

# **ASSESSMENT OF DIFFERENTIAL ITEM FUNCTIONING TO DETECT GENDER BIASED ITEMS IN ECONOMICS MULTIPLE CHOICE QUESTIONS IN SENIOR SCHOOL CERTIFICATE EXAMINATION**

**<sup>1</sup>Francis Elochukwu Ikeh, <sup>1</sup>Catherine C Ene, <sup>1</sup>Blessing Ojobo, <sup>4</sup>Mercy Ifunanya Ani, <sup>2</sup>Clementina Ifeoma Metu, <sup>1</sup>Felicia Chinyere Ugwu, <sup>1</sup>Olaniyi Owolawi, <sup>3</sup>Victoria Olasoji Omosowon, <sup>1</sup>Basil C Oguguo, <sup>1</sup>Jude Ifesinachi Ezugwu, <sup>1</sup>Janehilda Oluchi Agugoesi**

<sup>1</sup>Department of Science Education, University of Nigeria, Nsukka, Nigeria

<sup>2</sup>Department of Educational Foundations, Nnamdi Azikiwe University, Awka, Nigeria

<sup>3</sup>Department of Science Education, Enugu State College of Education Technical, Enugu Nigeria

<sup>4</sup>Department of Educational Foundations, Alex Ekwueme Federal University, Ndufe Alike Ikwo, Ebonyi State, Nigeria

**Corresponding Author:** Clementina Ifeoma Metu (ic.metu@unzik.edu.ng)

## **Abstract**

The study examined the application of Differential Item Functioning (DIF) in detecting gender biased items in Economics multiple choice questions in senior school certificate examination. Causal comparative or Ex-post factor research design was used for the study. The population for the study comprised 2,985 Senior Secondary School Three (SS3) Economics students in Nsukka education zone of Enugu State. Sample size of 339 SS3 Economics students was used for the study. The instrument used for the study was 2018 SSCE 50-item multiple Economics choice questions developed by West African Examination Council (WAEC). Reliability coefficient of 0.87 was obtained using Kuder-Richardson formula. The data collected from the study were analyzed using Logistic Regression procedure. The study revealed that: out of 50 items in 2018 WAEC Economics questions, 14 items which represent 28% of the test items displayed significant gender DIF at the 0.05 level of significance. Among the 14 items that displayed significant gender DIF, only 1 item representing 2% were identified to exhibit significant gender DIF in favour of male students while 13 items representing 26% differentially functioned in favour of female students. Based on the findings of the study, it was recommended among others that DIF analysis should be conducted by test experts and test developers on their test items. This will help to reveal items that exhibit Differential Item Functioning which could be revised or eliminated in order to produce items that are fairness.

**Keywords:** Differential Item Functioning (DIF), Gender, Item Bias, Economics

## **Introduction**

In every classroom setting, the teacher's efforts are to ensure adequate achievement of the set goals or objectives of the lesson content. According to Ugwuja and Igbokwe (2009), the extent to which the teacher is able to achieve the goals of any lesson content is a reflection of the degree of achievement by the teacher in attaining the expected change(s) in the behaviours of the learners after instruction. These expected changes in the behaviours of the learners are measured using different assessment tools like test. Test is one of the assessment instrument used in assessing students' academic achievement in any given instruction. Test according to Nworgu (2011) consists of sets of uniform questions or tasks with preferred response(s) to which a student or examinee is to respond independently and the result of which can be treated in such a way as to provide a quantitative comparison of the performances of different students.

