DIFFERENTIAL ITEM FUNCTIONING OF 2015/2016 BIOLOGY MULTIPLE-CHOICE ITEMS OF OSUN STATE JOINT PROMOTION EXAMINATION

BY

Bamidele Abiodun FALEYE: Department of Educational, Foundations and Counseling, Faculty of Education, Obafemi Awolowo University, Ile-Ife, Nigeria

&

Abdulwasiu Adeyemi RASHEED: Department of Educational, Foundations and Counseling, Faculty of Education, Obafemi Awolowo University, Ile-Ife, Nigeria

Abstract

Differential Item Functioning is one of the important and currently used methods to investigate test fairness. This study investigated the test items that function differentially based on gender and school location of the test takers in 2016 Biology multiple-choice items of Osun State Joint Promotion Examination (OSJPE). The research design adopted for this study was ex-post-facto design. The Sample for the study was 2,000 respondents who sat for OSJPE in 2016. The sample was selected using a multi-stage sampling procedure. Three Local Government Areas (LGAs) were selected from the three senatorial zones in Osun State, using a purposive sampling technique (where at least two schools were located at both rural and urban in each of the LGAs). From each Local Government Area, four schools (two from each of rural and urban centres) were selected using a stratified sampling technique, using location as a stratum. The students' responses of the 40 multiple-choice questions were subjected to the 3-parameter model of Item Response Theory analysis using the Calibrating process of analysis setting Mantel Haenszel as statistics. The result showed that One (item 5) out of the 40 multiple-choice Biology items functioned differentially for male students. Eight items (19, 20, 23, 28, 36, 37, 38 and 40) were biased against rural and urban students. The findings implied that there are incidences of differential item functioning in 2016 OSJPE Biology multiple-choice items. It was recommended that examination bodies should organise training for item developers on the construction of valid, reliable and fair tests especially in the area of DIF. In addition, items flagging DIF should be revised, modified or eliminated from the test.

Keywords: Differential item functioning, Item bias, Test fairness, OSJPE, Biology, Gender, Location

Introduction

The Osun State Joint Promotion Examination (OSJPE) is a state-based mock examination introduced in 2004 and is usually taken in Senior Secondary School II (SSS II) during the third term. The OSJPE serves two main purposes of preparing the students for better performance in external standardised examinations and for the State Government to pay for the West African Examination Council (WAEC) fees of those students who pass the OSJPE. The responsibility of conducting standardised examinations in Nigeria is put under the control of Examination bodies for certification of senior secondary schools students. The bodies include WAEC, the National Examination Council (NECO), National Business and Technical Board (NABTEB), and so on. It is incumbent on students to have good performance in one of these standardised examinations before they are admitted into higher institutions in Nigeria and abroad. External standardised examinations are so important that adequate preparation must be made by students to achieve excellently in them.

School administrators have engaged themselves in different methods to prepare their students for external standardised examinations. Among the methods is the use of mock examination conducted at school level and this has been faulted by many scholars as not being free from bias and teachers' manipulations. Students that take the same external standardised examination are not exposed to the same mock examination whose standards do not differ from one school to another. The psychometric properties of items are not ascertained in the examination. All these make it impossible to predict the performance of students in external standardised examinations.

Osun State government introduced an examination called Osun State Joint Promotional Examination (OSJPE) to adequately prepare the ground for all students in performing excellently in anyone of standardised examinations. The OSJPE is taken throughout the state and under the same examination conditions as these external examinations are marked centrally by the teachers who indicated interest and are paid like examiners of external examinations. The Biology test of OSJPE is to predict students' achievement in external standardised Biology examinations. It has been observed that students' performance in OSJPE Biology items is not so encouraging and students find it difficult to pass the subject. Biology as one of relevant science subjects, students who fail it will not avail themselves the opportunity to study biology-related courses such as Medicine, Agriculture, Zoology, Biochemistry, Dentistry, Microbiology, etc. in higher institution. Besides, Biology is the main player in understanding some latest biological issues such as genetic engineering, biotechnology, waste disposal, and food security. A shortage of experts in Biology-related professions will result if there is a continuous failure of Biology subject in external standardized examinations as OSJPE is to predict the student's performance in external examinations.

