistic regression statistics detected items that have DIF against subgroups such as public and private schools examinees, and it was revealed that out of the sixty items in NECO economics examinations question paper, ten items showed DIF these items are item 9, 13,14,16,43,46,49,50 and 59. Six item which are item 9, 13, 44 , 46, 49,and 59, favored private school students while the public school student were disadvantaged, while four items which are item 14,16,43, and 50 favored public schools than the private schools. The private schools on these items were disadvantaged.

Logistic regression also detected items that have DIF against subgroup such as urban and rural school students, and it was revealed that out of the sixty items in NECO economics examinations question paper, eight items showed DIF these items are 3, 7, 9, 11, 17, 47, 52 and 54.

From the findings, its 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. The findings of this study agrees with the work of Pedrajita, (2009) when he used Logistic regression to detect test items bias in Chemistry Achievement", the result from the study revealed that there is school type bias in the Chemistry Achievement test that was administered to the testees out of 22 items that were biased 11 items favoured public schools while eleven also favored private schools.

Nworgu, (2011), revealed that current research evidence has implicated test used in national and regional examination as functioning differently with respect to different subgroups. This means that students' scores in such examinations are determined largely by the group to which an examinee belongs and not by ability. Adedoyin (2010) in his study on investigating gender baised items in public examinations, he found that out of 16 test items that fitted the 3PL item response theory statistical analysis, 5 items were gender biased.

The finding of this study agrees with Felder, Mohr, Dietz and Ward (1994) who find out that urban student enjoy greater success than rural student, a result also supported by Tremblay, Ross and Berthelot, (2001), Kolcic, (2006) and Considine and Zappala, (2002). On the other hand the findings of this study disagree with Lee and McIntire, (2001) whose findings revealed that there is no significant difference between performance of rural students and urban students. This implies that items used in assessing student ability has element of biasness that disadvantaged the rural school examinees and favors the urban schools examinees.

CONCLUSION

Based on the forgoing findings the following conclusion were made

There were presences of school type and school location bias in NECO economics questions. On the basis of the findings and conclusion, the following recommendations are made:

1. 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.
2. A study of this should be conducted to provide further empirical evidence on the validity of the method in detecting biased test items.

Table 1. Logistic regression to detect school type 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	.351	.837	.577	1.216
8.	-.339	.191	.075	.712	.490	1.035
9.	.417	.195	.033*	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	.234	.797	.549	1.158
13.	.663	.201	.001*	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.	-.123	.235	.601	.884	.558	1.402
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	2.350
47.	-.209	.280	.455	.811	.469	1.404
48.	.263	.207	.203	1.301	.867	1.953

Table 1. Continued

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	.953	.986	.609	1.595
59.	.564	.207	.007*	1.758	1.171	2.639
60	-.060	.195	.760	.942	.643	1.381

Table 2. Logistic regression of sixty NECO item for school location.

Items	B	S.E.	Sig	Exp(B)	95% Lower	For Exp (B) Upper
1	.007	.223	.975	1.007	.650	1.560
2	-.226	.236	.340	.798	.502	1.268
3	1.220	.200	.000*	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	.300	.805	.535	1.212
7	-.840	.194	.000*	.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	2.341
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	.799	.948	.651	1.379
34	-.329	.218	.131	.720	.470	1.103

Table 2. Continued

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	.806	.948	.621	1.449
54	.598	.194	.002*	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	-.251	.196	.201	.778	.530	1.143

3. Evaluators and educational practitioners who are engaged in the development of assessment tools should use logistic regression for bias correction
4. Measurement practitioners should make use of logistic regression for developing a valid , reliable gender fair test school type fair test with biased items revised or replaced
5. The subject curriculum should be made clear for teachers to be able to teach the concept effectively
6. Teachers should exposure learners to more than one textbooks.

REFERENCES

Adedoyin OO (2010). Using IRT approach to detect gender biased items in public examinations: educational research and reviews academic journals. 5(7): 385-399

Ajayi K, Ogunyemi B (1990). The relationship between instructural resources and soci-economics status in selected population of high school. Dissertation Abstract International, 25(2): 22-23.

Ajayi IA (1999). Unit Cost of Secondary Education and student academic performance in Ondo State, Nigeria (1991 – 1995). Ph. D Dissertation. University of Ibadan, Ibadan

Camilli G, Shepard LA (1994). Methods for identifying biased test items. Newbury Park, CA: Sage.