Test is used to determine how much the students have covered or learned a particular content. According to Joshua (2005), tests are used to gain useful information about test-takers' knowledge, skills and progress; it helps each professional to perform his or her work effectively; are used in promotion, placement, selection, certification and decision making. Students, individuals and group of individuals who are tested by standardized or teacher-made tests are known as examinees and in testing, it is expected that examinees of the same latent trait should respond to test item correctly irrespective of their different subgroups (Ogbebor, 2012). This implies that a test has to be fair to all the examinees because a fair test according to Roever (2005) affords all examinees an equal opportunity to demonstrate the skills and knowledge which they have acquired and which are relevant to the test purpose. However, irrespective of efforts, time and consideration put in the construction of test by test developers and experts in test construction; it may not be perfect (Walker, 2011). This is because they may be susceptible to human errors and on analysis they may be found to be ambiguous, too simple, overly difficult and discriminating. This imperfection in the constructed test may result to test bias.

Test bias is the existence of some irrelevant elements present in items that causes differences in the achievement for student of the same ability but from different subgroups like gender (Ogbebor & Onuka, 2013). A test item is considered bias if compared to other test items and it is relatively more difficult for one group than for another (Abedalaziz, Leng & Alahmadi, 2014). The presence of bias in an item is a cause for concern for educators since tests are used as a gatekeeper for educational opportunities, and it is a very important issue that test items are fair for every examinee (Adedoyin, 2010). In line with the statement, Ahmad, Mokshein and Husin (2018) opined that the presence of biased items is alarming as testing is usually used as a controller for educational opportunities. This is an indication that students who get a minimum achievement in any given examination have chances of continuing their studies while unsuccessful students who failed to reach the minimum requirement may be restricted from continuing their studies (Ahmad, Mokshein & Husin, 2018). It is therefore, a clear indication that test items should be fair to every student. This is because a fair test gives all examinees the opportunity to demonstrate the skills and knowledge they acquired in relation to the purpose of the test (Adedoyin, 2010). Therefore, a test should be constructed in such a way that even if there are differences in the academic achievements of students, it cannot be attributed to the fact that those students belong to a particular subgroup like gender.

Gender is the psychosocial aspect of maleness and femaleness. According to Keightley (2011), gender is concerned with masculinity and femininity as categorized to each sex in the society. It is a range of characteristics used to distinguish between male and female, particularly in the cases of men and women, masculine and feminine attributes assigned to them. In line of the statement, Bronfenbrenner (2005) refers to gender as social differences and relations between men and women. Gender is a social construct, it is not biologically determined but a concept equivalent to race or class (Offorma, 2004). This definition suggests that gender is socially or culturally constructed characteristics and role, which are associated with males and females in the society. According to Lee (2001), gender is ascribed attribute that differentiates feminine from masculine.

Different studies as reported below pointed out that females usually make external attributions for successes and failures, and that when they make internal attributions, these refer not so much to effort, but to ability. However, males usually attribute success to stable internal causes like effort, thus showing an attribution pattern which enables them to enhance their own image of self (Smith, Sinclair & Chapman, 2002). The differences in the achievements of males and females are generically attributed to biological cause and or to culture and stereotypes (klein, 2004). Some studies (Simkim & Kuechler, 2005; Jiboku 2008) have shown low performance of female students in education. Educators have therefore expended tremendous efforts in the study of the personal factors affecting academic achievement especially in the sciences and social sciences. Eryilimaz (2004) observed that gender contributes in achievement of students in Economics.

However, the issue of gender differences in students' performance in their academics has raised the concern of various researchers. For instance, Orluwene and Otuata (2017) in their study reported that gender significantly influenced the differential effects of objective test formats on students' achievement in Economics. Amuda, Domiya and Durkwa (2016) reported that there was no significant gender difference in the academic performance of students in SSCE, WAEC and NECO in Economics, while SSCE, WAEC and NECO for

2006/2007 session in Economics for 2008 to 2010 sessions showed consistent significant gender difference in the academic performance of students in favour of male students. Contrarily, Okpala (2016) reported no significant difference between male and female senior secondary school students' achievement in Economics. These differences in the performance of male and female students that are likely to be both content and ability dependent are regarded as Differential Item Functioning (DIF).

