AFRICAN JOURNAL OF EDUCATION AND TECHNOLOGY

African Journal Of Education And Technology
Volume 5 Number 1 (2015) 33-49
ISSN 2046-6927 (Print) ISSN 2046-6935 (Online)
Publishers: Sacha & Diamond, England, United Kingdom
www.sachajournals.com
Cumulative Impact Factor: 37.0

THE PROBLEMS OF DETECTING GENDER AND LOCATION BIASED TEST ITEMS IN JUNIOR SECONDARY CERTIFICATE EXAMINATIONS IN RIVERS STATE: ADOPTING THE ITEM RESPONSE THEORY APPROACH

ASIEGBU, Chinelo Nneka[1]
and
JOE , A.I (Ph.D)[1]

[1]Department of Educational Psychology,
Guidance and Counselling, Faculty of Education,
University of Port Harcourt, Nigeria

ABSTRACT

This study identified gender and location biased test items in the Rivers state JSSCE Business Studies conducted in 2009 using Item Response Theory Approach. The Sample comprises 4000 male and female JSS III students that sat for the JSSCE Business Studies Examination in 2009 drawn from schools in the urban and rural locations in Rivers state which were selected through stratified proportionate random sampling. The instruments for data collection were the test items in JSSCE Business Studies conducted in 2009 and a result collection sheet. Three null hypotheses drawn from three research questions were tested at 0.50 probability of correct response. The assumptions of unidimensionality and local independence were tested using factor analysis. The students' responses to the test items were calibrated with an IRT statistical software named Ex-calibre 4.2.2 developed by Assessment Systems Corporation. The results revealed that the test items met the assumptions of unidemensionality and local independence and there is gender and location bias in the Rivers state JSSCE Business Studies test items for 2009. Based on the findings, recommendations were made which includes restructuring of the business studies curriculum to accommodate the new technological advances which may be one of the reasons for the presence of gender and location bias in the Business Studies test conducted in 2009 in Rivers State.

Keywords: Evaluation, Measurement, Education, Item Bias, Nigeria.

1. INTRODUCTION

Achievement test is one of the means through which educators assess learning outcomes. It sometimes, includes items that function differentially for different sub – groups of a population like males, females, urban and rural students. It is important to trace and identify the items as they may lead to unfair interpretation of test results for the groups concerned. Test results help educators know how students learn and provide feedback for restructuring and modifying the teaching – learning process. Tests are also often used to make important decisions about people. In educational system, institutions use scores from tests to make decisions on admissions like placement of students in gifted schools. Some companies also make job decisions about people based on test scores. There is therefore the need to ensure that tests administered by different examination bodies are free of bias and are as fair as possible for all subgroups of a population.

Bias then, according to Hambleton & Rogers (2012) is the presence of some characteristics of an item that result in differential performance for individuals of the same ability but from different ethnic, sex, cultural or religious groups. This means that an item is considered biased when different groups but with the same ability level, have nevertheless, different probabilities of answering the item correctly. Biased items can lead to biased measurement of ability because the measurement is affected by so-called nuisance factors (Ankerman, 2002).The presence of bias in test items threatens the content validity of such tests. This means that the items are measuring attributes that are not necessary or relevant to the construct being measured by the test.

A valid and reliable examination indicates that the exam has been carried out in accordance with its target and test items were prepared in line with the purpose of the exam. If the item in the test provides an advantage to any of the group taking the test because of various features like gender and location, it can be said that the test has bias in favour of that group. This will negatively affect the validity of the decisions made based on the test scores. Bias can also lead to systematic error which may, according to Zumbo (2012), distort the inferences made in the classification and selection of students. Tests should be used to help not to harm, they are supposed to benefit the examiners. Biased test items can be explained away as real differences as a result of gender and location. This can be damaging to students for it could lead to individual needs not being diagnosed and denied opportunities. Interpretation of biased test results creates a lot of problem as many candidates who are supposed to be qualified in certain things like promotion are sometimes disqualified. This implies that biased test item blurs the reason for testing which is to reveal the latent ability of examiners. There is therefore the need for different examination bodies like the Rivers State Ministry of Education to construct test items in such a way that they will be free of bias so that any disparity between different sub – group will be attributed to real differences in what the test intends to measure.

Business Studies is an arm of vocational education which prepares the youth for future membership and participation in the life of the society and for its maintenance, growth and development. It is meant for skill acquisition and is designed to prepare the students for the world of work. At the Junior Secondary level, it has five arms which include commerce, office practice, typing, shorthand and book keeping.A biased Business Studies Test may be due to the presence of irrelevant constructs related to gender and location, and daily life experiences. It has also been observed that most questions in the Rivers State JSSCE Business Studies are often repeated yearly. If those items are biased and are not screened, the validity of the test items will be jeopardized. More so, in the analysis of the JSSCE Business Studies results, gender and location are not equally represented. Discrepancies have been observed in the result in favour of one sub-group of the population or the other. Therefore, it is necessary to ensure that Business Studies test do not contain items that will differentially in different sub-groups of the population and which may cause bias to students with regards to gender and location.

2. THEORETICAL FRAMEWORK

Item response theory provides a unified framework for conceptualizing and investigating bias at the item level. Test developers and educational researchers have developed a number of item bias detection procedures. Ertuby & Russele (1999), Ojerinde (2012) and Osterlind (2014) suggested that because of their high level of sophistication, IRT, procedures provide the best results for detecting biased items. This means that biased items in Rivers State JSSCE Business Studies can be detected using an IRT approach.

Mellenberg (1994) posited that "IRT is a model for expressing the association between an individual's response to an item and the underlying latent variable often called ability or trait being measured by the instrument. IRT models use item responses to obtain scaled estimations of θ as well as to calibrate items and examine their properties". In agreement to this, Lord (1980) states that the main task in most testing work is to infer to the examinees ability level or skill. In order to do this, something must have to be known about how an examinees knowledge or skill determines his response to an item. Thus IRT starts with a mathematical statement of how response depends on the level of ability or skill. Osterlind (2014) believes that it is a collection of models that provide information about the properties of items and scales they comprise of through analysis of individual item response. IRT is also called a latent theory because the theory assumes the existence of a latent trait which is a tester characteristic that leads to a consistent performance in a test. Reeve (2003) believes that IRT refers to a set of mathematical models that describe in probabilistic terms the reason between a person's response to an item and his\her level of the latent variable being measured. In order words, IRT assumes that there is an underlying probability that someone with a particular level on the latent trait will endorse a given item. For example, a person with a high ability level in Business Studies will have a high likelihood of responding correctly most of the time to items in a Business Studies achievement test, whereas a person with a low ability level will have a low likelihood of responding correctly except for guessing. This means that IRT depicts a more thorough picture of item functioning as it predicts how an examinee will perform on a test item. This, according to Crocker & Angina (1986), Zhu (2006) allows for comparisons to be made between examinees that have taken different tests and also allows for bias detection.

