Examining Differential Item Functioning in a Chemistry Achievement Test for Students in Nigeria

***Memory Queensoap (PhD)**
Department of Science Foundations,
School of Foundation Studies,
Bayelsa State College of Health Technology,
Otuogidi, Ogbia-Town, Yenagoa
*mqueensoap@gmail.com

Goodness W. Orluwene (PhD)
Department of Educational Psychology, Guidance & Counselling,
Faculty of Education,
University of Port Harcourt,
Choba, Port Harcourt
Rivers State

Abstract
This study used Differential Item Functioning (DIF) approach to detect item bias in Chemistry Achievement Test (CAT). The CAT was administered to 400 students of senior secondary III in Niger State, Kwara State, Anambra State and Bayelsa State. A descriptive-comparative research design was used to describe and compare examinees of the four ethnic groups. A statistical and content analysis was done with logistic regression statistics. The logistic regression statistical indices revealed 46 items exhibiting biases between the focal group (Ijaw) and reference groups (Hausa, Igbo and Yoruba). Based on the analysis, the study established that the CAT has some items that showed biases and not all items that exhibited differential functioning were flagged biased. Therefore, the study concludes that there was an ethnic bias in the CAT. The study recommends, among others, that Item Response Theory should be used in educational testing.

Keywords: *Differential Item Functioning (DIF), logistic regression statistics (LRS), Ethnicity, Item bias.*

1. Introduction

To avoid test or item bias, and achieve test fairness, item analysis methods should be used to design reliable, valid and usable scales or test. Wiersma and Jurs (1990) posited that item analysis helps to make better decisions about the students (test takers), the instruction, and the test items. Various methods have been designed for item analysis in either in the classical test theory (CTT) or latent trait theory (LTT).

Anastasi and Urbina (1997) emphasized that items can be analyzed qualitatively, in terms of their content and form, and quantitatively in, terms of their statistical properties. Qualitative analysis includes the consideration of content validity as well as the evaluation of items in terms of effective item writing procedures. Quantitative analysis includes principally the measurement of item difficulty and item discrimination. Nevertheless, as test bias became a sensitive concern to the community of test makers or developers, several procedures are generated to eliminate biases in test. Among, such procedures, is differential item functioning approach or method (DIF). This is a method that investigates the test items in a test, one at a time, for signs of interaction with sample characteristic. Pedrajita (2009) denotes, differential

item functioning as the probabilities of success on an item of examinees of the same ability but belonging to different groups; that is, when examinees from different groups have a different probability of answering an item correctly after controlling for overall test performance. The author further states that, DIF may be attributed to item-bias but may also reflect performance difference that the test is designed to measure. In the same vein, Zumbo (2007) identified three generation of DIF analysis from the more commonly used term as item bias to its praxis. DIF methods permit test users to judge whether items (and ultimately the test they constitute) are functioning in the same manner in various groups of examinees. DIF assumes to answer question like, is the test performing in the same manner for each group of examinees? Similarly, Oshima and Morris (2008) stated that DIF analyses are vital in the field of test and measurement because DIF sets to address equivalence across subgroups of examinees.

In Nigeria, terminal examinations are done from one level to the other especially the primary and secondary levels. These examinations are being conducted by different bodies such as West African examination council, National Examination Council, Ministry of Education etc. Different examinees of the same ability from different languages, culture, sex etc. are made to undertake these examinations. Are these test items fair enough for all groups? How will one know that such national examination or test items are not fair? When, it is widely known that Nigeria is composed of more than 250 ethnic groups. It is obvious, therefore, that most central or terminal examination being administered to Nigerian student may not be fair to one ethnic group or the other if methods that will refine test items devoid of ethnic, cultural and gender biases are not put into consideration, especially differential item functioning (DIF) methods. It was at this instance that the researcher purposely embarked on this study, to apply the differential item functioning approach to detect item bias in Chemistry Achievement Test (CAT) among the major ethnic groups in Nigeria, such as Yoruba, Igbo, Hausa, and Ijaw. For this study, Ijaw students group was considered as the focal group while the Yoruba, Hausa and Igbo students were considered reference group.