Differential Item Functioning is a statistical tool used to assess the existence of item bias in a test. DIF is a statistical technique used in identifying differential item response patterns between groups of examinees such as gender (male and female) as well as verifying potentially biased items (Madu, 2012). Chahine and Childs (2010) is of the view that DIF analysis is typically used to identify test items that are differentially difficult for respondents who have the same level of knowledge, skill or ability but differ in ways that should be irrelevant to their performance on the test (e.g., females versus males, students in rural versus urban schools, students in public schools versus students in private schools among others). Scott et al (2010) opined that differential item functioning methods are a range of techniques that are increasingly being used to evaluate whether different subgroups respond differently to particular items within a scale, after controlling for group differences in the overall assessment. For Abedi, Leon and Kao as cited in Queensoap and Orluwene (2017), DIF analysis is often used to examine group differences between specific racial or ethnic groups or between male and female. DIF items can lead to biased measurement of ability because the measurement is affected by so-called "nuisance factors" (Ackerman as cited in Özdemir, 2015).

Awareness of this bias is of great importance to educators where scale scores are used to investigate gender differences and ensure that derived scores are comparable across groups (Zampetakis, Bakatsaki, Litos, Kafersios & Moustakis, 2017). This may be the reason why Pedrajita (2009) opined 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, Oshima and Morris (2008) stated that DIF analyses are vital in the field of test and measurement since it sets to address equivalence across subgroups of examinees which can occur in external examinations like West African Examination Council (WAEC). This examination is taken by different examinees of the same ability from different subgroup such as culture, religion, ethnicity, gender among others. Moreover, since this examination is used for gaining admission into the higher institutions, one will begin to as if the test items are fair enough for all groups. On this basis, it is therefore important to apply differential item functioning (DIF) to detect gender biased items in Economics multiple choice questions in senior school certificate examination.

## **Method**

This study employed causal comparative or ex-post facto research design. This is because the researchers worked with non-manipulative independent variable such as gender. The population of the study comprised 2563 Senior Secondary School (SSS) III Economics students in Nsukka Education Zone of Enugu State, Nigeria. The population comprised 1141 males and 1422 female students. The sample size of 444 SSS3 Economics students drawn from government 46 co-education schools in Nsukka Education Zone of Enugu State, Nigeria was used for the study. The sample comprised 198 male and 246 female students.

The instrument for data collection was 2018 multiple choice Economics questions adopted from West African Examination Council (WAEC). The instrument consists of two sections; section A and section B. Section A focused on the demographic data of the respondent such gender. Section B of the instrument consists of 50-item questions. Each item consists of a question and a list of possible answers lettered A - D of which only one option is the correct answer. The test items of the instrument were scored 1 for correct option and 0 for wrong option with maximum score of 50 and minimum of 0.

The instrument were constructed and validated by experts in the department of examinations and quality control of the West African Examination Council (WAEC) and therefore require no further validation since the instrument was adopted. To determine the reliability of the instrument, the instrument was administered to 25 Economics students in Nsukka education zone of Enugu State who were not part of the study. The responses obtained from the students were subjected to Kuder-Richardson 20 (KR-20) formula to determine the internal

consistency of the instrument. A reliability coefficient of 0.93 was obtained. The choice of Kuder-Richardson (KR-20) was because the instrument was dichotomously scored.

To collect pertinent data needed for the study, instrument was administered to SSS3 Economics students in each of the sampled schools with the help of the Economics teachers in the sampled schools. To ensure that students put in their best, the researchers through their Economics teacher informed the students ahead of time about the exercise and the need to be prepared because it formed part of their continuous assessment. Logistic Regression Analysis Procedure was used to analyze the data obtained. It was used to answer the research questions posed for the study.

**Result**

**Research Question**

What test items on standardized WAEC 2018 SSCE May/June multiple choice Economics questions function differentially in terms of gender?