Critics have attributed the poor performance in the subject to non-validation of the examination items. To them, it is disputed whether or not the items of the examination are validated before administering them. In test validation, item fairness is one of the areas that are examined. A fair test is the one that refers to a test which gives equal opportunity of answering questions to all groups and individual (Roever, 2005, as cited in Perrone, 2006). In other words, according to Perrone (2006), similar performance on individual examination regardless of their gender, culture, ethnicity, or race must be scored by test takers with the same knowledge of all items in a test. Also, fairness is the degree at which teachers and testers treat every student the same by finding test questions, administration procedures, scoring methods, and reporting policies that enhance the chances that each student will receive equal and fair treatment (Brown, 2005). The examinees' scores are used to make important decisions; therefore educational measurement must be directed to give 'fair treatment' of test items across sub-groups of testees. Nworgu (2011) observed a test as a holistic approach for measuring students' responses independently to provide 'a quantitative description' of the performance in different students.

Test fairness can be investigated using various methods and procedures. This study, however, focuses on one of the important and currently used methods known as Differential Item Functioning (DIF). According to Beatrice, Imo and Friday (2014), DIF occurs when a measurement of a certain response on a test or questionnaire is characterized by "measurement bias" due to the different probability of giving a certain response by people of different groups (commonly gender, or ethnicity) with the same latent traits (ability/skills). According to Scheuneman (2009), a test item is described as differentially functioning when the probability of correct response is not the same for all persons of a given ability irrespective of their group membership. No individual or group answering a question should be disadvantaged in any way. Construction of assessment instruments needs to be free of bias, so that students of equal ability drawn from the same population but belonging to different subgroups such as male or female, urban or rural students will have the same probability of getting an item correct. This is tampered with when an item is biased.

A Test item that differentially affects groups or individuals from showing their true abilities and thereby measuring irrelevant construct is known as bias test item. This type of bias item that exhibits differential item functioning (DIF) systematically underestimates or overestimates the value of the variable the items are designed to measure (Reynolds, 2006). Penfield and Camilli (2007) also reported that DIF exists when a test item denies students from two sub-populations with the same ability level to have the same expected scores on the same item despite controls for overall test performance. Studies by Williams et al., (2011) and Hirschfeld et al., (2013) have shown that poor performance of students in some subject areas such as mathematics and biology could be attributed to the factors that have to do with gender in which

the males are favoured against females. In 2012 NECO Biology multiple-choice questions, Amuche and Fan (2014) identified 10 items that exhibit DIF in relation to school type and 8 items in related to school location. These findings imply that NECO Biology examinations questions have incidences of Differential Item Functioning (DIF).

Wang (2010) investigated DIF of 2006, 2007 and 2008 Biology test of the College Entrance Examination of Taiwan and the result showed that (i) Almost half of the items flagged DIF and this number of items that favoured the males and the females showed both sexes experienced the same disadvantages; (ii) The DIF estimated for all the year showed no clear cut of difference despite the decrease in the DIF-flagged from year to year, the amount of DIF did not display a significant difference among the three consecutive years; (iii) The different scoring rubrics did have significantly different effects on the amount of DIF displaying in each of the three years' data.

Obinne and Amali (2014) carried out studies of DIF on biology examination test items administered by the West African Examination Council (WAEC) and those administered by the national examination council (NECO) for the years 2000 and 2001. The results of the analysis showed that some of the items in the examinations functioned differently which indicated the existence of DIF effects thus, measuring what they were not supposed to measure. Odili (2004) also investigated DIF of Biology paper II objective test used by WAEC in 1999, 2000 and 2001 examinations. The study revealed that Biology multiple-choice questions used by WAEC in the SSCE contained test items that significantly function differently for students from high and low socio-economic status, urban and rural geographical locations and male and female testees. It was also found that simplifying the language of Biology test items brought about a significant reduction in differential item functioning. In another study, Adediwura (2013) conducted a study to identify differential item functioning in items concerning gender and students' course of study using IRT and GLM methods to compare the nature of DIF identified by the two models in a dichotomous test. The study adopted a descriptive survey design with a population consisting of all the part three students in the Faculty of Education of Obafemi Awolowo University Ile-Ife. However, both methods detected six (6) items as exhibiting uniform DIF while four (4) items were identified as exhibiting non-uniform DIF.

