A Comparison of Three Methods for Detecting DIF in Osun State Qualifying Examination for Senior Secondary School Students

By

AJEIGBE Taiwo Oluwafemi
taiaje215@yahoo.com
Department of Educational Foundations and Counselling, Faculty of Education,
Obafemi Awolowo University, Ile-Ife, Nigeria

AFOLABI Eyitayo Rufus Ifedayo
eriafolabi@oauife.edu.ng
Department of Educational Foundations and Counselling, Faculty of Education,
Obafemi Awolowo University, Ile-Ife, Nigeria

Abstract
The study compared the relative ability of chi-square, transformed item difficulty and b-parameter statistical methods in detecting DIF in the examination in terms of sex and school location. The population for consisted of responses of students to 50 multiple-choice English Language and 50 multiple-choice Mathematics items of the Osun State Qualifying Examination (OSQE) for 2008.The sample consisted of 4156 students' responses from four stratified secondary schools from three purposively selected local government areas of the state. Data collected were analyzed using chi-square, transformed item difficulty and b-parameter, percentages and one-way analysis of variance. The results showed that there was a significant difference in the relative ability of the three methods of detecting DIF in the OSQ examination ($F_{(2, 3)} = 9.301, p<.05$).Chi-square method was most effective than Transformed Item Difficulty (TID) and B-parameter in detecting DIF items in Mathematics in terms of sex. Chi-square, TID and B-parameter identified 15(30%), 0(0%), and 4(8%) items as exhibiting DIF respectively. Also, B-parameter was most effective in detecting DIF items in Mathematics in terms of school location. Chi-square method was most effective in detecting DIF items in English Language with respect to sex. While Chi-square method was most effective in detecting DIF items in English Language in terms of school location. The study concluded that the Chi-square method demonstrated greater statistical power in detecting DIF items than TID and B-parameter in detecting DIF in Osun State Qualifying Examination.

Keywords: Differential Item Functioning (DIF), Chi-square (χ^2), Transformed Item Difficulty (TID) and B-parameter

1. Introduction

The study of Differential Item Functioning (DIF) is an integral part of determining the validity and reliability of standardized tests. In the context of tests, DIF occurs when students with similar abilities but from different subgroup of a population respond differently to test items. For example, if boys display higher probability of answering mathematics test items correctly more often than girls of equal ability level because the contents in the test items, than the items are biased against girls, hence, the items are said to exhibit DIF and should be considered for modification or removal from the test. The presence of DIF can inform test experts about the effect gender, location or ethnicity on the testees' ability test scores (Hambleton, Swaminathan & Rogers, 1991).

When standardized tests are administered on test takers, the test-taking population could vary on a number of personal and educational characteristics such as age, gender, first language, environment, and academic discipline. From the researcher's personal experience and observations, some test developers do not always take into cognizance the diversities that characterized the test takers before administering such test. This could result into various kinds of errors especially scoring error that inflates scores for one group at the expense of the other. Consequently, such test may be regarded as unreliable or lack test fairness. In a standardized test, item characteristics such as difficulty index, discrimination index, reliability, and validity must have been determined or established before they are administered on students. In this study, the Osun State Qualifying Examination (OSQE) was examined to ascertain the extent to which the test items are DIF free, bias free.

The state examination was introduced in 2004 as an intervention measure to arrest the decline and enhance better performance of students in public examinations. It is a qualifying examination for SS II students in public secondary schools and only the students that pass the State Qualifying Examination (SQE) are privileged to sit WAEC SSCE at government expense. This was to reduce students' failure rate in the secondary school certificate public examination to motivate and students to be more serious and diligent in their studies. The performance of students in the State qualifying examination since inception has been encouraging such that, the State has not reneged to set aside sufficient fund annually for adequate preparation, administration, and grading of scripts.

However, despite the huge amount being expended by the state government students' performance in public examinations has been generally unsatisfactory, especially in core subjects such as Mathematics and English Language. Given that the teachers and students have put in efforts in academic preparation because of the high stake attached to the examination, it is important to address the quality of the test items used for the state examination. Therefore, it is pertinent to direct attention towards quality of the test items used by the Osun State Ministry of Education to prepare its students for public examinations. In Nigeria, Mathematics and English language are core and pre-requisite subjects for gaining admission into higher institutions of learning.

