RESEARCH ARTICLE

USE OF MANTEL HAENSZEL DIFFERENTIAL ITEM FUNCTIONING IN DETECTING ITEM BIAS IN A CHEMISTRY ACHIEVEMENT TEST IN FOUR ETHNIC GROUPS IN NIGERIA

*,1Memory Queensoap and 2Goodness W. Orluwene

1Department of Arts Education, Faculty of Education, Federal University Otueke, Yenagoa, Bayelsa State
2Department of Educational Psychology, Guidance and Counselling, Faculty of Education, University of Port Harcourt, Choba, Port Harcourt, Rivers State

ABSTRACT

This study used mantel haenszel Differential Item Functioning (DIF) approach to detect item bias in a Chemistry Achievement Test (CAT) in four ethnic groups in Nigeria. The CAT was administered to 400 chemistry students of senior secondary III in Niger State, Kwara State, Anambra State and Bayelsa State representing Hausa, Yoruba, Igbo and Ijaw respectively. 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 Mantel Haenzel's adjusted – DIF.The Mantel-Haenszel adjusted-DIF statistics flagged all items as showing significant differential item functioning between the focal group (Ijaw) and reference group (Yoruba, Hausa & Igbo). It was also observed that item 20 and 36 showed NO DIF for Ijaw/Yoruba and Ijaw/Hausa respectively. 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 because some items fall within category A, (small effect) following the Educational Test Services (ETS) scale. Therefore, the study concludes that there was an ethnic bias in the CAT. The study recommends, among others, that Item Response Theory based DIF approaches should be used in educational testing.

INTRODUCTION

Generally, test is to assess the degree to which the terminal behavior of students differs from their initial behaviors. As such test is applied in every sphere of human life-spiritual, social, mental, physical, educational, political etc. This unavoidable quality of test has generated many Test-Makers or Developers to give seemingly insoluble definitions of test. Consequently, test has been categorized into teacher made test and standardized test. The former denotes test developed by the classroom teacher, policy makers, administrators or non-test experts while the latter category of test which can also be known as psychological test are made by test experts or psychometricians. Nevertheless, of who the test maker is, tests are constructed or designed to make important decision about a person, business, or society. This study boarder on the differential item functioning of test administered to students or test users which can detect whether test items are bias or not. Psychological test, which according to the American Psychological Association and National Council on Measurement in Onunkwo (2002), refers to a set of task or questions intended to elicit particular types of behaviors when

presented under standardized conditions and to yield scores that have desirable psychometric properties. Psychological tests are tools which are used to detect or identify some qualities, traits, attributes etc possessed by a person, an object or a thing. These tests are expected to measure what they are designed for. They should be fair to all test takers. In recent years the issue of test fairness has gained considerable attention. According to Roever (2005), a fair test is one that is comparably valid for all groups and individuals, and 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. The basic concern of assessment in Multilanguage nation like Nigeria is whether student's performance on test items is comparable when items are adapted to different languages. Infact, Federal Government of Nigeria (FGN) (2004) emphasized all national examinations to be as valid and fair as possible to all students. It is also true that the nation's education watchdog stipulated a must be equal education at the secondary school education. And a valid test should be free from biases either ethnic, gender or what have you. 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 foritem analysis in either in the classical test theory (CTT) or latent trait theory (LTT). 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 permits 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? Differential item functioning (DIF) detection methods are widely used for selection of potentially biased items.

The typical detection framework of two groups, a reference group and a focal group was used against which the studied item is suspected of being biased and the reference group provides the bases for comparison. In an ethnic DIF study, like in this work, the sample of the Ijaws is specified as the focal group and the other ethnic group samples serve as the reference group. Many statistical methods have been employed in the detection of test items that function differently among subgroups in a given sample of study. According to Brown (2012) there are two major statistical techniques used in analyzing DIF which are non-parametric and parametric statistics. The non-parametric statistics tools are Mantel-Haenszel statistics, Test Graf (non-parametric IRT) and Simultaneous Item Bias Test while the parametric DIF statistical techniques include Logistic regression, Item Response Theory methods and Structural Equation Modelling. Brown further stated that there are three pieces of information necessary for DIF analysis. These pieces of information include: * Group membership * Score on a matching variable * Response to an item. He identified that DIF is present when expected item scores differ across groups conditional on the matching variable and also DIF is present when group membership tells one something about responses to an item after controlling for the latent construct. Brown (2012) asserted that Mantel-Haenszel method is a popular DIF method since the late 1980's, and the method is still effective compared with newer methods.