Studies have also shown that there are significant differences in the academic performance of students from rural and urban areas. Adedoyin (2010) observed that there was a significant difference in the performances of students from rural and urban schools in their academic performance; he therefore, concluded that students from urban schools were superior to their rural counterparts. Owoeye (2000) also found out that there was a significant difference between academic performance of students in rural and urban areas in public examinations. In the different studies of Ajayi and Ogunyemi (1990) and Gana (1997) on the relationship between academic performance and school location revealed that there was no effect of school location (urban and rural) on students' performance while Ajayi (1999) was also on the same page with Ajayi and Ogunyemi (1990) and Gana (1997).

Scheuneman (2005), claimed that research in differential item functioning should make a step forward from identifying test items that display DIF to that of examining the underlying factors that bring about unfairness in test items and ways to improve or eliminate the defect items. The presence of DIF in Biology items of OSJPE is a cause for great concern considering that test results are to be used to predict students' performance in external standardised examinations as well as good indicators of peoples' ability level of performance in the subject. It is in furtherance to this background that the present study concerns itself with the investigation of the DIF of Biology items of Osun State Joint Promotion Examination (2015/2016) in relation to school location and gender.

Objectives of the Study

The specific objectives of the study were to:

(a) determine the test items that function differentially based on gender among the test test-takers and;
(b) investigate the test items that function differentially based on school location of the test-takers.

Research Questions

These research questions were raised to achieve the objectives of the study:

(1) What are the test items that function differentially based on gender among the test-takers?
(2) What are the test items that function differentially based on school location of the test-

Methodology

The research design used was ex-post-facto. What features in ex-post-facto design are; the existence of a control or comparison group, the use of intact groups and non-manipulated data. This design is closely connected to this study because it allows analysis to be performed on existing data. In this case, the responses of students to multiple-choice items in Biology of 2016 OSJPE constituted the data for the study. Also in ex-post-facto design, manipulation becomes impossible and data collected are near perfection since they are collected in a controlled environment. This technique enabled the researcher to predict accurately the characteristics or thoughts of a pre-defined group of students.

The population for the study consisted of 17, 612 Senior Secondary School II students who sat for Biology in the 2016 OSJPE conducted by Osun State Ministry of Education. Sample for the study was 2,000 responses of the students who sat for OSJPE in 2016. The sample for the study comprised 998 (49.7%) male and 1012 (50.3%) female students. All the students' responses to 40 multiple-choice Biology items of the 2015/2016 OSJPE were used in the study. The sample was selected using a multi-stage sampling procedure. Three Local Government Areas (LGAs) were selected from the three senatorial zones in Osun State, using a purposive random sampling technique. From each Local Government Area, four schools (two from each of rural and urban centres) were selected using a stratified sampling technique, using location as a stratum. Thus, a total number of 36 schools were selected for the study. The research instruments used were the 40 multiple-choice Biology questions of 2016 OSJPE and response sheets of all the students that wrote the Biology 40 item multiple-choice questions of OSJPE (during the 2015/2016 session) in the selected schools as contained in the Optical Mark Recorder (OMR) and their keys. All the students' responses to the 40 multiple-choice Biology items during the 2015/2016 OSJPE in the selected schools constituted the sample for the study as shown in Table 1.

Table 1: No of students' response sheets in the selected schools

School	No of response sheets	School	No of response sheets

1	50	19	77
2	64	20	51
3	53	21	57
4	61	22	42
5	69	23	65
6	45	24	68
7	70	25	71
8	44	26	37
9	60	27	41
10	54	28	75
11	40	29	67
12	35	30	75
13	70	31	62
14	63	32	38
15	39	33	70
16	43	34	35
17	50	35	58
18	59	36	42
Total	969		1031= 2000

Table 1 shows total number of selected schools and number of students' response sheets totaled 2000 after which some have been removed for the fact that the name on those sheets appeared unisex and some were without any response.