DIF analyses being part of the framework for probing item bias. For instance, if a focal group (e.g., black or female) performs lower on a specific item, when compared with a reference group (e.g., white or male), after controlling for the overall differences in their ability scores, then one could say that the item is biased against the focal group. Differential item functioning analyses compare the performance of two groups of the same level of ability in order to disentangle the effects of unfairness and ability level. Matching ability level is essential, since different groups may have different ability levels, in which case differences in performance are to be expected (Clauser & Mazor, 1998). Consistent differences between two groups of the same ability level would suggest that DIF is present

There are many competing approaches for the conduct of DIF analyses and many criteria for determining what constitutes significant DIF in items that are scored dichotomously. Although many DIF methods abound, a relatively small number of these methods are preferred based on their theoretical and empirical strengths (Clauser & Mazor, 1998). Three of the preferred methods frequently used to detect item with DIF are Chi-square test, Transformed Item Difficulty, b-Parameter. Literature had shown that there were divergent results on the effectiveness of these three methods based on their strength and power in identifying DIF items. For instance, in the study carried out by Mark, Michael & Terry (2000), it was concluded that out of the three methods, TID has highest efficacy. In another study carried out by Nabeel (2010) it was revealed that chi-square method displayed the highest efficacy. However, since strength and power of each method differs from one study to another, it is necessary to identify which of the methods will be more adequate and appropriate for DIF detection in state examinations such as that of OSQE.

There are many methods for DIF detection proposed over the past two decades. This study will focus on three of these methods; Chi-square method, Transformed Item Difficulty (TID) and b-parameter difference. This study will compare the relative effectiveness of these three methods when applied on Mathematics and English items of the Osun State Qualifying Examination of 2008. The three methods are briefly discussed below.

In the Chi-square method, items that are not flagged DIF are those which probability of a correct response is the same for all persons of equal ability, irrespective of group membership. The Transformed Item Difficulty (TID), involves computing the difficulty or p-value (proportion of subjects getting item right) for each item separately for each group. Using tables of the standardized normal distribution the normal deviate z is obtained corresponding to the (1-p) th percentile of the distribution, i.e., z is the tabled value having proportion (l-p) of the normal distribution below it. Then to eliminate negative z-values, a delta value is calculated from the z-value by the equation $\Delta = 4z + 13$. A large delta value indicates a difficult item. For two groups, there will be a pair of delta values for each item. These pairs of delta values can then be plotted on a graph, each item represented by a point on the graph. A line can be fitted to the plot of points; and the deviation of a given point from the line is taken as measure of that item's bias, (Subkoviak; Mack, Ironson, & Crag, 1987). This procedure has been used to study cultural differences in a wide variety of contexts (Angoff, 1993). An item's b parameter is the point on the ability scale corresponding to the location on the S-function item characteristic curve where the probability of a correct response is 0.5 (Hambleton, Swaminathan, & Rogers, 1991). In other words, b is determined by first locating the point on the ICC that corresponds to a 50% chance of getting the item right (0.5 on the y-axis), and then determining the value of θ (on the x-axis) that corresponds to that point on the ICC. Items that are difficult will have higher b values and will be located at the right or higher end of the θ scale, which indicates that a greater level of ability is required in order to answer them correctly. Conversely, easy items will have lower b parameter values, will stay to the left (lower) end of the θ scale, and will require less ability to answer them correctly (Hambleton et al., 1991).

2. Statement of the Problem

Differential Item Functioning can lead to an unfair advantage or disadvantage for certain subgroups in educational and psychological testing. There are many competing approaches for the conduct of DIF analyses and many criteria for determining what constitutes significant DIF in items that are scored dichotomously. Although many DIF methods abound, a relatively small number of these methods are preferred based on their theoretical and empirical strengths (Clauser & Mazor, 1998). Three of the preferred methods frequently used to detect item with DIF are Chi-square test, Transformed Item Difficulty, b-Parameter. Literature had shown that there were divergent results on the effectiveness of these three methods based on their strength and power in identifying DIF items.

3. Purpose of the Study

The study was conducted to:

(1) determine which of the three methods is most effective in detecting DIF in the selected subjects on the basis of sex and school location; and

(2) compare the relative ability of Chi-square, Transformed Item Difficulty and B-Parameter methods in detecting DIF in the examination in terms of sex and school location.

4. Research Questions

1. Which of the three methods of Chi-square, Transformed Item Difficulty (TID) and b-parameter is most effective in detecting DIF in OSQ Mathematics examinations in terms of sex (male/female) and school location (rural and urban)?
2. Which of the three methods of Chi-square, Transformed Item Difficulty (TID) and b-parameter is most effective in detecting DIF in OSQ English Language examinations in terms of sex (male/female) and school location (rural and urban)?