IRT methods have very great capabilities for diagnosing measurement problems. A test developer using an IRT technique generates a mathematical function to describe the relationship between test performance and ability or trait (Reid, Kolakowsky, Lewis & Armstrong, 2007). If a test has been developed using IRT, the level of the ability and trait measured can be estimated for an individual completing any subset of the test items. Administration of the entire test may no longer be necessary.

IRT also has potent tools for studying potential bias in assessment instruments. It is from IRT that the item characteristics curve (ICC) approaches to the detection and correction of test item bias are derived (Osterlind, 1983). It has been seen as particularly useful for studies of item bias because in theory, the ICC does not depend on the distribution of ability in the sample used to ascertain the parameters. Hence, if the parameters of an ICC are estimated separately for two samples drawn from the same population, the resultant curves are supposed to be the same even though the sample may differ in the distribution of abilities within them. If the curves are not the same, the conclusion of the bias may be necessary.

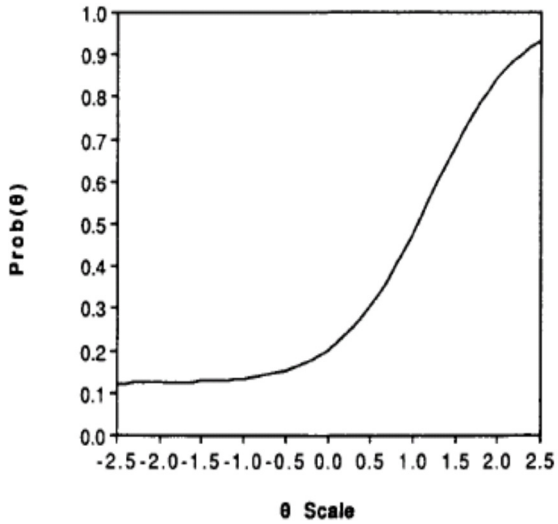
A variety of models have been developed from the IRT perspective and these models differ from each other in at least two important ways. One important difference among the measurement techniques is in terms of the item characteristics or parameters that are included in the model. A second important difference is in the nature or type of the response option format. For test items that are dichotomously scored, their IRT models are known as one, two and three parameter models, 1pl, 2pl and 3pl for short. The simplest IRT model is the 1pl which contains only a single parameter and is usually described as the Rasch model. The two parameter model adds a varying slope to the Rash model allowing item responses and person abilities to interact. In

this model, each item varies in average reliability. The 2pl uses both difficulty and discrimination parameters. The difficulty parameter (b) tells us how easy or how difficult an item is. It is the only parameter used in the 1pl model. The discrimination parameter is also known as the (a) parameter and it shows how efficiently an item can distinguish between highly knowledgeable students and less knowledgeable students. The 3pl uses a, b and c parameters. The c parameter is the guessing parameter and its value shows how likely the examinees are to obtain the correct answers by guessing. When the guessing parameter is taken into consideration, it portrays the fact that in some items, though the examinee may not know anything about the subject matter, he or she can still have a probability of getting the right answer. These models are designed to be used for binary outcomes as the response option.

3. ITEM CHARACTERISTIC CURVE (ICC)

Item characteristic curve (ICC) is the functional relationship between the proportion of correct response to an item and a criterion variable (Baker & Kim, 2004). Psychometricians use ICC to describe and evaluate characteristics of the items in a test. It reflects the probability with which the individuals across the range of trait levels are likely to answer each item correctly. It models the relationship between a person's probability for endorsing an item category and the level of construct measured by the scale. Osterlind (1983) sees it as a graphic presentation of a mathematical function that describes the probability of an examinee correctly answering a test item relating to the ability measured by the total set of items of the test. ICC shows the relationship between the individuals' ability on an item and the likelihood that the individual will respond correctly to that particular item (Suen, 2010). ICC is assumed by researchers to be the major concept of IRT as it is a monotonically increasing function which relates the relationship between examinees' item performance p(Θ) and the trait (Θ) underlining item performance. The regression of the score of examinees' abilities, otherwise known as the Item Response Function (IRF) is shown in Figure 1.

Figure 1: Item Characteristics Curve (ICC) from Harris (2012)



On the x-axis is the function together with the examinee's ability level while the probability to answer an item correctly is shown on the y-axis. Basically, every candidate or examinee is presumed to acquire some level of ability (represented by Θ) and is thus placed on the ability scale according to that level. For every examinee, there is a probability that he/she, according to his/her ability, will be able to answer an item correctly. This probability, represented by P(Θ), is lower for individuals with lower ability levels and higher for those with higher levels

in a given situation. Thus, a plot of probability function P(Θ) against (Θ) will produce an S-curve, which depicts the form of the ICC (Baker, 2012, Lord, 1980, Warm, 2008). This means that if the probability function of students' Achievement test in JSSCE Business Studies is plotted against their ability levels, a typical S-shaped form of the ICC (Figure 1) is supposed to appear. Each item in the test will have a separate ICC, the shape of which will be determined by the type of IRT model used.

IRT has caused significant changes in psychometric theory and test development. As noted by Hambleton & Slater in Marie (2004), it basically assumes that a single trait depicts an examinee's performance on a given test and that the probability of a correct response on an item is a monotonically ascending curve. From IRT, Differential Item Functioning (DIF) is derived which is a powerful method for investigating item bias. The ICC of an item shows graphically the probability of the correct response as a function of the magnitude or level of the underlying trait being measured. Osterlind (1983) describes ICCs as "the most elegant of all the models to tease out item bias".