The steps in the Mantel-Haenszel procedure, according to Brown, include

- Examine whether the Mantel-Haenszel statistic is statistically significant
- Examine the size of the common odds ratio. If the item does not show DIF, the ratio is expected to be 1.
- Use the Educational test services (ETS) classification scheme to judge the practical significance of DIF. The ETS Delta scale is computed a $\Delta MH = -2.35\lambda MH$ and the following cut-offs are used- Large DIF $|\Delta MH$ $|>1.5$ (Class C), Moderate DIF $1< |\Delta MH | \leq 1.5$ (class B) and Small DIF $| \Delta MH | \leq 1$ (Class A).

Mantel-Haenszel procedure, according to Schumacker (2005), is popular and very appropriate for a small sample size. For Narayanam and Swaminathan (1994), they established that There are two non-parametric procedures for detecting DIF which are; Mantel-Haenszel (MH) and the Simultaneous Item Bias procedure. The authors maintained that the MH statistic is used to compare the performance of two groups of examinees one item at a time. In Mantel-Haenszel's procedure, examinees are grouped into classes on the bases of matching variables. Most often, the matching variable is the total score on the test. The MH provides a significant test and also a measure of the effect size. This latter feature is useful because DIF detection alone does not always tell a complete story (Cohen, 2006).

By interpretation a value of zero for MHindicates no DIF in studied item, a positive value indicates the studied item favors the focal group, and a negative value indicates the studied item favors the reference group. This study, therefore, adopted the Mantel-Haenszel DIF to match the four ethnic groups in Nigeria. Among the over 250 ethnic groups, the four major ones are chosen, they are Igbo, Hausa, Yoruba and Ijaw. In this study, Ijaw is used as the focal group while the other ethnic groups are taken as reference groups.

MATERIALS AND 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. Mantel Heanszel-DIF was used to analyze the data obtained (Queensoap, 2014).

RESULTS

Table1 above shows result ofpair-wise comparison of examinee performance of different ethnic groups which include, Ijaw versus Hausa, Ijaw versus Yoruba, and Ijaw versus Igbo. It was observed that 59 items showed differential item functioning (DIF) between Ijaw versus Hausa, and Ijaw versus Yoruba while all the 60 items indicated DIF for Ijaw examinees and Igbo examinees. Table 1 indicated that out of the 60 items 7 items (3,21,30,33,34,40&45) and 8 items (6,15,17,24,28,32,53&56) were apparently flagged as large and moderate effect sizes respectively for Ijaw and Hausa examinees.

For Ijaw versus Yoruba, 5 items (3,4,30,32&40) and 6 items (5,7,8,21,38,&45) showed large and moderate DIF respectively while 4 items (6,25,40&44) and 6 items (1,11,35,42&56) showed large and moderate DIF respectively for Igbo ethnic group. It was also identified that item 20 and 36 showed NO DIF for Ijaw versus Yoruba and Ijaw versus Hausa respectively. Again table 1 clearly portrayed item 40 as showing large effect size DIF among focal and reference groups. Table 1 as well indicated 20 items (9,13,14,15,19,22,23, 26,29,41,46,49,50,51,54,57,58,59&60) showing DIF against the Ijaw examinees while 10 items (1,3,6,21,39,40,42,44, 45&56) were flagged having DIF against all reference groups.