5. Research Hypothesis

One research hypothesis was postulated for the study.

H_{o1}: There is no significant difference in the relative ability of Chi-square, Transformed Item Difficulty (TID) and b-parameter in detecting DIF in OSQ examination.

6. Methodology

The research design used was ex-post-facto. Ex-post-facto design is a method of testing out possible antecedents of events that have happened and cannot, therefore, be manipulated by the investigator. When translated literally, ex-post-facto means 'from what is done afterwards'. In the context of social and educational research, the phrase means 'after the fact' or 'retrospectively' and refers to those studies which investigate possible cause-and-effect relationships by observing an existing condition (Gay, Mills, & Airasian, 2006).

The population for this study consisted of the students who sat for 2008 Osun State Qualifying Examination. A sample of 4156 students' responses to 50 multiple-choice Mathematics and 50 multiple-choice English Language items of the OSQE for 2008 were used in the study. Three local government areas were purposively selected (where at least two schools are found to be located at both rural and urban areas in each LGA) from each of the three senatorial districts in the State. From each LGA, four schools (two each from rural and urban) were selected using stratified sampling technique. Thus, a total of 36 schools were used for the study. All the students' responses to 50 multiple-choice English Language and 50 multiple-choice Mathematics items of the OSQE for 2008 examination in the selected schools constituted the sample for the study. These represent 4,156 students' responses each for Mathematics and English Language respectively. A rural population for the purpose of this study was identified as communities with less than 20,000 people and where some of the basic infrastructures like electricity, pipe borne water, motorable roads, hospitals and post office are absent (National Population Commission, 1991).

Two research instruments were used in the study. The first instrument was responses of all the students to 50 multiple-choice English Language and 50 multiple-choice Mathematics examination in the selected schools in 2008 as contained in the Optical Mark Recorder (OMR) sheets. The second instrument was 50 multiple-choice English Language and 50 multiple-choice Mathematics questions of the OSQE for the year under consideration. The instruments are assumed to have been moderated and validated by the Stated Ministry of Education before they are administered on the students. In any case, the researcher found out that the 50 multiple-choice English Language and 50 multiple-choice Mathematics questions covered a wide range of topics in the Senior Secondary School (SSS) syllabus, showing that it had content validity. The reliability coefficients of the students' responses to the 50 multiple-choice English Language and 50 multiple-choice Mathematics questions using Cronbach's Alpha coefficient were found to be 0.87 and 0.88 respectively.

7. Results

Research Question 1: Which of the three methods (Chi-square, Transformed Item Difficulty and b-parameter) is most effective in detecting DIF in OSQ Mathematics and English Language examinations in terms of sex (male and female)?
The results are presented in Tables 1 and 2 below.

Table 1: Summary of Results from the Chi-square, TID, and B-parameter Methods of Identifying Differential Item Functioning on the OSQ Mathematics and English Language Examinations Based on Sex