The purpose of this study therefore is to detect items showing bias in JSSCE Business Studies conducted in Rivers State in 2009 with regards to gender and location using IRT approach. The following research questions guided this study:

a) To what extent do test items in Rivers State JSSCE Business Studies conducted in 2009 comply with the assumptions of unidimensionality and local independence?
b) To what extent do test items in Rivers State JSSCE Business studies conducted in the 2009 show up gender differences?
c) To what extent do test items in Rivers State JSSCE Business Studies conducted in 2009 show up location differences

The following null hypotheses were tested:

a) The test items in Rivers State JSSCE Business Studies conducted in 2009 will not differ significantly in complying with the assumptions unidimensionality and local independence.
b) The ICCs of students' responses to the test items in Rivers State JSSCE Business Studies conducted in 2009 will not differ significantly due to gender at 0.50 probability of correct response.
c) The ICCs of students' responses to the test items in Rivers State JSSCE Business Studies conducted in 2009 will not differ significantly due to location at 0.50 probability of correct response.

4. METHODOLOGY

Analytical descriptive research design was adopted for this study. The population comprised 52893 male and female students in the urban and rural location that sat for the JSSCE Business Studies conducted in Rivers State in 2009. Proportionate stratified random sampling was used to obtain a total of 4000 students consisting of 1948 males (1081 from urban location and 867 from rural location) and 2052 females (1204 from urban location and 848 from rural location). Data were extracted from the students' responses to the test items with the aid of staff members, exams and records department, Rivers State Ministry of Education with a result collection sheet. Factor analysis with the aid of SPSS was performed on the students' responses to the test items so as to check if the data met the assumptions of unidimensionality and local independence. An IRT software, Ex- caliber 4.2.2 developed by Assessment Systems Corporation was used to estimate the item parameters with the marginal maximum likelihood estimation technique and to also draw the ICCs.

5. RESULTS

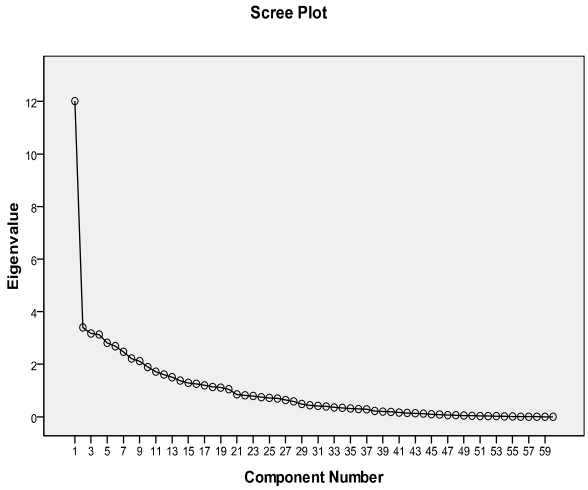


Fig. 2: Scree plot of the eigen values

From Appendix 1 and fig. 2, the hypotheses that the test items in Rivers State JSSCE Business Studies conducted in 2009 will not differ significantly in complying with the assumptions of unidimentionality and local independent therefore upholds. This is because the eigen value of the first factor as is shown in Apendix 1 and figure 2 accounts for 20% of the total variance which agrees with Reckase (1979) minimum criterion to assure unidimentionality of data. It is also large compared to the second factor and the eigen value of all the other factors are all about the same as suggested by Ojerinde (2012).

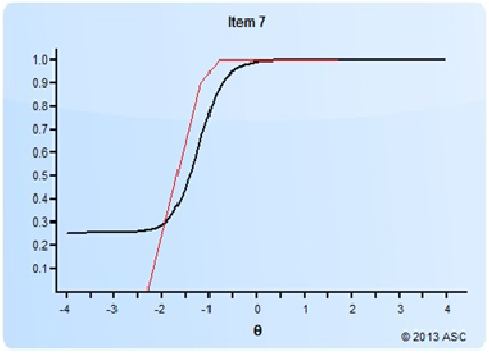
Only 8 out of the 60 items functioned equally between male and female students who sat for the JSSCE Business Studies conducted in Rivers State in 2009. All the other items were biased against one group or the other. This means that only 8 ICCs are similar. Hypotheses 3: Appendix 3 below shows the item parameters of the 60 test items while fig 4 (1 – 2) shows the ICCs of the items that are similar for students in the urban and rural location that sat for the JSSCE Business Studies in 2009.

Fig. 3: (1-8): Items functioned equally between male and female students who sat for the JSSCE Business Studies conducted in Rivers State in 2009