The 40 multiple-choice Biology questions were administered on Senior School II students in their respective schools under the supervision of the representatives of State Ministry of Education and appointed school supervisors in each school. The demographic data of each of the students such as sex, the school, local government area, centre number, candidate number were printed on the OMR sheet to ensure proper coding for computer analysis. In order to access the OMR sheets containing the students' responses to Biology items of OSJPE 2015/2016, the researcher visited the Osun State Ministry of Education. The researcher presented a letter of introduction and it was properly documented by the Ministry of Education and processed for further action. Later, appointment was given to the researcher to allow him search for the OMR sheets and questions for the subject in the store with the assistance of one of the members of staff in the Ministry of Education. Data collection of 2,000 responses of students took two days and all the data were found useful. The key that was used for the subject was also made available by the Ministry. The data were uploaded into the computer before the analysis was carried out by the researcher.

Data collected were scored dichotomously (right = 1, wrong = 0). The data file was prepared using Microsoft Excel 2010. A specialized Software (i.e. XCalibre 4.2) was used in order to analyse the Biology test items in this study. Research questions 1 and 2 were answered using calibrating process of analysis for the coded data in Xcalibre 4.2-IRT statistical software using 3PL, setting Mantel-Haenszel as statistics for analyzing possible item that may flag DIF with respect to school location and gender.

Results and Discussion

Research Question 1: What are the test items that function differentially based on gender among the test-takers?

To answer this question, the Optical Mark Readers' (OMR) of the students who sat for Biology during the 2016 Osun State Joint Promotion Examination were entered into an excel spreadsheet. The correct option was coded "1", while the incorrect option was coded "0" for the 40-multiple choice Biology items. The coded items were then moved to note pad for formatting order in the language that could be understood by Xcalibre 4.2 statistical software. The 40 Multiple-choice items were then calibrated under the 3PL, setting the Mantel Haenszel as statistics for analysing possible items that may flag Differential Item Functioning (DIF) in terms of gender. The results of the analyzed data are presented in Table 2.

Table 2: DIF test results of Mantel-Haenszel statistic of the multiple-choice Biology test items based on 3PL IRT models for gender

ITEM	M-H D	3PL P	M/F
1	0.311	0.306	N/A
2	0.243	0.42	N/A
3	0.244	0.413	N/A
4	-0.28	0.366	N/A
5	0.772	**0.011**	MALE
6	0.01	0.976	N/A
7	0.266	0.385	N/A
8	0.16	0.601	N/A
9	-0.36	0.254	N/A
10	-0.042	0.892	N/A
11	-0.111	0.726	N/A
12	0.251	0.422	N/A
13	0.348	0.266	N/A
14	0.322	0.304	N/A
15	0.14	0.675	N/A
16	0.461	0.137	N/A
17	0.565	0.064	N/A
18	-0.298	0.369	N/A
19	-0.073	0.821	N/A
20	-0.082	0.794	N/A
21	-0.192	0.586	N/A
22	-0.123	0.709	N/A
23	0.24	0.468	N/A
24	0.049	0.88	N/A
25	-0.099	0.767	N/A
26	0.234	0.465	N/A
27	-0.098	0.763	N/A
28	0.482	0.139	N/A
29	0.086	0.788	N/A

30	-0.18	0.628	N/A
31	0.091	0.8	N/A
32	0.537	0.1	N/A
33	-0.01	0.975	N/A
34	0.138	0.657	N/A
35	0.162	0.627	N/A
36	0.453	0.177	N/A
37	0.274	0.423	N/A
38	0.25	0.441	N/A
39	-0.006	0.984	N/A
40	-0.035	0.913	N/A

The OSJPE Biology test was calibrated using the IRT Mantel-Haenszel statistic to test if items exhibit DIF with respect to gender. The presence of DIF was determined under the three-parameter model (3PL). Items with a p-value of less than 0.05 revealed significant DIF. When this is applied to the OSJPE Biology test items, Mantel-Haenszel significant value less than 0.05 procedure under the three-parameter logistic models showed that 1 item (item 5) flagged DIF representing 2.5% of the 40 items as displaying DIF under the model.

Research Question 2: What are the test items that function differentially based on school location of the test-takers?

To answer this question, the Optical Mark Readers' (OMR) of the students who sat for Biology during the 2016 Osun State Joint Promotion Examination were entered into excel spread sheet. Correct option was coded "1", while incorrect option was coded "0" for the 40-multiple choice Biology items. The coded items were then moved to note pad for formatting order in the language that could be understood by X-Calibre 4.2 statistical software. The 40 Multiple-choice items were then calibrated under the 3PL, setting the Mantel Haenszel as statistics for analysing possible items that may flag Differential Item Functioning (DIF) in terms of school location. The results of the analysed data are presented in Table 3.