Items	Mathematics						English Language					
	Chi-Square		TID		B-parameter		Chi-Square		TID		B-parameter	
	χ^2	Odds Ratio	Δ	D_i	$\Delta b = F_F - b_R$	d Statistics	χ^2	Odds Ratio		D_i	$\Delta b = F_F - b_R$	d Statistics
1	0.43	0.96	11.00	0.01	-0.01	-0.31	0.39	1.06	11.00	0.07	0.01	0.31
2	0.13	1.02	14.20	0.06	0.01	0.18	0.00*	0.79	16.60	0.54	-0.04	-0.94
3	7.39*	1.19	13.80	0.25	0.04	1.39	0.23	0.93	12.60	0.68	-0.02	-0.47
4	8.83*	1.21	11.80	0.22	0.05	1.51	0.55	0.96	13.00	0.49	-0.01	-0.24
5	22.55*	1.35	11.80	0.42	0.07	2.42*	0.30	1.07	13.40	0.30	0.02	0.40
6	0.69	1.06	10.60	0.20	0.01	0.38	0.24	1.08	11.00	-0.53	0.02	0.42
7	0.91	1.06	11.80	0.22	0.01	0.48	0.90	1.00	10.60	-0.14	-0.00	-0.04
8	3.26*	1.12	13.00	0.24	0.03	0.93	0.18	0.91	11.00	-0.14	-0.02	-0.46
9	0.32	0.96	12.20	0.03	-0.01	-0.29	0.35	0.94	13.40	0.69	-0.02	-0.36
10	0.65	1.06	15.80	0.08	0.01	0.35	0.46	0.95	10.60	0.06	-0.01	-0.24
11	0.22	1.03	14.20	0.06	0.01	0.27	0.07	1.14	15.00	-0.07	0.03	0.64
12	0.24	1.03	14.60	0.06	0.01	0.23	0.31	1.07	10.60	0.06	0.01	0.36
13	7.80*	1.20	14.20	0.26	0.04	1.37	0.98	1.00	10.60	-0.34	0.00	0.01
14	1.99	1.11	15.80	0.28	0.02	0.61	0.04*	0.87	11.80	0.27	-0.03	-0.78
15	1.81	1.10	11.00	0.21	0.02	0.64	0.04*	0.85	17.00	-0.25	-0.02	-0.60
16	0.29	1.04	12.20	0.03	0.01	0.28	0.88	0.99	11.00	0.07	-0.00	-0.05
17	7.34*	0.84	14.20	-0.14	-0.04	-1.38	0.02*	0.86	13.40	-0.30	-0.04	-0.94
18	3.71	0.89	13.00	-0.16	-0.03	-0.99	0.72	0.98	10.60	-0.14	-0.01	-0.12
19	0.25	1.01	12.20	0.02	0.00	0.08	0.47	1.05	11.80	-0.13	0.01	0.28
20	3.66	1.31	11.80	0.22	0.03	0.96	0.23	1.08	14.20	-0.67	0.02	0.45
21	3.24	1.16	16.60	0.29	0.02	0.70	0.12	1.10	12.60	-0.11	0.02	0.60
22	2.57	0.90	12.20	-0.17	-0.02	-0.81	0.09	0.90	12.60	0.09	-0.03	-0.66
23	12.55*	1.25	13.40	0.24	0.05	1.80	0.70	0.98	13.40	-0.10	-0.01	-0.15
24	2.92	0.90	13.80	0.05	-0.03	-0.87	0.01*	0.84	13.40	0.10	-0.04	-1.07
25	7.01*	1.18	13.40	0.24	0.04	1.36	0.80	0.98	11.00	-0.72	-0.00	-0.09
26	0.06	0.98	14.20	0.06	-0.00	-0.12	0.24	0.93	12.20	-0.71	-0.02	-0.44
27	15.06*	1.28	12.20	0.23	0.06	2.00*	0.54	0.96	11.00	-0.33	-0.01	-0.22
28	7.47*	1.19	13.40	0.24	0.04	1.39	0.04*	0.88	12.20	-0.12	-0.03	-0.76
29	0.01	0.99	14.20	0.06	-0.00	-0.05	0.46	0.98	11.80	-0.51	-0.01	-0.28
30	0.04	0.99	12.60	0.03	-0.00	-0.10	0.03*	1.18	9.80	-0.34	0.03	0.69
31	1.54	0.93	13.00	0.04	-0.02	-0.64	0.50	1.04	12.20	0.48	0.01	0.26
32	0.48	0.96	13.00	0.04	-0.01	-0.36	0.08	1.14	15.40	2.85*	0.02	0.59
33	0.95	1.06	12.60	0.03	0.02	0.50	0.86	1.01	13.00	0.49	0.00	0.07
34	7.93*	1.20	13.40	0.24	0.04	1.44	0.71	0.98	13.00	0.09	-0.01	-0.14
35	0.10	1.02	14.20	0.06	0.01	0.16	0.56	1.04	13.00	0.29	0.01	0.23
36	6.62*	1.20	15.00	0.27	0.04	1.18	0.24	0.93	13.00	0.29	-0.02	-0.46
37	2.55	0.91	13.00	-0.16	-0.03	-0.82	0.19	0.92	14.20	-0.68	-0.02	-0.50
38	0.75	0.95	13.80	0.05	-0.01	-0.44	0.41	1.05	13.80	-1.27*	0.01	0.31
39	0.63	0.95	14.20	0.06	-0.01	-0.39	0.03*	0.86	15.00	-0.47	-0.03	-0.81
40	0.48	0.95	11.40	-0.18	-0.01	-0.33	0.47	1.05	11.40	0.07	0.01	0.27
41	15.64*	0.78	13.00	-0.16	-0.06	-2.05*	0.90	1.01	13.00	0.29	0.00	0.05
42	2.67	0.90	13.00	-0.16	-0.03	-0.85	0.00*	0.71	13.80	0.50	-0.08	-2.12*
43	0.91	1.08	15.80	0.28	0.01	0.41	0.68	0.97	12.20	-0.31	-0.01	-0.16
44	0.01	1.01	17.00	0.10	0.00	0.037	0.73	1.02	13.40	-0.10	0.01	0.13
45	0.88	0.94	13.40	0.04	-0.02	-0.48	0.07	0.89	13.00	-0.11	-0.03	-0.72
46	1.20	1.07	13.40	0.24	0.02	0.56	0.25	0.87	13.40	0.29	-0.04	-0.87
47	3.89	0.88	14.60	0.06	-0.03	-0.96	0.67	1.03	16.20	-0.06	0.01	0.13
48	0.10	0.98	15.00	0.07	-0.00	-0.14	0.10	1.11	14.20	0.31	0.03	0.62
49	16.21*	0.78	13.00	-0.16	-0.06	-2.09*	0.45	0.95	14.2	0.11	-0.01	-0.28
50	10.98*	0.81	13.00	-0.16	-0.05	-1.71	0.30	0.93	15.8	-0.26	-0.01	-0.35