Female

A b C

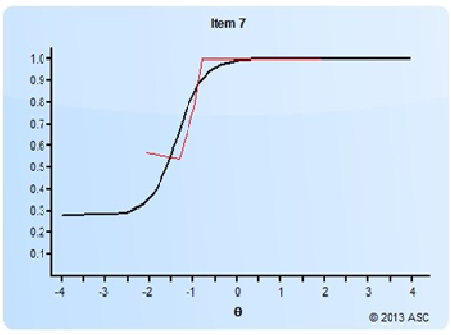
2.153 -1.176 0.252



Male

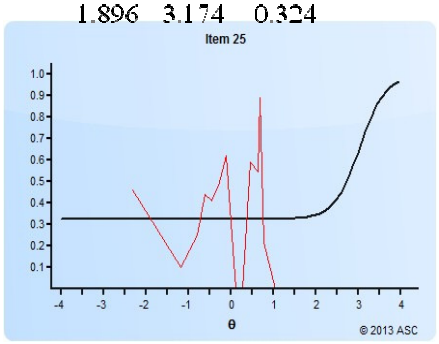
a b c

2.144 -1.172 0.282

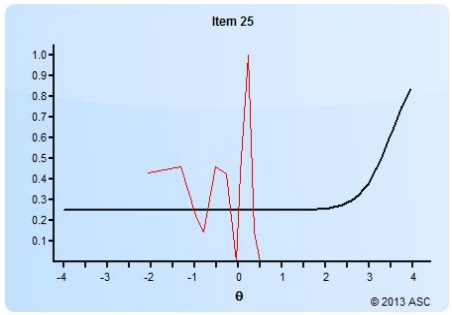


Female

a	b	c
1.896	3.174	0.324

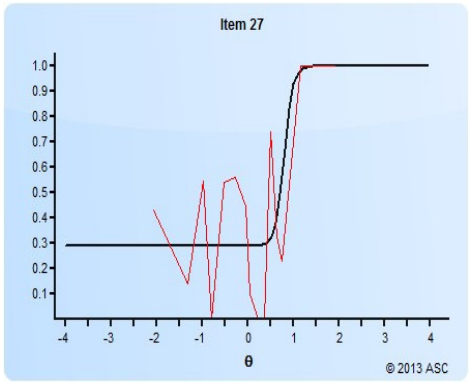


Male



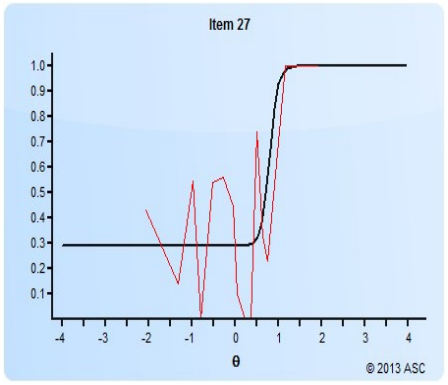
Female

a	b	c
6.000	0.886	0.288



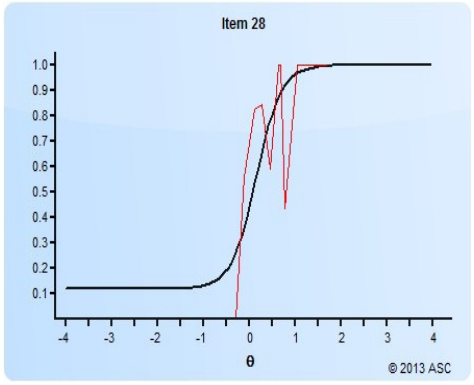
Male

a	b	c
6.000	0.886	0.288



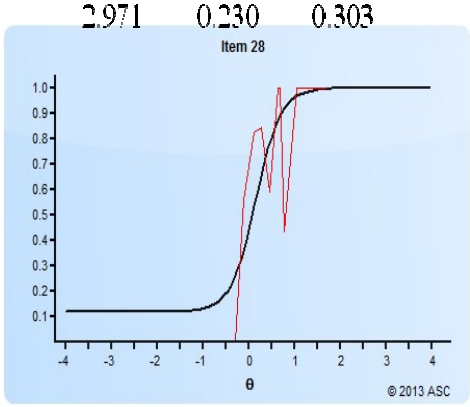
Female

a	b	c
2.971	0.230	0.303



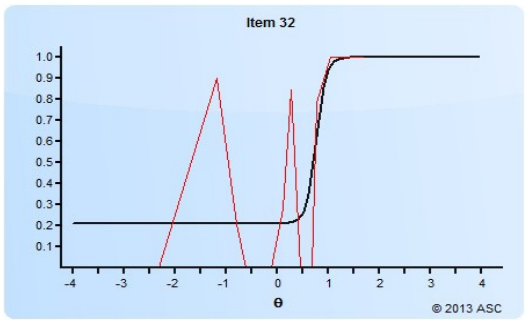
Male

a	b	c
2.971	0.230	0.303



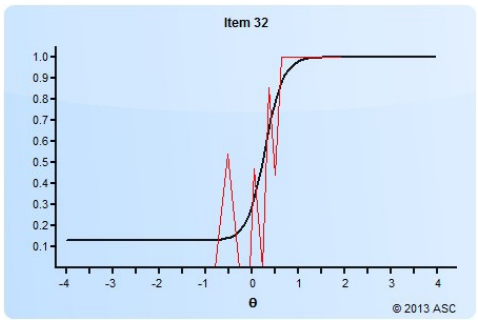
Female

A	b	C
6.000	0.852	0.211

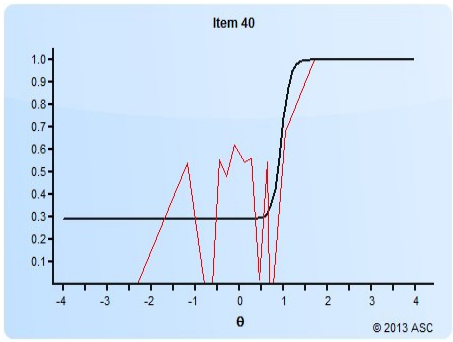


Male

a	b	c
5.003	0.864	0.129

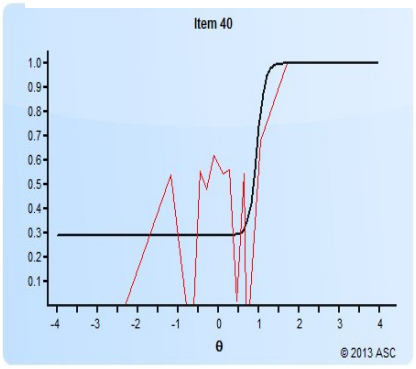


a	b	C
6.000	1.047	0.290



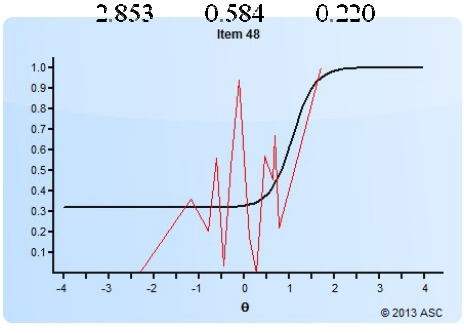
Male

a	b	
6.000	1.047	0.290



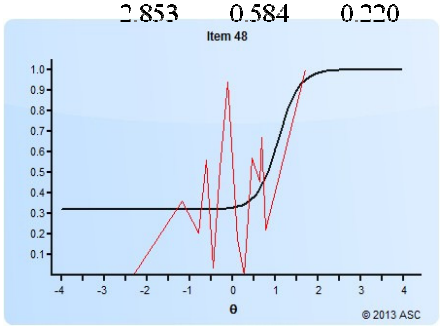
Female

a	b	C
2.853	0.584	0.220



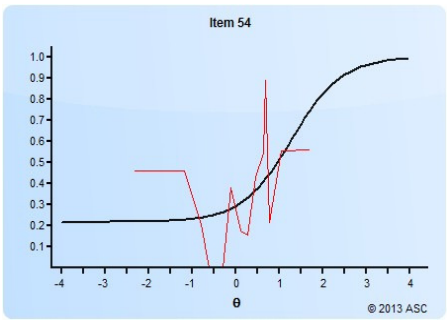
Male

a	b	c
2.853	0.584	0.220



Female

a b c
1.008 1.378 0.217



Male

a b c
1.008 1.378 0.217

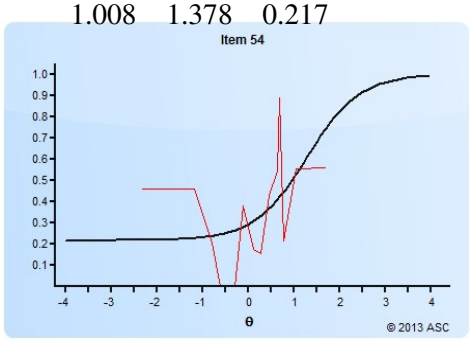
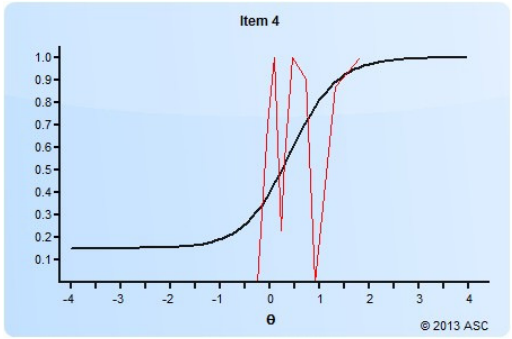


Fig. 4: (1-4) Only 2 of the items out of the 60 items have similar ICCs for the urban and rural students that sat for the JSSCE Business Studies conducted in 2009.

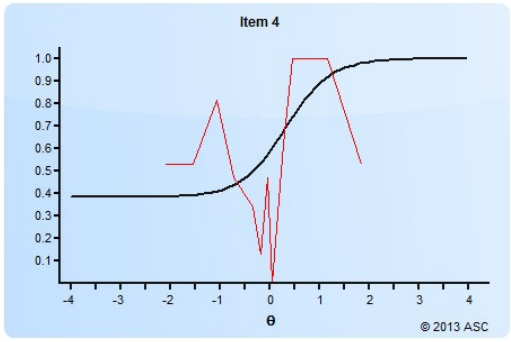
Rural

a b c
1.281 0.394 0.151



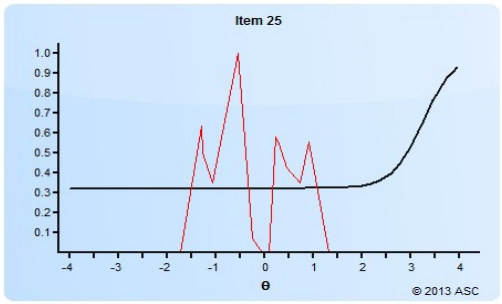
Urban

a b c
1.281 0.394 0.383



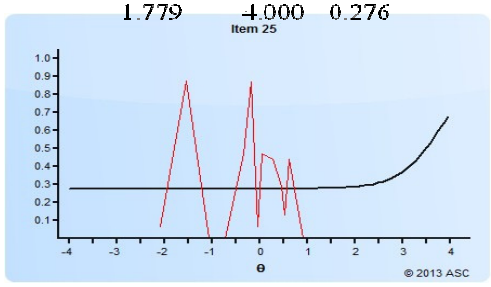
Rural

a b c
1.781 4.000 0.322



Urban

a b c
1.779 4.000 0.276



6. DISCUSSION

Following the results of the calibration of data, 35 items namely, items 2, 3, 4, 5, 6, 10, 12, 14, 16, 17, 18, 19, 20, 21, 22, 23, 24, 33, 34, 37, 39, 41, 44, 45, 47, 49, 51, 52, 53, 55, 56, 57 and 60 were biased against the females. On the other hand, 14 items (8, 9, 11, 15, 26, 30, 35, 36, 38, 42, 43, 46, 50, and 58) were biased against the males. Items 7, 25, 27, 28, 32, 40, 48, and 54 were equally difficult for both groups, which means they were not biased against any group. There was no variance for the males in items I, 29, and 59. Item 1 was so simple that all the males answered it correctly unlike items 29 and 59 which were extremely difficult. In fact both groups found the two items very difficult.