**Table 1:** DIF Analysis to Detect Gender Bias on 50 Multiple Choice Economics Questions

Item	B	S.E	Wald	Sig.	Exp (B)	95.0% C.I. for Exp (B)	
						Lower	Upper
1	.29	.20	2.27	.132	1.34	.92	1.96
2	.16	.20	.62	.432	1.67	.79	1.72
3	.43	.19	4.88	.027*	1.53	1.05	2.23
4	-.22	.20	1.23	.268	.81	.55	1.18
5	-.09	.19	.22	.643	.91	.63	1.34
6	-.03	.19	.02	.895	.98	.67	1.42
7	.10	.20	.24	.625	1.10	.75	1.63
8	-.38	.19	3.81	.051	.69	.47	1.00
9	.33	.20	2.57	.109	1.38	.93	2.06
10	.83	.23	13.24	.000*	2.30	1.47	3.59
11	.24	.20	1.36	.243	1.27	.85	1.89
12	.25	.20	1.57	.211	1.29	.87	1.91
13	.14	.20	.54	.462	1.16	.79	1.69
14	.20	.21	.94	.333	1.22	.81	1.85
15	.25	.19	1.67	.196	1.28	.88	1.87
16	.26	.19	1.87	.172	1.30	.89	1.89
17	.30	.19	2.38	.123	1.34	.92	1.96
18	.04	.20	.04	.838	1.04	.71	1.53
19	.30	.21	2.07	.150	1.34	.90	2.01
20	.43	.23	3.62	.057*	1.54	.99	2.39
21	-.07	.19	.12	.731	.94	.64	1.36
22	.18	.21	.74	.389	1.19	.80	1.79
23	.41	.20	4.37	.037*	1.51	1.03	2.22
24	.18	.19	.89	.345	1.20	.82	1.75
25	.13	.19	.47	.494	1.14	.78	1.67
26	-.24	.19	1.56	.211	.79	.54	1.15
27	-.02	.20	.01	.925	.982	.67	1.44
28	.59	.20	8.57	.003*	1.80	1.22	2.68
29	.45	.20	4.91	.027*	1.57	1.05	2.33

30	.37	.19	3.72	.054	1.45	.99	2.12
31	.22	.20	1.31	.252	1.25	.85	1.83
32	.53	.20	7.10	.008*	1.70	1.15	2.52
33	.45	.21	4.36	.037*	1.56	1.03	2.37
34	.02	.19	.008	.928	1.02	.70	1.49
35	.34	.20	2.85	.092	1.41	.95	2.09
36	-.60	.22	7.73	.005*	.55	.36	.84
37	-.07	.19	.12	.735	.94	.64	1.37
38	-.13	.19	.44	.506	.88	.60	1.28
39	.53	.20	6.74	.009*	1.70	1.14	2.53
40	.51	.19	6.84	.009*	1.66	1.14	2.42
41	-.10	.19	.24	.621	.91	.62	1.33
42	.67	.22	4.56	.033*	1.60	1.04	2.46
43	.15	.20	.56	.456	1.16	.78	1.74
44	.19	.19	.95	.329	1.21	.83	1.76
45	.09	.19	.20	.658	1.09	.75	1.58
46	.41	.20	4.09	.043*	1.51	1.01	2.24
47	.16	.19	.719	.397	1.18	.81	1.71
48	.46	.20	5.34	.021*	1.58	1.07	2.33
49	.21	.19	1.25	.263	1.24	.85	1.80
50	.58	.19	9.04	.003*	1.79	1.23	2.62

Result on Table 1 revealed that out of 50 items in WAEC 2018 SSCE May/June multiple choice Economics questions; DIF was present in 14 items. These items are item 3, 10, 23, 28, 29, 32, 33, 36, 39, 40, 42, 46, 48 and item 50. These items revealed significant DIF between male and female students with significant level less than 0.05. Among the 14 items that displayed significant gender DIF, only 1 item (item 36) representing 2% were identified to exhibit significant gender DIF in favour of male students while 13 items (3, 10, 23, 28, 29, 32, 33, 39, 40, 42, 46, 48 and item 50) representing 26% differentially functioned in favour of female students.