**Item reveals DIF ($p \leq 0.05$)*

Tables 1 provided information statistical analyses in respect of the calculated Chi-square, Transformed Item Difficulty and b-parameter values of 50 multiple-choice Mathematics and English Language items in terms of sex at 0.05 level of significance. The chi-square range of odds ratios signifies DIF in favour of males were from 1.12 to 1.35, whereas the range of odds ratio signifies DIF in favour of females were from 0.78 to 0.84. For Mathematics, this revealed that 15 items or 30% of the fifty items flaged DIF (3, 4, 5, 8, 13, 23, 25, 27, 28, 34, and 36 were in favour of males and items 17, 41, 49, 50 were in favour of females) and The Chi-square procedure flagged nine or 18% of the fifty items as indicating DIF (item 30 in favour of males and items 2, 14, 15, 17, 24, 28, 39, 42 were in favour of females) for English Language. For item to flay DIF in TID method Di values must be in excess of ± one unit. However, from Table 1 none of the Di values is in excess of one, hence, no item display DIF with respect to Mathematics, while 2 items or 4% of the fifty items as indicating DIF (item 38 in favour of male and item 32 in favour of female) with respect to English Language. For item to show DIF using b-parameter, the positive value of the difference indicate DIF favouring the males group, whereas a negative value of difference indicates DIF favouring the females group. Hence, a significant value of d greater than or equal to 1.96 indicates DIF favouring male students at 0.05 level, whereas, a significant value of d less than or equal -.1.96 indicates DIF favouring females students at the 0.05 level of significant. In Mathematics, four or 8% of the 50 items revealed DIF (the items: 5, 27 were in favour of male and the items: 41, 49 were in favour of females), while 1 item or 2% of the 50 items revealed DIF (i.e. items 42) in English Language.

Table 2: Percentage of DIF items Identified by each method based on sex for Mathematics and English Language.

Sex	DIF Methods (for Mathematics)			DIF Methods (for English Language)		
	χ^2	TID	b-parameter	χ^2	TID	b-parameter
Male	11 (22%)	0 (0%)	2 (4%)	1(2%)	1 (2%)	1 (2%)
Female	4 (8%)	0 (0%)	2 (4%)	8(16%)	1 (2%)	0 (0%)
Total	15 (30%)	0 (0%)	4 (8%)	19 (18%)	2(4%)	1 (2%)

( ) percentage

Table 2 showed the number of items that flagged DIF in Mathematics and English Language with respect to sex (male and female). Out of the three methods of detecting DIF, the chi-square method identified 15 items or 30% of the fifty items as showing DIF [11 items (22%) in favour of male and 4 items (8%) in favour of female] in Mathematics examination. The chi-square method also identified 9 items or 18% of the fift items as showing DIF [1 item (2%) in favour of students in the male students and 8 items (16%) in favour of female students] in English Language examinations. It can therefore be concluded that the Chi-square method is more effective than TID and b-parameter in detecting DIF items in Mathematics and English Language examinations items base on sex.

Research Question 2: Which of the three methods of Chi-square, Transformed Item Difficulty (TID) and b-parameter is most effective in detecting DIF in OSQ Mathematics and English Language examinations in terms of school location (rural/urban)?

In order to answer this question, the three methods were statistically applied on the 50 multiple-choice Mathematics and English Language items in order to identify the items that flag DIF based on school location. The summary of the statistical analyses of the three methods are presented in Tables 3 and 4.