Generally, from the results of the analyses, the male students found the items easier than their female counterparts, 35 items favoured them while only 14 items favoured the females. A closer look at these questions reveals that the female students excelled in items that are related to office management while the males excelled in those items that involved problem solving and application. This agrees with the findings of Kuram (2012), Predjata & Talisayon (2009), Adedoyin (2010) and Mokobi & Adedoyin (2013). However, some of the items like items 25 and 33, though favourable for the male students, were extremely difficult for both groups. It calls for re-wording, revision or removal from the test items. More so, a closer look at the items also reveals that there's no gender clue in the way some of the items that were flagged as biased were phrased. It could therefore be concluded that most of the items that functioned differently between male and female students may not be because the items were biased but because of differential item impact or other factors.

Furthermore, the result of the analysis for students' in the urban and rural location showed that out of the 60 test items, bias was not found in items 4 and 25. There was no variance for the urban group in item 1 and for the rural group in item 59. A total of 28 items which are 2, 8, 10, 11, 13, 14, 15, 17, 18, 20, 21, 22, 23, 24, 27, 29, 30, 31, 35, 39, 40, 42, 46, 48, 51, 52, 54, and 58 were in favour of students in the rural location while items 3, 5, 6, 7, 9, 12, 16, 19, 26, 28, 32, 33, 34, 36, 37, 38, 41, 43, 44, 45, 47, 49, 50, 53, 55, 56, 57, and 60 favoured students in the urban location. Simply put the ICCS's of students' responses to the 60 items of the Business Studies test differed significantly in 56 items due to location. Bias was found in 28 items in favour of students in the rural location while the other 28 were in favour of students in the urban location. Both groups found items 4 and 25 equally difficult. This finding agrees with Asiegbu (2007) where students in the rural location achieved as high as in the urban location. It, however, disagrees with the findings of researchers like Mobito and Adedoyin (2013), Amuche & Fan, (2014), Uremu & Adams (2013) and Predjata & Talisayon (2009). The results of these researchers showed that most of the items that were biased were to the disadvantage of students in the rural location. They claimed that students in the urban location are more exposed than their counterparts in the rural location and as a result, bias is inevitable.