Considine G, Zappala G (2002). Influence of social land economics disadvantaged in the academic performance of school students in Australia. Journal of sociology , 38, 129 – 148. Retrieved on August 16, 2007 from http:// jos. Sage pub.com

Embretson S, Reise SP (2000). Item response theory for psychologists. Mahwah New Jersey: Lawrence Erlbaum Associates Publishers.

Felder RM , Mohr PH, Dietz EJ, Ward LB (1994). A longitudinal study of engineering student performance and retention. II differences between students from rural and urban backgrounds. Journal of Engineering Education, 83(3):15–21,

Gana ES (1997). The effect of using visual designed training model on the learning of mathematics at J.S.S Ph. D. Thesis, Unpublished. University of Ibadan,

Gierl MJ (1999). Test bias, Sage publications, Inc., California.

Gierl MJ, Bisanz BG, Boughton KA, Khaliq SN (2001). Using differential bundle functioning to identify and interpret gender difference on science achievement tests. Paper presented at the annual meeting of the American education research association.

Hambleton R, Rodgers J (1995). Item bias review. Practical Assessment, Research, and Evaluation, 4(6). Retrieved March 18, 2009, from http://PAREonline.net/getvn.asp?v=4andn=6.

Hambleton R, Rodgers J (1995). Item bias review. Practical Assessment, Research, and Evaluation, 4(6). Retrieved November 18, 2006, from http://PAREonline.net/getvn.asp?v=4&n=6

Hambleton RK, Swaminathan H (1995). Item response theory: Principle and application. Boston: Kluwer.

Hambleton RK, Swaminathan H, Rogers HJ (1991). Fundamentals of item response theory. Newbury Park, California: Sage Publications.

Hirschfeld M, Moore RL, Brown E (1995). Exploring the Gender Gap on the GRE Subject Test In Economics. *Journal of Economic Education*, 26, Pp. 316

Kimball MM (1989). A new perspective on women's math achievement. Psychol. Bull. 105: 198-214.

Kolcic IOP (2006). Academic performance and scientific involvement of final year medical students coming from urban and rural background of Andjija stampar school of public Heaith, medical school, university of Zagreb, Croatia htt://rrh. Deakin.edu.au/

Kunnan AJ (2000). Fairness and justice for all. In A. J. Kunnan (Ed.), Fairness and validation in language assessment: Selected papers from the 19th Language Testing

Lee J, Mclritire W (2001). Interstate variation in the mathematics achievement of rural and non-rural students. Journal of research in rural education, 16 (3)

Nworgu BG (2011). Differential item functioning :Acritical issue in regional quality assurance. Paper presented in NAERA conference.

Nworgu BG, Odili JN (2005). Analysis of DIF in WAEC SSCE Biology Multiple Choice Test. Review of Education 16, (2) 140-152.

Ogbebor UC (2012). Differential Item Functioning Economics Question Paper of National Examinations Council in Delta State Nigeria. Unpublished M.ed thesis. University of Ibadan.

Owoeye JS (2000). The effect of integration of location, facilities and class size on academic achievement of secondary school students in Ekiti State, Nigeria. Ph.D thesis, unpublished. University of Ibadan

Pedrajita JQ (2009). using logistic regression to dectect Biased test items. The international of educational and psychological assessment journal

Perrone M (2006). Differential item functioning and item bias: Critical consideration in test fairness. Applied Linguistics, 6(2): 1-3.

Roever C (2005). "That's not fair!" Fairness, bias, and differential item functioning in language testing. Retrieved November 18, 2006, from the University of Hawai'i System Web site: http://www2.hawaii.edu/~roever/brownbag.pdf

Schumacker R (2005). Test bias and differential item functioning. Retrieved November 18, 2006, from http://www.appliedmeasurementassociates.com/White%20Papers/TEST%20BIAS%20AND%20DIFFERENTIAL%20ITEM%20FUNCTIONING.pdf

Shohamy E (2000). Fairness in language testing. In A. J. Kunnan (Ed.), Fairness and validation in language assessment: Selected papers from the 19th Language Testing

Tremblay S, Ross NA, Berthelot JM (2001). "Individual and community factors affecting grade three academic performance: a multi-level analysis." *Education Quarterly Review,* 7 (4), p. 25-36.

Williams ML, Walaver C, Duggal V (1992). Gender Differences In Economic. An Extension of the Analysis. Journal of Economic Education. No. 23 Vol.3. Pp

Williams V (1997). The "unbiased" anchor: Bridging the gap between DIF and item bias. Applied Measurement in Education, 10, 253-267.

Zumbo BD (1999). A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and Likert-like (ordinal) item scores. Ottawa, Canada: Directorate of Human Resources Research and Evaluation.