Table 3:Summary of Results from the Chi-square, TID, and b-parameter Methods of Identifying Differential Item Functioning on the OSQ Mathematics and English Language Examinations Based on School Location

Items	Mathematics						English Language					
	Chi-Square		TID		B-parameter		Chi-Square		TID		B-parameter	
	Sig.	Odds Ratio	Δ	D_i	$\Delta b = F_F - b_R$	d Statistics	Sig.	Odds Ratio	Δ	D_i	$\Delta b = F_F - b_R$	d Statistics
1	0.06	0.88	11.00	0.03	0.67	0.31	0.14	0.98	11.40	-0.39	0.01	0.14
2	0.00	1.52	13.80	-0.10	-2.42*	-0.94	4.69	1.19	16.20	0.39	-0.03	-0.68
3	0.17	0.92	14.20	0.02	0.52	-0.47	47.28	1.55	12.20	0.79	-0.11	-2.69*
4	0.00	0.71	12.60	0.08	2.02*	-0.24	25.56	1.38	12.60	0.39	-0.08	-1.99*
5	0.00*	0.70	12.60	0.09	2.16*	0.40	1.22	0.93	13.40	0.00	0.02	0.43
6	0.06	0.88	11.00	0.03	0.66	0.42	105.52	0.48	11.80	-1.58	0.15	3.74*
7	0.22	0.08	11.80	-0.02	-0.47	-0.04	26.95	0.68	11.00	-0.79	0.07	1.79
8	0.00*	0.61	13.80	0.12	3.06	-0.46	29.67	0.67	11.00	-0.79	0.08	1.91
9	0.00	0.80	12.20	0.05	1.32	-0.36	64.67	1.67	13.00	0.79	-0.13	-3.15*
10	0.00	0.68	16.20	0.07	1.76	-0.24	0.94	0.93	10.60	-0.39	0.01	0.33
11	0.03	1.15	14.20	-0.03	-0.82	0.64	2.36	0.67	15.40	-0.79	0.08	2.02*
12	0.00	1.24	14.60	-0.05	-1.18	0.36	3.72	0.82	11.00	-0.39	0.04	0.99
13	0.00*	1.24	14.20	-0.05	-1.18	0.01	47.86	0.60	11.00	-1.18	0.10	2.37*
14	0.72	1.03	15.80	-0.01	-0.11	-0.78	1.02	1.07	11.80	0.00	-0.02	-0.39
15	0.25	0.92	11.40	0.02	0.42	-0.60	3.39	0.61	17.00	-1.18	0.07	1.81
16	0.00*	0.80	12.60	0.05	1.34	-0.05	2.72	0.83	11.40	-0.39	0.04	1.01
17	0.82	1.02	13.80	-0.00	-0.09	-0.94	3.98	0.56	13.80	-1.18	0.15	3.64*
18	0.00	0.75	13.00	0.07	1.75	-0.12	1.46	0.74	11.00	-0.79	0.06	1.40
19	0.00	0.83	12.60	0.04	1.11	0.28	3.12	1.12	11.80	-0.79	-0.03	-0.67
20	0.04	1.14	11.80	-0.03	-0.78	0.45	1.73	0.42	15.40	-1.97	0.20	4.91*
21	0.00*	1.84	16.20	-0.08	-2.02	0.60	0.76	0.64	13.00	-0.79	0.11	2.77*
22	0.08	0.89	12.20	0.03	0.68	-0.66	2.44	0.84	12.60	-0.39	0.04	1.07
23	0.01	1.18	13.80	-0.04	-0.97	-0.15	1.49	0.80	13.80	-0.79	0.06	1.38
24	0.00	0.43	14.60	0.20	5.06*	-1.07	0.69	0.95	13.40	-0.39	0.01	0.33
25	0.00*	1.42	13.00	-0.08	-2.12*	-0.09	1.18	0.45	11.80	-1.97	0.16	3.89*
26	0.24	1.08	14.20	-0.02	-0.45	-0.44	0.63	0.45	12.60	-1.97	0.18	4.53*
27	0.00	0.69	13.00	0.09	2.27*	-0.22	4.75	0.61	11.40	-1.18	0.10	2.49*
28	0.00	1.31	13.40	-0.06	-1.60	-0.76	0.03	0.74	12.20	-0.79	0.07	1.69
29	0.42	1.06	14.20	-0.01	-0.30	-0.28	3.62	0.52	12.60	-1.58	0.15	3.67*
30	0.00*	0.70	13.00	0.09	2.21	0.69	3.11	0.56	10.60	-1.18	0.10	2.38*
31	0.00	0.55	13.40	0.15	3.75*	0.26	1.86	1.22	12.20	0.39	-0.05	-1.22
32	0.01	0.85	13.00	0.04	1.01	0.59	4.95	0.85	10.20	5.12	0.03	0.74
33	0.00	1.72	12.20	-0.13	-3.33*	0.07	39.24	1.49	12.60	0.39	-0.10	-2.46*
34	0.00	0.69	14.20	0.09	2.27*	-0.14	2.20	0.91	13.00	-0.39	0.02	0.59
35	0.00	0.62	15.00	0.11	2.81*	0.23	0.36	0.96	13.00	0.00	0.01	0.24
36	0.00	1.59	15.00	-0.09	-2.20*	-0.46	1.24	0.93	13.00	0,00	0.02	0.44
37	0.00*	0.57	13.00	0.14	3.42*	-0.50	190.85	0.41	15.00	-1.97	0.21	5.30*
38	0.00*	1.43	13.40	-0.08	-2.11*	0.31	400.97	0.27	15.40	-3.15	0.31	7.73*
39	0.00	0.23	15.80	0.33	8.35*	-0.81	2.58	0.27	15.40	-1.58	0.17	4.14*
40	0.00	0.40	11.80	0.18	4.54*	0.27	7.75	0.83	11.80	-0.39	0.04	1.04
41	0.00	0.39	13.80	0.23	5.80*	0.05	0.20	1.03	13.00	0.00	-0.01	-0.18
42	0.00	0.59	13.40	0.13	3.32*	-2.12*	8.87	1.21	13.40	0.39	-0.05	-1.17
43	0.03*	0.84	16.20	0.03	0.72	-0.16	2.93	0.65	12.60	-1.18	0.10	2.51*
44	0.40	1.08	17.00	-0.01	-0.23	0.13	3.17	0.64	13.80	-0.79	0.11	2.75*
45	0.10	0.90	13.40	0.03	0.65	-0.72	49.53	0.64	13.00	-0.79	0.11	2.77*
46	0.38	1.06	13.40	-0.01	-0.34	-0.87	0.10	1.02	13.00	0.00	-0.01	-0.13
47	0.00*	0.69	15.00	0.08	2.06*	0.13	39.33	0.61	16.60	-0.79	0.08	1.97*
48	0.00*	0.52	15.80	0.13	3.30*	0.62	6.17	0.85	14.20	0.00	0.04	0.94
49	0.00*	0.37	13.80	0.24	6.01*	-0.28	3.13	0.89	14.60	-0.39	0.03	0.67
50	0.00*	0.47	13.00	0.18	4.54*	-0.35	60.64	0.57	16.20	-1.18	0.10	2.60*