A close look at the findings of this research also reveals that there are some items that both groups found very difficult. Examples of such items are 25 and 59. Students in the urban location also found items from typewriting more difficult than students in the rural location. This might be due to the impact of the computer age on the present day society and humans. Most schools in the urban location no longer purchase typewriters, they see it as outdated, rather they provide their students with computers but schools in the rural location are still using typewriters to date. Meanwhile, subject experts prepare their test items based on the knowledge of typewriter and its usage.

Item 59 had only 4 distracters and most students did not answer it. This may have given room for the high level of difficulty parameters. In item 25, most students may not have known what stencil looks like, how much more the pen with which to write on it. Only very few students scored this item right, perhaps because they were privileged to have come across either the stencil or the pen. Item 1 was very easy for both groups, almost all the students in the urban location scored it right. These items call for revision or rephrasing.

7. CONCLUSION AND RECOMMENDATIONS

The findings from this study reveal that there is gender and location bias in the Rivers State JSSCE Business Studies conducted in 2009. More items were biased against the female students. It was also discovered that female students excelled in items that were based on office management while male students excelled in those requiring computation and analyses. Furthermore, students in the rural location endorsed items in typewriting more correctly than students in the urban location. On the basis of these findings, the following implications are deduced:

a) JSSCE Business Studies conducted in 2009 favours one sub group more than the other. As some of the items are repeated yearly, it therefore calls for revision, rephrasing or rewording of the faulty items.
b) Some aspects of the Business Studies curriculum should be restructured or revised.
c) There may be other factors outside bias affecting students' performance in Business Studies as some of the test items do not have either gender or location clues.

The researchers therefore recommend that test developers in the state and other examination bodies should ensure that their items are bias free by utilizing IRT approach, most especially the Item Characteristic Curve in teasing out biased test items which will subsequently be rephrased or removed entirely from the item bank. Secondly, Business Studies curriculum should be restructured so as to accommodate the new technological advances which may be one of the reasons for location bias in the business studies test conducted in 2009.

REFERENCES

Adedoyin, O.O. (2010): Using IRT Approach to Detect Gender Biased Items in Public Examination: A case Study of the Botswana Junior Certificate Examination in Mathematics. *Educational Research and Reviews,* Vol. 5 no. 7. pg 385-399.

Amuche, C. I. & Fan, A. F. (2014): An Assessment of Item Bias using Differential Item Functioning Technique in NECO Biology Conducted examination in Tanabe State Nigeria. *American International Journal of Research in Humanities, Arts and Social Sciences*. Volume 14 No. 84, pg 95-100.

Ankerman, T.A (2002): Practical Applications of Item Response Theory, sponsored by the Department of Measurement, Statistics and Evaluation Seminar and Training Center at the University of Maryland. College Park.

Asiegbu, C.N. (2007). Influence of Attitude towards Vocational Education on Student's Achievements in vocational subjects in Rivers State. Unpublished M.Ed Thesis, University of Port Harcourt.

Baker, F.B. (2012). *The Basics of Item Response Theory. Eric Cleaning house on assessment and Evaluation*. University of Maryland and College part. Retrieved on 6/06/2012 from *http;//edres.orggirl./baker/.*

Baker, F.B., & Kim, Seock-Ho (2004*): Item Response Theory, parameter Estimation Techniques*. New York. Marcel Decker, Inc

Crooker, L. & Angina, J. (1980): *Introduction to Classical and Modern Test Theory.* New York: Holt Rinehart & Winston.

Ertuby, C. & Russell, R.J.H. (1996): *Dealing with Compatibility Problems of cross Cultural Data.* Paper presented at the 26th International Congress of Psychology. Montreal. August 1996.

Gregory, R.J. (2004): *Psychological Testing, Principles and Applications*. Boston. Allyn and Bacon.

Hambleton, R. & Rogers, J. (2012): *Item Bias Review.* Practical Assessment, Research, and Evaluation. Retrieved on 12/06/2012 from *http://PAIEonline.nrt/geton.aspv=48 n=six.*

Hambleton, R.K (1993): *Principles and Selected Applications of Item Response Theory.* In R.L Linn (Ed). Educational Measurement (3rd ed). Phoenix Arizona. American Council on Education and the Oryx press.

Hambleton, R.K. and Swaminathan, H. (1985) *Item Response Theory, Principles and Applications*. Boston. Kluwer-Nijhof Publishing co.

Harris, D. (2012); Comparison of One, Two, Three, parameter IRT Models. The Instructional topics in Educational Measurement Series. Pg 153-163. Retrieved on 06/06/2012 from *http://www.ncme.org/pubs/items.cfm*

Kuram, U.E. (2012); An Investigation of Item Bias in Science and Technology Subtests and Mathematics Subtest in Level Determination Exam (LDE) Retrieved on 6/06/2012 from *http://www.highbeam.com*

Lord, F.M. (1980); *Applications of Item Response Theory to Practical Testing Problems*. Mahwah, New Jersey. Lawrence Erlbaum Associates.

Marie, D. B. (2004): Use of Differential Items Functioning (DIF) analysis for Bias Analysis in Test construction. *SA Journal of Industrial Psychology*. Vol .30, No. 4 p. 52-58.

Mellenberg, G.J. (1994): A One-dimensional Trait Model for Continuous Item Response. Multivariate Behavioral Research. 29. pg 223-236.

Mokobi, T. & Adedoyin, O.O. (2014): Identifying Location Biased Items in the 2010 Botswana Junior certificate Examination Mathematics Paper One Using Item Response Characteristic Curves. *International Review of Social Sciences and Humanities.* Vol. 7, No 2 pg 63-82

Ojerinde, D. (2012): *Introduction to Item Response Theory, Parameter Models, Estimation and Application*. Abuja- Nigeria, Marvellous Press.

Osterlind, S.J (1983); *Test Item Bias*. Newbury Park, London. Sage Publications.