Table 1. Summary of Mantel-Haenszel DIF Analysis

Item No.	Ethnic Group	Odd Ratios	MH–DIF	ETS Classification	Against
1	Hausa	0.67	0.94	A	Hausa
	Yoruba	0.88	0.31	A	Yoruba
	Igbo	0.58	1.27	B	Igbo
2	Hausa	1.40	-0.80	A	Ijaw
	Yoruba	0.98	0.05	A	Yoruba
	Igbo	2.91	-2.51	A	Ijaw
3	Hausa	0.37	2.33	C	Hausa
	Yoruba	0.31	2.75	C	Yoruba
	Igbo	0.68	0.89	A	Igbo
4	Hausa	0.66	0.96	A	Hausa
	Yoruba	0.42	2.04	C	Yoruba
	Igbo	1.03	-0.07	A	Ijaw
5	Hausa	1.36	-0.73	A	Ijaw
	Yoruba	0.65	1.01	B	Yoruba
	Igbo	1.01	-0.02	A	Ijaw
6	Hausa	0.60	1.20	B	Hausa
	Yoruba	0.84	0.40	A	Yoruba
	Igbo	0.41	2.09	C	Igbo
7	Hausa	0.66	0.96	A	Hausa
	Yoruba	0.64	1.06	B	Yoruba
	Igbo	3.41	-2.89	A	Ijaw
8	Hausa	1.11	-0.23	A	Ijaw
	Yoruba	0.65	1.01	B	Yoruba
	Igbo	1.08	-0.19	A	Ijaw
9	Hausa	1.97	-1.60	A	Ijaw
	Yoruba	1.77	-1.34	A	Ijaw
	Igbo	1.80	-1.39	A	Ijaw
10	Hausa	0.78	0.59	A	Ijaw
	Yoruba	0.73	0.73	A	Yoruba
	Igbo	1.57	-1.06	A	Ijaw
11	Hausa	0.82	0.47	A	Ijaw
	Yoruba	1.29	-0.59	A	Ijaw
	Igbo	0.64	1.06	B	Igbo
12	Hausa	3.93	-2.87	A	Ijaw
	Yoruba	1.53	-1.01	A	Ijaw
	Igbo	0.69	0.87	A	Ijaw
13	Hausa	1.57	-1.06	A	Ijaw
	Yoruba	1.19	-0.40	A	Ijaw
	Igbo	2.33	-1.97	A	Ijaw
14	Hausa	1.50	-0.94	A	Ijaw
	Yoruba	1.42	-0.82	A	Ijaw
	Igbo	1.02	-0.05	A	Ijaw
15	Hausa	0.59	1.24	B	Hausa
	Yoruba	1.68	-1.22	A	Ijaw
	Igbo	1.43	-0.85	A	Ijaw
16	Hausa	3.22	-2.75	A	Ijaw
	Yoruba	6.19	-4.28	A	Ijaw
	Igbo	0.77	0.59	A	Igbo
17	Hausa	0.61	1.15	B	Ijaw
	Yoruba	1.25	-0.52	A	Ijaw
	Igbo	0.95	0.12	A	Ijaw
18	Hausa	0.97	0.07	A	Ijaw
	Yoruba	0.88	0.31	A	Yoruba
	Igbo	2.11	-1.76	A	Ijaw
19	Hausa	2.31	-1.97	A	Ijaw
	Yoruba	2.57	-2.21	A	Ijaw
	Igbo	4.57	-3.57	A	Ijaw
20	Hausa	0.84	0.40	A	Ijaw
	Yoruba	1.00	0.00	N/D	Ijaw
	Igbo	1.87	-4.39	A	Ijaw

21	Hausa	0.31	2.75	C	Hausa
	Yoruba	0.56	1.36	B	Yoruba
	Igbo	0.72	0.77	A	Ijaw
22	Hausa	2.29	-1.95	A	Ijaw
	Yoruba	1.36	-0.73	A	Ijaw
	Igbo	2.25	-1.90	A	Ijaw
23	Hausa	2.35	-2.00	A	Ijaw
	Yoruba	1.30	-0.61	A	Ijaw
	Igbo	2.29	-1.95	A	Ijaw
24	Hausa	0.54	1.46	B	Hausa
	Yoruba	0.84	0.40	A	Yoruba
	Igbo	1.74	-1.32	A	Ijaw
25	Hausa	0.68	0.89	A	Ijaw
	Yoruba	1.31	-0.63	A	Ijaw
	Igbo	0.28	3.01	C	Igbo
26	Hausa	1.04	-0.09	A	Ijaw
	Yoruba	1.07	-0.16	A	Ijaw
	Igbo	1.23	-0.49	A	Ijaw
27	Hausa	0.69	0.87	A	Ijaw
	Yoruba	1.33	-0.68	A	Ijaw
	Igbo	1.59	-1.08	A	Ijaw
28	Hausa	0.57	1.32	B	Ijaw
	Yoruba	2.17	-1.81	A	Ijaw
	Igbo	1.59	-1.08	A	Ijaw
29	Hausa	1.82	-1.41	A	Ijaw
	Yoruba	1.53	-1.01	A	Ijaw
	Igbo	1.30	-0.63	A	Ijaw
30	Hausa	0.21	3.67	C	Hausa
	Yoruba	0.16	4.30	C	Yoruba
	Igbo	1.04	-0.09	A	Ijaw
31	Hausa	3.62	-3.03	A	Ijaw
	Yoruba	0.93	0.16	A	Ijaw
	Igbo	2.26	-1.90	A	Ijaw
32	Hausa	0.56	1.36	B	Hausa
	Yoruba	0.31	2.75	C	Yoruba
	Igbo	1.59	-1.08	A	Ijaw
33	Hausa	0.34	2.54	C	Hausa
	Yoruba	0.85	0.38	A	Ijaw
	Igbo	1.38	-0.75	A	Ijaw
34	Hausa	0.36	2.40	C	Hausa
	Yoruba	1.03	-0.07	A	Ijaw
	Igbo	2.18	-1.70	A	Ijaw
35	Hausa	1.31	-0.63	A	Ijaw
	Yoruba	0.91	0.24	A	Ijaw
	Igbo	0.57	1.32	B	Ijaw
36	Hausa	1.00	0.00	N/D	Ijaw
	Yoruba	1.25	-0.52	A	Ijaw
	Igbo	0.75	0.68	A	Ijaw
37	Hausa	0.72	0.77	A	Ijaw
	Yoruba	1.36	-0.73	A	Ijaw
	Igbo	0.52	1.53	C	Ijaw
38	Hausa	1.41	-0.80	A	Ijaw
	Yoruba	0.55	1.41	B	Yoruba
	Igbo	0.76	0.63	A	Ijaw
39	Hausa	0.79	0.54	A	Hausa
	Yoruba	0.80	0.52	A	Yoruba
	Igbo	0.92	0.19	A	Igbo
40	Hausa	0.32	2.68	C	Hausa
	Yoruba	0.36	2.40	C	Yoruba
	Igbo	0.48	1.71	C	Igbo
41	Hausa	3.04	-2.61	A	Ijaw
	Yoruba	1.61	-1.13	A	Ijaw
	Igbo	2.36	-2.03	A	Ijaw
42	Hausa	0.77	0.61	A	Ijaw
	Yoruba	0.87	0.33	A	Yoruba
	Igbo	0.55	1.41	B	Ijaw
43	Hausa	1.49	-0.94	A	Ijaw
	Yoruba	0.70	0.85	A	Yoruba
	Igbo	0.86	0.35	A	Ijaw
44	Hausa	0.81	0.49	A	Ijaw
	Yoruba	0.68	0.92	A	Yoruba
	Igbo	0.38	2.28	C	Ijaw
45	Hausa	0.45	1.88	C	Hausa
	Yoruba	0.59	1.25	B	Yoruba
	Igbo	0.59	1.24	B	Ijaw
46	Hausa	2.58	-2.23	A	Ijaw
	Yoruba	1.67	-1.20	A	Ijaw
	Igbo	2.43	-2.09	A	Ijaw