**Item reveals DIF* ($p \leq 0.05$)

Tables 3 provided information statistical analyses in respect of the calculated Chi-square, Transformed Item Difficulty and b-parameter values of 50 multiple-choice Mathematics and English Language items in terms of school location (rural and urban) at 0.05 level of significance. For Mathematics, the Chi-square method flagged 11 items or 28% of the fifty items as indicating DIF (13, 21, 25, 30, 38, were in favour of urban students and items 5, 8, 16, 37, 43, 47, 48, 49, 50 were in favour of rural students). The Chi-square method flagged 38 items or 76% of the fifty items as indicating DIF (2, 3, 4, 9, 31, 33, 42 were in favour of urban students and items 6, 7, 8, 11, 12, 13, 15, 16, 17, 18, 20, 21, 22, 23, 25, 26, 27, 28, 29, 30, 32, 37, 38, 39, 40, 43, 44, 45, 47, 48, 50 were in favour of rural students) for English Language. Using TID five or 10% of the fifty items as indicating DIF (7, 39, 40, 41, 49 were in favour of rural students and none of the item in favour of urban students) with respect to Mathematics, while 15 or 30% of the fifty items as indicating DIF (item 32 in favour of rural students and items 6, 15, 17, 20, 25, 26, 27, 29, 30, 37, 38, 39, 43, 50 were in favour of urban students) with respect to English Language. In Mathematics, using b-parameter, 21 or 42% of the 50 items revealed DIF (the items: 4, 5, 24, 27, 31, 34, 35, 37, 39, 40, 41, 42, 47, 48, 49, 50 were in favour of urban students and the items: 2, 25, 33, 36, 38 were in favour of rural students) while Twenty two or 44% of the 50 items revealed DIF (the items: 6, 11, 13, 17, 20, 21, 22, 25, 26, 27, 29, 30, 37, 38, 39, 43, 44, 45, 47, 50 were in favour of urban students and the items: 3, 4, 9, 33 were in favour of rural students) in English Language.