Osterlind, S. B. (2014): *Constructing Test Items*. Boston. Kluwer Academic Publisher.

Pedrajita, J.O. & Talisayon, M.Y. (2009): Identifying Biased Test Items by Differential Item functioning Analysis using Contingency Table Approaches. *Education Quarterly*, Vol. 67 no 1 pg 21-43.

Reckase, M. (1979): Unifactor Latent Trait Models Applied to Multifactor Tests: Results and Implications. Journal of Educational Statistics, Vol. 4, p 203 – 230.

Reeve, B.B. (2003): Item Response Theory Modeling in Health Outcomes and Measurement. *Expert Review of Pharmacoeconomics and Outcomes Research*, Vol. 3 no. 2. Pg 131-143.

Suen, H.K. (2010): *Principles of Test Theories*. Hillsdale. N.J. Lawrence Erlbaum.

Uremu, O. & Adams, O (2013): Differential Item Functioning Method as an Item Bias Indicator. *Journal of Educational Research*. Vol. 4 No 4. p. 367-373.

Warm, T.A. (2008): *A Primer of Item Response Theory (Technical Report*). Oklahoma City. US Coast Guard Institute.

Zhu, W. (2006): Constructing Test Items Using Item Response Theory in Wood T. and Zhu W, *Measurement Theory and Practice in Kinesiology.* Champaign. Illinois Human Kinetics.

Zumbo, D. B. (2012): A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-type Item Scores. Retrieved on 06/06/2012 from: *http://edu.ubc.ca/faculty/zumbo/dif/handbook.pdf.*

APPENDICES

Appendix 1: INITIAL EIGEN VALUES

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	12.006	20.009	20.009	12.006	20.009	20.009
2	3.406	5.676	25.685	3.406	5.676	25.685
3	3.162	5.271	30.956	3.162	5.271	30.956
4	3.129	5.215	36.171	3.129	5.215	36.171
5	2.810	4.683	40.853	2.810	4.683	40.853
6	2.684	4.474	45.327	2.684	4.474	45.327
7	2.472	4.120	49.447	2.472	4.120	49.447
8	2.224	3.707	53.154	2.224	3.707	53.154
9	2.115	3.526	56.679	2.115	3.526	56.679
10	1.894	3.157	59.836	1.894	3.157	59.836
11	1.718	2.863	62.700	1.718	2.863	62.700
12	1.605	2.675	65.375	1.605	2.675	65.375
13	1.501	2.501	67.876	1.501	2.501	67.876
14	1.379	2.299	70.174	1.379	2.299	70.174
15	1.291	2.151	72.326	1.291	2.151	72.326
16	1.256	2.093	74.419	1.256	2.093	74.419
17	1.195	1.992	76.411	1.195	1.992	76.411
18	1.139	1.898	78.308	1.139	1.898	78.308
19	1.111	1.852	80.160	1.111	1.852	80.160
20	1.050	1.750	81.911	1.050	1.750	81.911
21	.849	1.414	83.325			
22	.816	1.361	84.686			
23	.792	1.319	86.005			
24	.748	1.247	87.253			
25	.716	1.193	88.446			
26	.697	1.162	89.608			
27	.642	1.071	90.679			
28	.581	.969	91.647			
29	.486	.811	92.458			
30	.438	.729	93.188			
31	.413	.689	93.876			
32	.393	.655	94.532			
33	.359	.599	95.130			
34	.339	.565	95.695			
35	.316	.527	96.222			
36	.295	.492	96.714			
37	.288	.480	97.193			
38	.220	.367	97.560			
39	.199	.332	97.891			
40	.188	.313	98.205			
41	.164	.274	98.479			
42	.145	.241	98.720			
43	.133	.221	98.941			
44	.119	.198	99.139			

45	.094	.156	99.296			
46	.086	.143	99.439			
47	.067	.112	99.551			
48	.059	.099	99.650			
49	.050	.083	99.732			
50	.041	.068	99.800			
51	.032	.053	99.853			
52	.027	.046	99.899			
53	.022	.036	99.935			
54	.014	.024	99.958			
55	.009	.015	99.973			
56	.007	.012	99.985			
57	.005	.008	99.993			
58	.004	.006	99.999			
59	.000	.001	100.000			
60	-1.109E-15	-1.848E-15	100.000			

Extraction Method: Principal Component Analysis.

Hypotheses 2: Appendix 2 shows the item parameters of the 60 test items while fig 3 (1 – 8) shows the ICCs of the items that are not biased for the male and female students that sat for the JSSCE Business Studies in 2009.