Pedrajita (2009) revealed the use of Logistic Regression analysis for the detection and testing for *DIF/*Item bias for each comparison group. The author explained Logistic Regression as "a- kind of regression analysis often used when the dependent variable is dichotomous or categorical and scored 0 or 1 (it can also be used when the dependent variable has more two categories). It is usually used for predicting whether something will happen or not - anything that can be expressed as Event/Non-Event. Independent variables may be categorical or continuous (p.57)". Logistic regression was further enunciated that it uses the logistic curve that best approximates the set of data. In other words it is based on transforming data by taking their natural logarithms so as to reduce nonlinearity.

Similarly, Swaminathan and Rogers (1990) carried out a work on detecting Differential Item Functioning using Logistic Regression Procedures. 'This study asserted use of logistic regression procedure to identify non-uniform DIF which MH procedure is not designed for and may not be powerful in detecting non-uniform DIF, remarked Swaminathan and Rogers. The paper explained uniform DIF as the probability of answering the item correctly is greater for one group than the other uniformly over all levels of ability. That is, uniform DIF exists when there is no interaction between ability level and group membership while non-uniform DIF exists when there is an interaction between ability level and group membership. This in essence is defined as the difference in the probabilities of a correct answer for the two groups is not the same at all ability levels. The authors stated that non-parallel item characteristic curves indicate the presence of non-uniform DIF with regards to item response theory. Result of the study indicated that the logistic regression procedure is more powerful than the Mantel-Haenszel procedure for detecting non-uniform DIF and as powerful in detecting uniform DIF.

For Abedi, Leon and Kao (2007) DIF analysis is often used to examine group differences between specific racial or ethnic groups or between males and female. The authors stated that "there are several statistical procedures that can be used to identify differentially functioning test items, including the Mantel-Haenszel statistic, logistic regression, SIBTEST, the Standardization procedure, and various item Response Theory-based approaches (p.3)". Nevertheless, the writers subscribed to the use of logistic regression approaches; arguing that it is easier to be employed and is more suitable for answering research questions.

2. Methods

This study adopted a descriptive-comparative research design. Four matched groups were compared in terms of their probability of success on each item of the Chemistry Achievement Test (CAT). The matched groups were Yoruba, Igbo, Hausa, and Ijaws ethnic groups drawn from Kwara State, Anambra State, Niger State and Bayelsa State in Nigeria respectively. The Instrument used for this study was titled Chemistry Achievement Test (CAT). It was adopted from Orluwene (2007). The two parameter latent trait model was used in preparing the CAT. The instrument was made up of two Sections A and B. Section A was designed to elicit personal information from the respondents such as age, gender, school type etc. Section B consisted of 60 items with 5 options (multiple choices) answers to the questions (Orluwene, 2007). The instrument was dichotomously scored, 1 or 0 and was marked over 60. The instrument was revalidated through content and face validity. It was observed that the instrument was 82.62% reliable and 17.38% unreliable which was a good indication of consistency and high precision of measurement (Orluwene, 2007). Nevertheless, a pilot survey was done on a sample size of 20, made up of all ethnic groups at the Federal Government College, Odi. After administration and retrieval of the instrument, scoring and analysis was done with Cronbach Alpha reliability determination method which gave a reliability coefficient of 0.80. Since the reliability coefficient is consistent with Orluwene (2007) the instrument was therefore considered to be reliable. Four research assistants were recruited to administer the CAT instrument to respondent-examinees in their states. Marking scheme was prepared which, was in line with test designer's answers for the CAT. The data gathering procedures involved simultaneous administration of the test to the original intact classes and checking and scoring the test. Data gathered were not made up or influenced. Logistic regression analysis was used for the detection and testing for differential item functioning/item bias for each comparison group (Queensoap, 2014).