### **Discussion of Finding**

#### **WAEC Economics Multiple Choice Test Items of the Senior School Certificate Examination that function differentially in terms of Gender**

The results of analysis of students' responses to Economics multiple choice test items used in WAEC SSCE in 2018 revealed that Economics tests contain some items with significant gender differential functioning. This implies that the test contain items that measured different things for male and female students with the same ability in Economics. From the findings, these items that showed DIF may be due to the structure of the questions and stem, thus; these could be the characteristics that affected the test takers response to getting the item correctly. This is in line with the submission of Nworgu (2011) who stated that current research evidence has implicated test used in national and regional examination as functioning differently with respect to different subgroups. This means that students' scores in such examinations are determined largely by the group to which an examinee belongs and not by ability.

The findings of the study is in agreement with the findings of Abedlaziz, Ismail and Hussin (2011) whose finding reports that out of 30 multiple choice mathematics items administered to grade ten students in Jordan, 17 items showed DIF based on gender. Also in agreement, was the study conducted by Odili (2004) who reported that in 1999, 2000 and 2001 Biology multiple choice questions, 8, 2 and 3 items respectively, differentially functioned significantly for examinees based on gender. Also in agreement was the finding of Madu (2012) who reported that out of 50 test items set by WAEC in mathematics, male and female examinees perform differentially in 39 items and

none in 11 items. The findings of Pedrajita (2009), Obinne and Amali (2013) also found in their respective studies that the test items contained items that functioned differentially for male and female students. Meanwhile, the findings of Umoinyang, Ndifon and Idiku (2014) that assessed gender based differential item functioning of 2010 junior secondary school mathematics examination in Southern education zone of Cross River State disagreed with the findings of the present study. The findings showed that none of the detection method identified items that functioned differentially between male and female students.

### **Conclusion**

Based on the findings of the study, the study concludes that Economics multiple choice questions used by WAEC in 2018 are gender biased as they functioned differentially for examinees in Nsukka education zone of Enugu State. This is because majority of the items showed evidence of gender differential item functioning.

### **Recommendations**

Based on the findings of this study, the following recommendations were made by the researcher:

1. Differential item functioning analysis should be conducted by test experts and test developers like West Africa Examination Council, National Examination Council among others on their test items. This will help to reveal items that exhibit Differential Item Functioning which could be revised or eliminated in order to produce items that are fairness.
2. Government through the ministry of education should ensure that test items in standardized tests are free of DIF across factors such as gender, location, culture, ethnic group and religion background. This will help to ensure that no segment will be unfairly panelized when taking standardized tests.
3. Examination bodies especially West Africa Examination Council should write test items in Economics that would not favour one group against the other. They should be sensitive to the heterogeneous nature of Nigeria.

### **REFERENCES**

- Abedalaziz, N., Leng, C. H. & Alahmadi, A. (2014). Detecting a gender-related differential item functioning using transformed item difficulty. *The Malaysian Online Journal of Educational Science*, 2 (1) 16-22
- 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 Reviews*, 5 (7), 385-399,
- Ahmad, H., Mokshein, S. E., & Husin, M. R. (2018). Detecting Item Bias in an Anatomy & Physiology Test for Nursing Students using Item Response Theory. *International Journal of Academic Research in Progressive Education and Development*, 7(1), 97–109
- Amuda, B. G., Domiya, G. A. & Durkwa, H. (2016). *Gender difference in academic performance in SSCE Economics subject among senior secondary school students in Maiduguri Metropolis, Borno State, Nigeria*. Retrieved from <http://pubs.sciepub.com/education/4/3/10/index.html>
- Bronfenbrenner, S.Y. (2005). Self-concepts, domain value, and self esteem: Relations and changes at early adolescence. *Journal of Personality*. 59, 224-232
- Chahine, S. & Childs, R.A. (2010). Detecting differential item functioning and differential step functioning due to differential that should matter. *Journal of Practical Assessment, Research and Evaluation*, 15(10)
- Eryilimaz, H. (2004). *The effect of peer instruction on high school students' achievement and attitude towards Physics*. PhD Thesis of Middle East University
- Jiboku, A.O. (2008). Gender and self-concept as predictors of academic self-efficacy of students. *Ogun Journal of Counselling Studies*. 2 (2), 89-94.
- Joshua, M.T. (2005). Test item bias in psychological testing: Evidence in Nigerian system. *A paper presented at the annual Conference of Nigerian Association of Educational Psychologists held at Ahmadu Bello University Zaria from 24th-28th March*.