2009							
	FEMALE				MALE		
Items	A	B	c		A	b	C
1	2.786	-1.063	0.634				
2	2.172	1.101	0.134		2.172	1.101	0.134
3	0.692	-0.068	0.400		0.763	-1.987	0.015
4	6.000	1.042	0.323		0.798	-0.573	0.082
5	4.139	1.767	0.342		6.000	0.847	0.474
6	0.528	-0.209	0.317		1.702	-0.992	0.224
*7	2.153	-1.176	0.252		2.144	-1.172	0.282
8	0.723	0.065	0.255		0.583	1.339	0.203
9	1.590	0.539	9.142		0.210	1.448	0.250
10	3.995	1.877	0.359		2.202	0.892	0.112
11	2.240	0.580	0.091		4.266	1.870	0.315
12	0.953	-0.264	0.258		0.257	-2.591	0.274
13	1.983	3.181	0.076		4.443	1.844	0.136
14	0.227	0.994	0.296		0.434	0.290	0.285
15	6.000	0.992	0.246		3.969	1.453	0.280
16	1.971	3.175	0.076		3.969	1.453	0.280
17	6.000	1.022	0.181		6.000	0.883	0.392
18	4.260	1.228	0.234		1.470	0.574	0.098
19	1.975	3.174	0.194		4.392	1.403	0.176
20	4.368	1.244	0.272		6.000	1.116	0.177
21	6.000	1.274	0.242		2.129	1.138	0.157
22	0.678	-0.213	0.238		0.262	-0.504	0.272
23	1.445	1.069	0.198		2.353	0.009	0.222
24	1.886	3.160	0.292		4.448	1.846	0.134
*25	1.896	3.174	0.324		1.682	3.654	0.249
26	6.000	0.017	0.120		5.736	0.166	0.326
*27	6.000	0.886	0.288		6.000	0.886	0.288
*28	2.971	0.230	0.303		2.971	0.230	0.303
29	1.980	3.171	0.132				
30	5.786	1.266	0.208		4.340	1.849	0.161
31	3.263	0.741	0.083		2.962	-0.717	0.228
*32	6.000	0.852	0.211		5.003	0.864	0.129
33	1.945	3.170	0.161		0.917	2.526	0.073
34	0.338	1.644	0.217		0.920	-0.519	0.283
35	0.576	0.304	0.254		1.594	0.514	0.094
36	0.887	0.219	0.251		4.184	0.422	0.419
37	1.330	1.600	0.098		6.000	1.115	0.241
38	0.419	0.595	0.244		3.066	1.807	0.387
39	3.925	1.720	0.207		1.396	1.264	0.361
*40	6.000	1.047	0.290		6.000	1.047	0.290
41	1.794	3.164	0.193		5.489	0.831	0.338
42	4.764	0.522	0.074		6.000	1.126	0.310
43	1.003	1.279	0.162		1.577	3.654	0.242
44	0.476	1.238	0.260		1.245	-0.030	0.264
45	6.000	0.697	0.433		0.709	-1.743	0.259
46	6.000	0.919	0.124		0.317	2.522	0.228
47	2.853	0.584	0.220		0.405	-1.719	0.260
*48	2.853	0.584	0.220		2.853	0.584	0.220
49	0.595	2.309	0.144		6.000	1.133	0.387
50	0.736	1.026	0.161		0.196	2.311	0.295
51	1.901	3.165	0.223		3.772	1.887	0.387
52	1.020	2.185	0.189		4.475	1.840	0.079
53	1.831	3.156	0.325		0.147	1.529	0.268
*54	1.008	1.378	0.217		1.008	1.378	0.217
55	0.164	0.372	0.280		0.531	-0.538	0.263
56	0.119	2.669	0.312		6.000	0.935	0.254
57	0.866	0.640	0.219		0.407	-0.542	0.222
58	6.000	0.542	0.195		1.615	3.654	0.189

59	1.985	3.166	0.075				
60	1.801	0.859	0.135		0.511	-0.487	0.281

*items that have similar ICC. Table 2

Appendix 3: items that have similar ICC.

2009							
	RURAL				URBAN		
Items	A	B	C		a	b	C
1	6.000	-1.512	0.186				
2	6.000	0.476	0.102		5.069	1.422	0.149
3	0.393	-1.734	0.334		0.976	-1.808	0.148
*4	1.281	0.394	0.151		1.281	0.394	0.383
5	1.077	1.089	0.256		5.069	0.956	0.401
6	0.824	-1.078	0.277		1.142	-0.134	0.402
7	1.026	-1.566	0.276		0.999	-2.008	0.212
8	0.594	0.230	0.180		0.190	2.006	0.255
9	2.967	1.063	0.142		0.459	1.015	0.138
10	2.967	1.063	0.142		0.387	1.950	0.166
11	1.312	0.478	0.187		0.383	3.624	0.182
12	1.000	-0.781	0.241		0.329	-1.390	0.229
13	5.725	1.780	0.117		1.281	4.000	0.085
14	3.996	0.389	0.549		0.190	0.492	0.302
15	6.000	1.193	0.168		1.285	4.000	0.202
16	5.498	1.791	0.192		5.069	1.454	0.164
17	6.000	0.894	0.237		5.069	1.256	0.356
18	6.000	1.007	0.170		0.508	1.565	0.132
19	5.224	1.809	0.296		5.069	1.441	0.093
20	6.000	1.018	0.251		4.187	1.828	0.185
21	4.180	1.129	0.192		4.187	1.828	0.185
22	0.733	-0.969	0.239		0.240	0.306	0.245
23	1.386	-0.064	0.227		4.828	0.536	0.136
24	5.394	1.797	0.225		1.302	4.000	0.203
*25	1.781	4.000	0.322		1.779	4.000	0.276
26	4.474	0.078	0.196		0.816	-0.401	0.153
27	5.139	0.903	0.282		5.069	1.164	0.267
28	1.743	0.222	0.160		0.740	0.065	0.201
29	1.830	3.396	0.085		1.334	4.000	0.133
30	6.000	1.196	0.204		1.212	4.000	0.151
31	2.640	-0.077	0.131		5.069	0.132	0.097
32	6.000	1.012	0.213		1.158	0.535	0.099
33	1.829	3.396	0.085		2.646	1.688	0.151
34	0.470	0.747	0.216		0.606	-0.489	0.190
35	1.446	1.038	0.313		0.290	1.528	0.210
36	3.460	0.474	0.348		1.131	0.411	0.412
37	2.274	1.173	0.112		1.172	1.151	0.120
38	0.671	1.240	0.227		0.106	0.430	0.285
39	0.610	1.158	0.204		1.638	1.266	0.188
40	6.000	1.199	0.244		1.638	1.266	0.188
41	4.934	1.804	0.258		5.069	0.806	0.224
42	1.797	0.454	0.130		0.604	1.161	0.138
43	0.734	2.973	0.193		0.483	1.987	0.143
44	0.812	0.100	0.268		5.069	0.700	0.377
45	0.599	-0.400	0.262		0.749	-1.145	0.242
46	1.083	1.203	0.112		0.282	2.740	0.199
47	2.269	0.383	0.378		1.458	-0.334	0.134
48	6.000	0.683	0.238		0.668	1.005	0.216
49	1.091	1.852	0.295		0.643	0.977	0.145
50	0.357	2.078	0.243		0.480	0.387	0.174
51	5.162	1.811	0.294		1.075	3.936	0.321
52	5.346	1.422	0.061		1.310	4.000	0.180
53	1.799	3.396	0.286		0.418	0.226	0.179
54	0.662	0.567	0.206		5.069	0.186	0.191
55	0.214	0.414	0.290		0.869	-0.144	0.319

56	2.160	1.318	0.280		5.069	1.202	0.273
57	0.497	0.530	0.178		1.675	0.137	0.307
58	0.710	1.441	0.147		1.137	4.000	0.277
59					1.339	4.000	0.067
60	6.000	1.021	0.347		1.024	-0.067	0.126