3. Results

Table1: Summary of Logistic Regression Analysis

Items	Ethnic Groups	Logistic Regression statistics				Effect size
		Odds ratios	Regression coefficient	Chi square (x^2)	Effect size	
1	Hausa	1.011	0.011	0.00	0	Small
	Yoruba	1.029	0.028	0.00	0	Small
	Igbo	2. 126	0.754	4.41*	0.15	Small
2	Hausa	1.880	0.631	3.42	0.13	Small
	Yoruba	1.092	0.088	0.06	0.02	Small
	Igbo	4.797	0.568	22.85*	0.34	Medium
3	Hausa	0.895	-0.111	11.44*	0.24	Small

	Yoruba	0.410	-0.891	7.80*	0.20	Small
	Igbo	1.790	0.582	4.14*	0.14	Small
4	Hausa	1.650	0.501	3.11	0.12	Small
	Yoruba	0.515	-0.663	5.22*	0.16	Small
	Igbo	2.312	0.838	8.35*	0.20	Small
5	Hausa	1.686	0.522	3.20	0.14	Small
	Yoruba	0.697	-0.361	1.32	0.08	Small
	Igbo	1.555	0.441	2.26	0.11	Small
6	Hausa	0.570	-0.562	3.22	0.13	Small
	Yoruba	0.944	-0.058	1.43	0.08	Small
	Igbo	0.480	-0734	5.19*	0.16	Small
7	Hausa	0.702	-0.354	1.51	0.09	Small
	Yoruba	0.714	-0.337	1.37	0.08	Small
	Igbo	5.748	1.695	26.71*	0.36	Medium
8	Hausa	2.888	1.060	11.61*	0.24	Small
	Yoruba	0.752	-0.285	0.66	0.06	Small
	Igbo	2.556	0.938	9.04*	0.21	Small
9	Hausa	2.878	1.057	109.52*	0.74	Small
	Yoruba	0.507	0.507	2.39	0.11	Small
	Igbo	1.099	1.099	12.02*	0.24	Small
10	Hausa	21.128	0.120	0.18	0.03	Small
	Yoruba	0.523	-0.648	2.42	0.04	Small
	Igbo	1.857	0.619	0.38	0.04	Small
11	Hausa	1.128	0.666	5.38*	0.16	Small
	Yoruba	1.439	0.364	1.60	0.09	Small
	Igbo	0.892	-0.114	0.01	0.00	Small
12	Hausa	4.218	1.439	18.67*	0.30	Medium
	Yoruba	1.673	0.515	2.14	0.10	Small
	Igbo	1.564	0.447	0.20	0.03	Small
13	Hausa	3.080	1.125	13.34*	0.26	Small
	Yoruba	1.327	0.283	0.78	0.06	Small
	Igbo	3.471	1.244	16.31*	0.28	Small
14	Hausa	2.013	0.700	2.02	0.10	Small
	Yoruba	1.513	0.414	0.65	0.06	Small
	Igbo	2.378	0.866	3.17	0.12	Small
15	Hausa	0.642	-0.443	1.28	0.08	Small
	Yoruba	1.698	0.529	2.45	0.11	Small
	Igbo	1.759	0.565	2.81	0.12	Small
16	Hausa	5.952	1.784	30.13*	0.39	Medium
	Yoruba	5.388	1.684	26.84*	0.37	Medium
	Igbo	0.819	-0.200	0.29	0.04	Small
17	Hausa	2.261	0.816	7.35*	0.39	Small
	Yoruba	1.401	0.337	1.37	0.37	Small
	Igbo	1.203	0.185	0.42	0.04	Small
18	Hausa	1.836	0.607	4.29	0.15	Small
	Yoruba	0.937	-0.065	0.04	0.01	Small
	Igbo	2.334	0.847	8.36*	0.20	Small
19	Hausa	5.697	1.740	21.08*	0.32	Medium
	Yoruba	2.918	1.071	7.42*	0.19	Small
	Igbo	8.877	2.183	33.52*	0.41	Medium