Table 3: DIF test results of Mantel-Haenszel statistic of the multiple-choice Biology test items based on the 3PL IRT models for school location

		3PL	
Item	**M.H.D**	**P**	**U/R**
1	0.263	0.387	N/A
2	0.255	0.395	N/A
3	0.011	0.971	N/A
4	0.267	0.372	N/A
5	0.375	0.218	N/A
6	0.512	0.092	N/A
7	0.194	0.524	N/A
8	0.376	0.211	N/A
9	0.028	0.928	N/A
10	0.456	0.125	N/A
11	0.415	0.177	N/A
12	0.257	0.407	N/A
13	-0.103	0.745	N/A

ISSN 2384-7662
E-ISSN 2705-2508

14	0.501	0.104	N/A
15	-0.631	0.068	N/A
16	0.189	0.545	N/A
17	-0.056	0.856	N/A
18	0.232	0.47	N/A
19	0.643	**0.038**	URBAN
20	0.762	**0.011**	URBAN
21	0.637	0.059	N/A
22	0.001	0.998	N/A
23	-0.779	**0.026**	RURAL
24	0.116	0.718	N/A
25	0.594	0.063	N/A
26	0.243	0.444	N/A
27	0.055	0.863	N/A
28	-0.844	**0.016**	RURAL
29	0.516	0.099	N/A
30	0.02	0.956	N/A
31	-0.429	0.244	N/A
32	-0.297	0.379	N/A
33	0.163	0.609	N/A
34	-0.295	0.349	N/A
35	-0.045	0.892	N/A
36	-0.915	**0.011**	RURAL
37	-1.133	**0.002**	RURAL
38	-1.837	**0**	RURAL
39	-0.167	0.596	N/A
40	-0.818	**0.013**	RURAL

The OSJPE Biology test was evaluated using the IRT Mantel-Haenszel statistic to test if items exhibit DIF with respect to school location. The three parameter model (3PL) was used. For item to flag DIF, the Mantel-Haenszel significant value must be less than 0.05. When this is applied to the OSJPE Biology test, the resulting Mantel-Haenszel statistic for DIF, (Table 4.4.1) showed items that flag DIF, p-value less than 0.05 procedure flagged eight (8) items representing 20% of 40 items exhibited DIF under the three-parameter model. Out of these items, six (6) items (the items: 23, 28, 36, 37, 38 and 40) were biased against examinees of rural area and two (2) items (the items: 19 and 20) were biased against examinees of urban area. In conclusion, the results of DIF using the three parameter logistic model showed that eight (8) items out of the 40 Multiple-Choice Biology items functioned differently by school location (the items: 19, 20, 23, 28, 36, 37, 38 and 40). Eight items representing 20% exhibited DIF under the 3PL model.

The findings of research questions 1 indicated that 1 (item 5) out of the 40 multiple choice Biology items functioned differentially for male students across the 3 models. The finding of this study agrees with the findings of Adedoyin (2010), who in his study investigated gender-biased items in public examinations, and found that out of 16 items that fitted 3PL item response theory statistical analysis, 5 items were gender biased. The finding also agreed with that of Adebule (2013) that out of 40 items examined for the

first-factor program structure in computer science, only seven items representing 17.5% displayed DIF, comparing male and female examinees.

Finally, research question 2 showed that 8 items were biased against rural and urban students under the 3PL model. Items 19 and 20 were biased against urban students under 3PL model. Six items were found to function differentially for rural students under the model. However, if test items are biased, the learners would be prevented from demonstrating appropriate knowledge, skills and attitude they had acquired. To be able to confirm this, this study is in line with Adebule (2013); he found out in his study that the testees in the various groups (gender, age, parents' qualification and location) are of comparable ability levels. Hence, these biased items were considered to have defects and are, therefore, needed to be reviewed especially if the items are to be used in subsequent examinations.