Continue……..

47	Hausa	1.17	-0.38	A	Ijaw
	Yoruba	0.90	0.26	A	Ijaw
	Igbo	0.87	0.33	A	Ijaw
48	Hausa	1.05	-0.12	A	Ijaw
	Yoruba	0.75	0.68	A	Yoruba
	Igbo	2.19	-1.83	A	Ijaw
49	Hausa	1.31	-0.63	A	Ijaw
	Yoruba	2.01	-1.65	A	Ijaw
	Igbo	1.75	-1.32	A	Ijaw
50	Hausa	2.40	-2.04	A	Ijaw
	Yoruba	2.69	-2.33	A	Ijaw
	Igbo	2.25	-1.90	A	Ijaw
51	Hausa	1.21	-0.45	A	Ijaw
	Yoruba	2.07	-1.72	A	Ijaw
	Igbo	2.76	-2.37	A	Ijaw
52	Hausa	2.10	-1.74	A	Ijaw
	Yoruba	1.29	-0.59	A	Ijaw
	Igbo	1.03	-0.07	A	Ijaw
53	Hausa	0.56	1.36	B	Hausa
	Yoruba	1.04	-0.09	A	Ijaw
	Igbo	0.75	0.68	A	Igbo
54	Hausa	1.98	-1.64	A	Ijaw
	Yoruba	2.09	-1.74	A	Ijaw
	Igbo	1.71	-1.27	A	Ijaw
55	Hausa	1.29	-0.59	A	Ijaw
	Yoruba	1.07	-0.16	A	Ijaw
	Igbo	0.97	0.07	A	Ijaw
56	Hausa	0.63	1.08	B	Ijaw
	Yoruba	0.82	0.47	A	Ijaw
	Igbo	0.62	1.13	B	Igbo
57	Hausa	4.54	-3.55	A	Ijaw
	Yoruba	1.15	-0.33	A	Ijaw
	Igbo	4.50	-3.52	A	Ijaw
58	Hausa	3.09	-2.65	A	Ijaw
	Yoruba	1.10	-0.24	A	Ijaw
	Igbo	2.71	-2.35	A	Ijaw
59	Hausa	3.45	-0.96	A	Ijaw
	Yoruba	1.50	-0.96	A	Ijaw
	Igbo	2.89	-2.49	A	Ijaw
60	Hausa	1.36	-0.73	A	Ijaw
	Yoruba	1.06	-0.14	A	Ijaw
	Igbo	1.34	-0.68	A	Ijaw

MHD – DIF = -2.35 x log λMH where λMH = odds ratio.