20	Hausa	1.513	0.414	1.46	0.08	Small
	Yoruba	1.090	0.086	0.06	0.02	Small
	Igbo	3.121	1.138	12.26*	0.25	Small
21	Hausa	0.633	-0.457	1.53	0.09	Small
	Yoruba	0.640	-0.446	1.45	0.08	Small
	Igbo	1.852	0.616	3.70	0.13	Small
22	Hausa	1.047	0.046	0.02	0.00	Small
	Yoruba	1.441	0.365	1.17	0.08	Small
	Igbo	4.385	1.478	21.33*	0.32	Medium
23	Hausa	2.178	0.778	5.69*	0.17	Small
	Yoruba	1.368	0.313	0.85	0.06	Small
	Igbo	4.048	1.398	19.09*	0.31	Medium
24	Hausa	0.713	-0.338	0.48	0.05	Small
	Yoruba	0.918	-0.085	0.03	0.01	Small
	Igbo	2.049	0.717	3.12	0.12	Small
25	Hausa	1.092	0.088	0.05	0.01	Small
	Yoruba	1.362	0.309	0.66	0.06	Small
	Igbo	0.241	-1.423	5.98*	0.17	Small
26	Hausa	2.334	0.847	7.92*	0.20	Small
	Yoruba	1.192	0.176	0.31	0.04	Small
	Igbo	2.852	1.048	12.12*	0.24	Small
27	Hausa	1.152	0.141	0.15	0.03	Small
	Yoruba	1.475	0.389	1.23	0.09	Small
	Igbo	3.925	1.367	17.37*	0.29	Small
28	Hausa	1.165	0.153	0.16	0.03	Small
	Yoruba	2.096	0.740	4.42*	0.15	Small
	Igbo	4.351	1.470	18.80*	0.30	Small
29	Hausa	4.606	1.527	16.06*	0.28	Small
	Yoruba	1.820	0.599	2.11	0.10	Small
	Igbo	1.680	0.519	1.55	0.09	Small
30	Hausa	0.417	-0.875	6.08*	0.17	Small
	Yoruba	0.237	-1.440	12.21*	0.25	Small
	Igbo	2.019	0.703	5.64*	0.17	Small
31	Hausa	8.355	2.123	33.10*	0.40	medium
	Yoruba	1.022	0.022	0.00	0.00	Small
	Igbo	4.941	1.597	18.73*	0.30	Medium
32	Hausa	0.971	-0.029	0.00	0.00	Small
	Yoruba	0.401	-0.914	7.53*	0.12	Small
	Igbo	3.664	1.298	19.10*	0.31	Medium
33	Hausa	0.886	-0.121	0.16	0.03	Small
	Yoruba	1.031	0.030	0.01	0.00	Small
	Igbo	2.838	1.431	23.85*	0.34	Medium
34	Hausa	0.592	-0.524	2.50	0.11	Small
	Yoruba	1.185	0.170	0.31	0.04	Small
	Igbo	3.016	1.104	13.82*	0.26	Small
35	Hausa	2.773	1.020	10.76*	0.23	Small
	Yoruba	1.085	0.081	0.06	0.02	Small
	Igbo	2.454	0.898	8.28*	0.20	Small
36	Hausa	1.968	0.677	3.68	0.13	Small
	Yoruba	1.431	0.358	0.13	0.02	Small