Table 4: Percentage of DIF items Identified by each method based on School Location for Mathematics and English Language

School Location	DIF Methods			DIF Methods		
	χ^2	TID	b-parameter	χ^2	TID	b-parameter
Urban	5(10%)	0 (0%)	16 (32%)	7(14%)	4(8%)	18 (36%)
Rural	9(18%)	5 (10%)	5 (10%)	31(62%)	1(2%)	4 (8%)
Total	14 (28%)	5 (10%)	21 (42%)	38 (76%)	5 (10%)	22 (44%)

( ) percentage

Table 4 showed the number of items that flagged DIF in Mathematics and English Language with respect to school location (urban and rural). Out of the three methods of detecting DIF, b-parameter identified 21 items or 42% of the fifty items as showing DIF [16 items (32%) in favour of students in the urban location and 5 items (10%) in favour of students in the rural location] in Mathematics examination. The chi-square method identified 9 items or 18% of the fifty items as showing DIF: 1 item (2%) in favour of students in the male students and 8 items (16%) in favour of female students in English Language examinations. It can therefore be concluded that the Chi-square method is more effective than TID and b-parameter in detecting DIF items in Mathematics and English Language examinations items base on school location.

Research Hypothesis 1: There is no significant difference in the relative ability of Chi-square, Transformed Item Difficulty (TID) and b-parameter in detecting DIF in OSQ examination.

To test this research hypothesis, One-way Analysis of Variance was carried.

Table 5: ANOVA showing the difference in the relative ability of the three methods of detecting DIF in the examination

Source	Sum of Square	df	Mean Square	F	p
Between Group	1029.33	2	514.67	9.301	<. 05
Within Group	166.00	3	55.33		
Total	1195.33	5			

.

The result presented in Table 5 showed that the difference in the relative ability of the three methods of detecting DIF in the OSQ examination is significant ($F = 9.301$, df1 = 2 df2 =3, $p<.05$). Hence, it can be concluded that there was significant difference in the relative ability of Chi-square, Transformed Item Difficulty (TID) and b-parameter in detecting DIF in OSQ examination.

8. Discussion

The results revealed that some items flagged DIF in favour of male and female with varying percentage across the three methods. The chi-square method identified more items as showing DIF than TID and b-parameter in terms of sex.This is in agreement with the findings of Nabeel (2010). This result therefore suggests that B-parameter method was more effective in detecting DIF with respect to school location. This result gives greater confidence in asserting the effectiveness of Chi-square method in detecting DIF with reference to sex. This is in agreement with the findings of Andrea, Karen and Launa (2006). The chi-square method was also found to be more effective in detecting items that flaged DIF on the basis of school location. Finally, the study established that there was significant difference in the relative ability of Chi-square, Transformed Item Difficulty (TID) and b-parameter in detecting DIF in OSQ examination.

9. Conclusion and Recommendations

The study concluded that the Chi-square method demonstrated greater statistical power in detecting DIF items than TID and B-parameter in detecting DIF in Osun State Qualifying Examination.

DIF analysis should be used during trial testing of achievement tests by our national examination bodies such as the West African Examination Council (WAEC), National Examination Council (NECO), National Business and Technical Examination Board NABTEB), and Joint Admission and Matriculation Board (JAMB) to overcome consistent measurement error in testing.

References

Andrea, Gotzmann., Karen, Wright., & Launa, Rodden. (2006). A Comparison of Power Rates for Items Favouring the Reference and Focal group for the Mantel-Haenszel and SIBTEST Procedures. Paper presented at *American Educational Research Association* (AERA) in SanFrancisco, California. 1-36.

Angoff, W. (1993). Perspective on differential item functioning methodology. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 3–24). Hillsdale, NJ: Lawrence Erlbaum Associates.

Gay, L.R., Mills, G.E., & Airasian, P. (2006). *Educational Research:* Competencies for Analysis and Applications (8^{th} ed.). Upper Saddle River, NJ: Pearson Education, Inc.

Hambleton, R.K., Swanninethan, H. & Rodgers, H.J. (1991). Fundamentals of Item Response Theory. Newbury: Sage.

Nabeel, A. (2010). A Gender-Related Differential Item Functioning of Mathematics Test Items. *The International Journal of Educational and Psychological Assessment*, 5, 101-116.

Subvokiak, M., Mack, J., Ironson, G., & Craig, R. (1987). Empirical comparison of selected item bias detection procedures with bias manipulation. *Journal of Educational Measurement, 21*, 49-58.