DISCUSSION

The analysis was done with Mantel-Haenszel adjusted DIF statistics. The focused group (Ijaw) was used to compare with the reference groups, (Igbo, Hausa, and Yoruba) separately. The test statistics identified the items with either positive or negative values. Cohen (2006) in his work established that a zero for adjusted Mantel–Haenszel (MH) will indicate "NO DIF" in the studied item. A positive value indicates that the studied item favours the focal group while negative value flagged the studied item favouring the reference group. consequently, the findings of the study revealed items 9, 13, 14, 15, 19, 22, 23, 26, 29, 41, 46, 49, 50, 51, 52, 54, 57 58, 59 and 60 as exhibiting DIF that favours the reference groups (Hausa, Yoruba & Igbo) against the focal group because they have a negative value while item 1, 3, 6, 21, 39, 40, 42, 44, 45 and 56 favours the focal group against the reference groups whereas items other than the ones mentioned above were flagged DIF either in favour or against of the Ijaw and the Hausa, Yoruba or Igbo. Again, using Brown (2012) ETS Delta Scale cut off some items are flagged as displaying DIF were observed within the "Class A", 'B' or 'C'. This signifies that the items were having either small, moderate or large DIF. Cohen (2006) asserted that the use of effect size measure. λMH=log (αMH) is plausible to the use of Mantel – Haenszel approach. He added that their feature is useful because DIF detection alone does not always tell the complete story. This idea was supported in the works of Zumbo (2007 and 1999),

Abedalaziz (2011), Schumacker (2005) and Williams (1997), stating that DIF is required but not sufficient enough to flag an item biased. Yang and Jones (2007) used Mantel–Haenszel method to detect item response bias which showed similar findings with that of this research work. Similarly, the work of Fidalgo, Ferreres and Muniz (2004) showed Mantel–Haenszel results agreeing with SIBTEST. The reliability of using Mantel– Haenszel approach was asserted in works of Yang and Jones (2007). Schumacker (2005) agreed that Mantel–Haenszel method is suitable with smaller sample sizes therefore considering the sample size of this study, the result giving with MH was admissible.

Conclusion

The Mantel-Haenszel DIF method revealed 59 items of the CAT items showing differential item functioning for Ijaw examinees versus Hausa as well as Yoruba examinees while all 60 items were flagged showing DIF between Ijaw and Igbo examinees. The ETS classification classified item 40 as showing large effect size among comparing groups. The study has made us to understand that responses to some of the Chemistry Achievement Test items were influenced by ethnicity. All the approaches revealed this fact to a very statistically significant extent. The study has made us to appreciate that, even though, DIF/IRT methods seem very complicated (require special software), they are theoretically appropriate and inferentially superior.

Recommendation

The Mantel Haenszel DIF should be used to detect for ethnic bias test items in Nigeria, so that, test items should be free from bias towards a particular group. The development, scoring and analysis of public examinations, especially quiz competitions and unity examinations in Nigeria, should adopt the item response theory based approach.

REFERENCES

Abedalaziz, N. 2011. Detecting DIF using item characteristics curve approaches. The Internationals Journal of Educational and Psychological Assessment, 8(2), 1 – 15.

Brown, A. 2012. Measurement invariance: university of Cambridge, department of psychology

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

Federal Republic of Nigeria 2004. National policy on education. Abuja.

Narayanam, P. and Swaminathan, H. 1994. Performance of the mantel-Haenszel and simultaneous item bias procedures for detecting differential item functioning. *Applied Psychological Measurement*, 18(4), 315-325.

Onunkwo, G.I. N. 2002. Fundamentals of evaluation. Owerri: Cope

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.

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. Application of differential item functioning in detecting item bias in chemistry achievement test in Nigeria. Unpublished doctorial dissertation, University of Port-Harcourt, Choba, Port-Harcourt. Rivers State. Nigeria.

Roever, C. 2005. That's not fair, bias, and differential item functioning in language testing. Retrieved October 11, 2010, from http://www.hawii.edu/roever/brownbag.pdf.

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

Wiersma, W. and Jurs, S. G. 1990. Educational measurement and testing (2nded.). Massachusetts: Allyn and Bacon.

Williams, V.S.L. 1997. The unbiased anchor: bridging the gap between DIF and item bias. *Applied measurement in Education,* 10(3), 253-267.

Yang, F.M. and Jones, R. N. 2008. Center for epidemiological studies – depression scale (CES.D) item response bias found with mantel – haenszel method successfully replicated using latent variable modeling. Retrieved march 30, 2011, from. file://f:/center for epidemiologic studies - depression scale (CES.D) item response bias…

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.

****