	Igbo	2.739	1.007	1.01	0.07	Small
37	Hausa	2.135	0.758	4.02*	0.14	Small
	Yoruba	2.405	0.877	5.47*	0.16	Small
	Igbo	4.317	1.462	16.40*	0.28	Small
38	Hausa	1.564	0.447	1.59	0.09	Small
	Yoruba	0.554	-0.590	1.92	0.10	Small
	Igbo	1.480	0.392	1.20	0.08	Small
39	Hausa	0.792	-0.233	0.33	0.04	Small
	Yoruba	0.878	-0.130	0.11	0.02	Small
	Igbo	0.723	-0.324	0.62	0.05	Small
40	Hausa	0.660	-0.415	1.18	0.08	Small
	Yoruba	0.405	-0.904	4.42*	0.15	Small
	Igbo	0.551	-0.596	2.26	0.11	Small
41	Hausa	5.605	1.724	30.14*	0.38	medium
	Yoruba	1.797	0.586	3.40	0.13	Small
	Igbo	4.945	1.598	26.23*	0.36	Medium
42	Hausa	1.360	0.307	1.14	0.07	Small
	Yoruba	0.988	-0.012	0.00	0.00	Small
	Igbo	1.474	0.388	0.15	0.03	Small
43	Hausa	1.935	0.660	4.28*	0.14	Small
	Yoruba	0.798	-0.226	0.40	0.04	Small
	Igbo	2.399	0.875	7.67*	0.19	Small
44	Hausa	2.236	0.805	5.06*	0.16	Small
	Yoruba	0.793	-0.232	0.31	0.04	Small
	Igbo	1.345	0.296	0.61	0.05	Small
45	Hausa	0.601	-0.509	2.68	0.11	Small
	Yoruba	0.711	0-0.341	1.25	0.08	Small
	Igbo	1.361	0.308	1.43	0.07	Small
46	Hausa	5.779	1.754	22.71*	0.34	medium
	Yoruba	1.860	0.620	2.43	0.11	Small
	Igbo	5.323	1.672	20.63*	0.32	Medium
47	Hausa	2.519	0.924	10.01*	0.22	Small
	Yoruba	1.078	0.075	0.06	0.02	Small
	Igbo	2.519	0.924	10.01*	0.22	Small
48	Hausa	1.771	0.571	2.56	0.11	Small
	Yoruba	0.878	-0.130	0.11	0.02	Small
	Igbo	4.718	1.551	21.06*	0.32	Medium
49	Hausa	1.360	0.307	0.62	0.05	Small
	Yoruba	2.211	0.793	4.67*	0.15	Small
	Igbo	3.969	1.378	15.24*	0.27	Small
50	Hausa	2.996	1.097	11.32*	0.24	Small
	Yoruba	3.047	1.114	11.67*	0.24	Small
	Igbo	4.862	1.581	23.81*	0.34	Medium
51	Hausa	2.105	0.744	6.721*	0.18	medium
	Yoruba	2.555	0.938	10.39*	0.23	Small
	Igbo	5.404	1.687	28.50*	0.38	Medium
52	Hausa	4.285	1.455	22.90*	0.34	Small
	Yoruba	1.433	0.251	0.66	0.06	Small
	Igbo	2.852	1.048	12.12*	0.24	Small
53	Hausa	0.630	-0.462	1.98	0.10	Small

	Yoruba	1.133	0.125	0.17	0.03	Small
	Igbo	0.832	-0.184	0.34	0.04	Small
54	Hausa	2.827	1.039	10.95*	0.23	Small
	Yoruba	2.338	0.849	7.17*	0.19	Small
	Igbo	3.597	1.280	16.72*	0.29	Small
55	Hausa	1.770	0.571	3.16	0.12	Small
	Yoruba	1.216	0.195	0.34	0.04	Small
	Igbo	2.712	0.998	10.10*	0.22	Small
56	Hausa	1.133	0.125	0.07	0.02	Small
	Yoruba	1.010	0.010	0.00	0.00	Small
	Igbo	0.653	-0.426	0.61	0.05	Small
57	Hausa	7.111	1.963	28.61	0.38	Small
	Yoruba	1.318	0.276	0.44	0.05	Small
	Igbo	8.696	2.163	34.74*	0.41	Small
58	Hausa	3.302	1.194	12.70*	0.25	Small
	Yoruba	1.252	0.225	0.38	0.04	Small
	Igbo	6.049	1.800	29.21*	0.38	Medium
59	Hausa	7.034	1.951	29.70*	0.38	medium
	Yoruba	1.709	0.536	1.91	0.10	Small
	Igbo	7.034	1.951	29.70*	0.38	Medium
60	Hausa	1.534	0.428	1.51	0.09	Small
	Yoruba	1.092	0.088	0.06	0.02	Small
	Igbo	1.299	0.261	0.54	0.05	Small