Conclusion

The study concluded that the multiple-choice items of 2016 OSJPE Biology Examination were not fair to all groups of the test takers. As a result of these item defects, the 40 multiple-choice Biology items of 2016 OSJPE were considered to have low psychometric quality and this could be basis for the failure of students in the examination. Beside gender and school location-related DIF, the poor performance of item such as the difficulty and discrimination, and lack of unidimensionality of test items may also be a cause to the weak performance of students. Similarly, the results of this study shows the need to improve the OSJPE biology items especially in developing valid and reliable items, through the mandatory involvement of expert in test, measurement and evaluation in the process.

Recommendations

From the findings and conclusion of the study, the following recommendations are made:

(1) Further study needs to be conducted with different test items and should include differential item functioning analysis to ensure that valid and reliable measuring items are used in OSJPE.

(2) Examination bodies should organise training for item developers on the construction of valid, reliable and fair tests especially in the area of DIF. In addition, items flagging DIF should be revised, modified or eliminated from the test.

References

Adebule, S.O. (2013). A study of differential item functioning in Ekiti State unified mathematics examination for senior secondary schools. *Journal of Education and Practice*, (17), 43-46.

Adediwura, A. A. (2013). A Comparative Study of Item Response Theory and Generalized Linear Model Methods of detecting differential item functioning in dichotomous test. *Research Journal in Organizational Psychology and Education Studies*, *2*(6), 308-316.

Adedoyin, O. O. (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 structural Resources and socio-Biology status in selected population of high school (Abstract). *Dissertation Abstract International*, *25*(2), 22-23.

Amuche, C. I. & Fan, A. F. (2014). An Assessment of Item Bias Using Differential Item Functioning Technique in NECO Biology Conducted Examinations in Taraba State Nigeria. *American International Journal of Research in Humanities, Arts and Social Sciences*, *14*(84), 95-100.

Beatrice, O. N., Imo, E. U. & Friday, O. I. (2014). *Differential Item Functioning of 2010 Junior Secondary School Certificate Mathematics Examination in Southern Educational Zone of Cross River State, Nigeria*. Unpublished manuscript.

Brown, J. D. (2005). *Testing in language programs, a comprehensive guide to English Language assessment*. New York: McGraw-Hill College.

Gana, E. S. (1997). *The effect of using visual designed training model on the learning of mathematics at J.S.S* (Unpublished doctoral dissertation). University of Ibadan, Ibadan.

Hirschfeld, M., Moore, R. L. & Brown, E. (2013). Exploring the Gender Gap on the GRE Subject Test in Biology. *Journal of Economic Education, 26*, 316.

Nworgu, B. G. (2011). Differential item functioning: A critical issue in regional quality assurance. *Paper presented in NAERA conference*.

Obinne, A. D. & Amali, A. O. (2014). Differential Item Functioning: The Implication for Educational

Testing in Nigeria. *International Review of Social Sciences and Humanities, 7* (1), 52-65.

Odili, G. A. (2004). *Differential item functioning of Biology Paper II objective test used by WAEC in 1999, 2000 and 2001 examinations.* Lagos: Rex Charles and Patrick Limited.

Owoeye J. S. (2000). *The effect of integration of location, facilities and class size on academic achievement of secondary school students in Ekiti State, Nigeria* (Unpublished doctoral dissertation). University of Ibadan, Ibadan.

Penfield, R. D. & Camilli, G. (2007). *Differential item functioning and item bias*. Newbury Park, CA: Sage.

Perrone, M. (2006). Differential item functioning and item bias: Critical considerations in test fairness. *Columbia University Working Papers in TESOL & Applied Linguistics, 6* (2), 1-3.

Scheuneman, J. (2005). An experimental study of causes of bias in test items. *Journal of Educational Measurement, 4*, 97-118.

Scheuneman, J. (2009). A method of assessing bias in test items. *Journal of Educational Measurement, 16*, 143-152.

Reynolds, R. C. (2006). *Measurement and Assessment in Education*. Boston: Pearson.

Roever, C. (2005). *"That's not fair!" Fairness, bias, and differential item functioning in language testing*. Retrieved from http://www2.hawaii.edu/~roever/brownbag.pdf

Wang, C. (2010). *Differential item functioning analysis of biology test of the College Entrance Examination Taiwan* (Unpublished doctoral dissertation). University of Illinois, Urbana-champaign.

Williams M. L, Walaver, C, & Duggal V. (2011). Gender Differences In Economic: An Extension of the Analysis. *Journal of Economic Education, 23*(3), 23-31.