*Critical X^2 =3.84, *Means Calculated X^2 was significant at .05 alpha level.*

Table 1 identified 46 biased items against the focal group (Ijaw examinees). The measure of bias is the significance of the chi-square value. A significance of chi-square value indicates that the odd of getting an item right were different for the reference groups coded 1 (Hausa), 2 (Yoruba) and 3 (Igbo) against the focal group (Ijaw). The biased items include; 1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 13, 16, 17, 18, 19, 20, 22, 23, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 37, 40, 41, 43, 44, 46, 47, 48, 49, 50, 51, 52, 54, 55, 57, 58 and 59 while the items that do not show significant difference were 5,10,14,15,21,24,36,38,39,42,45,53, 56 and 60. Of the forty-six biased items four (11, 12, 17, and 44) were biased items between Hausa and Ijaw alone, one was biased items between Yoruba and Ijaw alone, while ten (1, 2, 6, 7, 20, 22, 25, 27, 48, and 55) were biased items between Igbo and Ijaw student examinees. Also, the Table 1 indicated that item 3, 30, 37, 50, 51 and 54 were biased items between the three reference groups and the focal groups, that is 10% of the items showed significant difference between Ijaw and all other reference groups, 1% was Ijaw versus Yoruba, 11% was Ijaw/Igb0,7% was Ijaw and Hausa,42% was Ijaw and, either Hausa or Igbo or Yoruba while 23% showed no significant difference between the focal group and all other reference groups.

Table 1 identified also the effect size of the significant chi-square value, of the forty-six biased items, there was one large effect size (item 9 = 0.74) and twenty medium effect sizes (item 2, 7, 12, 16, 19, 22, 23, 28, 31, 32, 33, 41, 46, 48, 50, 52, 57, 58, 59), while the other items indicated a small size effect of biases.

4. Discussion

The result of this research question showed that there are 46 items that were detected showing item bias. The logistic regression statistics revealed that 46 items calculated Wald chi-square X^2 was greater than critical X^2 (3.84), df_1 at 0.05 alpha level of significance, indicating that the odds of getting an item right were different for the focal/reference group

compared (check values, table 2). The logistic statistics revealed that 10% of the 60 items showed item bias between the focal group and all the reference groups, 1% was focal group against Yoruba as reference group, 17% was Ijaw/Igbo, 7% was Ijaw versus Hausa, 42% was Ijaw versus either Hausa or Yoruba or Igbo while 23% showed significance difference between the focal group (Ijaw) and reference groups (Hausa, Yoruba & Igbo). Similarly, Pedrajita (2009) detected biased test items with the logistic regression method. In his study, 22 biased items were identified between the public and the private examinees. Also 7 items were detected biased between male and female students. Pedrajita concluded that "the two groups had not had equal opportunity to learning experience related to the content of the biased items (p.67)". This finding supported the finding of this research work, the focal group might not had equal opportunity to learning experience related to the content of the biased item or must have been influenced by the factors like language, poor calculation ability, omission and wrong use of units (WAEC, 2004 & Adedoyin, 2010).

Zumbo (2007 & 1999), Oshima and Morris (2008), Cohen (2006) and others had identified logistic regression method as a superior tool in analyzing items for differential functioning. Cohen (2006) rightly observed that the terms item bias and differential item functioning are interchangeably used. Schumacker (2005) emphasized a borderline difference between item bias and differential item functioning, establishing that an item may be flagged DIF but may not be biased item while an item that is flagged biased item always show differential item functioning. These conceptual frameworks proved that the 46 items flagged as biased can be likewise termed items showing differential item functioning. Because the same method used in detecting item bias is also plausible for making DIF analysis. However, Zumbo (1999) raised a salient but very important DIF operational policy issue that how much DIF do one need to see before one is willing to consider the item as displaying DIF?

The researcher treated this issue by using the Cramer's phi coefficient to know how much significant was the significance? The findings revealed that there was only one item that shows large effect size, twenty items were observed with medium effect while twenty-five showed small effect size. That is to say, 21 items can be flagged as actually displaying DIF/ item bias, since small effect is negligible (Coolican, 2009).

5. Conclusion

Based on the result of the study, the following conclusions were drawn;

- Using the logistic regression statistic as a differential item functioning approach to detect biased items on the 60 chemistry achievement test, 46 items were identified having a significant Chi-square value, indicating that the odds of getting an item right were different for the reference/focal groups compared. Hence, there was ethnic bias in the CAT items.
- Among the 46 items that showed significant chi-square value, it was found that one large effect size item, twenty (20) medium size effect and twenty-five (25) small size effects were identified with Cramer phi relation showing degree of the significance. There were varying degrees of biases among matching groups.
- It was found that, 10% of the item showed significance difference between Ijaws and other ethnic groups, 1% was Ijaw versus Yoruba, 17% was Ijaw/Igbo, 7% was Ijaw versus Hausa, 42% was Ijaw versus either Hausa or Yoruba or Igbo while 23% showed no significance difference between focal group (Ijaw) and reference groups (Hausa, Yoruba and Igbo). Ethnic difference was more apparent between Ijaw and Igbo examinees while Ijaw and Yoruba showed least ethnic difference.

- The study has made it clear that the Chemistry Achievement Test that was developed by Orluwene (2007) with Item Response Theory was successfully analyzed using DIF/IRT. So, DIF/IRT methods are appropriate in developing achievement test.

6. Recommendation

We put forward the following recommendation,

- Test experts and developers should explore the use of item bias detection method, particularly the logistic regression in DIF analysis of achievement test.
- Since IRT theoretical framework is more of identifying the latent trait of an examinee IRT should be incorporated by examination Boards into educational test and measurement practices in Nigeria.

References

Abedi, J., Leon, S., & Kao, J. (2007). *Examining differential item functioning in reading assessments for students with disabilities*. Minneapolis, MN: University of Minnesota, Partnership for Accessible Reading Assessment.

Adedoyin, O. O. (2010). Using IRT approach to detect gender biased items in public examinations: A case study from the Botswana junior certificate examination in Mathematics. *Educational Research and Review, 5(7), 385-399.*

Anastasi, A. & Urbina, S. (1997). *Psychological Testing*. India: Prentice – Hall.

Cohen, L.A. (2006). Item bias and differential item functioning. In T.M. Wood & W. Zhu (Eds), *Measurement theory and practice in kinesiology.*

Coolican, H. (2009). *Research methods and statistics in psychology*. Great Britain: Hodder Education.

Orluwene, G.W. (2007). Application of *two-parameter Latent trait model in development and validation of chemistry achievement test.*Unpublished Ph.D Dissertation, University of Port – Harcourt.

Oshima, T.C & Morris, S.B.(2008). An NCME instructional module on Raju's differential functioning of items and test (DFIT). *Educational measurement: issues and practice*. 43-50.

Pedrajita, J.Q. (2009). Using Logistic regression to detect biased test items. *The Internationals Journal of Educational and Psychological Assessment 2, 54-73.*

Queensoap, M. (2014). A*pplication of differential item functioning in detecting item bias in chemistry achievement test in Nigeria. U*npublished doctoral dissertation, University of Port-Harcourt, Port-Harcourt, Rivers State, Nigeria.

Schumacker, R.E. (2005). Test bias and differential item functioning. Retrieved October 11, 2010, from *http://www.appliedmeasurementassociates.com,pdf.*

Swaminathan, H., & Rogers, H.J. (1990).Detecting differential item functioning using logistic regression procedures. *Journal of educational measurement, 27,361-370.*

West African Examination Council (2004). *The examinations May/June 2004 Chief Examiners' reports (Nigeria).*

Wiersma, W. & Jurs, S. G. (1990).*Educational measurement and testing* (2nd ed.). Massachusetts: Allyn and Bacon.

Zumbo, B.D. (1999). *A handbook on the theory and methods of deferential item functioning (DIF):logistic regression modeling as a unitary frame work for binary and likert-type (ordinal) item scores*. Canada: Directorate of Human Resources Research and Evaluation.

Zumbo, B.D. (2007). Three generations of DIF analyses: considering where it has been, where it is now, and where it is going. *Language Assessment Quarterly. 4 (2), 223-233.*