- Keightley, J. (2011). Influence of gender identities on achievement of Boys and Girls in Schools. *Economics of Education Review*. 24, 410-431
- Klein, J. (2004). Who is most responsible for gender differences in scholastic achievements: Pupils or teachers? *Educational Research*, 46 (2) 183 –193.
- Lee. J. (2001). Inter State variation in rural students’ achievement and schooling conditions. Retrieved from <http://www.ericdigest.org/2002>
- Madu, B. C. (2012). Using transformed item difficulty procedure to assess gender-related differential item functioning of multiple-choice mathematics items administered in Nigeria. *Research on Humanities and Social Sciences*, 2 (6)41-55
- Nworgu, B. G. (2011). Differential item functioning: A critical issue in regional quality assurance. *Paper presented in National Association of Educational Research and Assessment conference*.
- Offorma, G. C. (2004). Language and gender. *International Journal of Arts and Technology Education*, 3, 62-75.
- Ogbebor, U. & Onuka, A. (2013). Differential item functioning of economics question papers of National Examinations Council in Delta State, Nigeria, *Nigerian Journal of Educational Research and Education*, 12 (1) (2013), 45-60.
- Ogbebor, U.C. (2012). Differential item functioning of biology question paper of National Examinations Council in Taraba State Nigeria. (Unpublished M.Ed project), University of Ibadan, Oyo state Nigeria
- Okpala, L. C. (2016). *Influence of school type and gender on secondary school students’ academic achievement and interest in Economics in Enugu state, Nigeria*. Retrieved from <https://oer.unn.edu.ng/read/influence-of-school-type-and-gender-on-secondary-school-students-academic-achievement-and-interest-in-economics-in-enugu-state-nigeria?rdr=1>
- Orluwene, G. W. & Otuata, E. A. (2017). Gender influence on the efficacy of multiple-choice, alternate response and completion objective test formats on students’ achievement in Economics. *European Journal of Research and Reflection in Educational Sciences*, 5 (1), 71-80.
- 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.
- Özdemir, B. (2015). A comparison of IRT-based methods for examining differential item functioning in TIMSS 2011 mathematics subtest. *Procedia - Social and Behavioral Sciences* 174 (2015) 2075 – 2083
- 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. & Orluwene, G. W. (2017). Examining differential item functioning in a chemistry achievement test for students in Nigeria. *International Journal of Education and Evaluation*, 3 (7) 49-57
- Roever, C. (2005). “That’s not fair!” *Fairness, bias, and differential item functioning in language testing*. Retrieved from <http://www2.hawaii.edu/~roever/brownbag.pdf>
- Scott1, N.W., Fayers, P.M., Aaronson, N.K., Bottomley,A., Graeff, A., Groenvold, M., Gundy, C., Koller, M., Morten, A.P.. & Sprangers, M.A.G. (2010). Differential item functioning (DIF) analyses of health-related quality of life instruments using logistic regression. *Health and Quality of Life Outcomes* 8 (81) 1-9
- Simkim, M. G. & Kuechler, W. L (2005). Multiple choice test and students understanding: What is the connection? *Decision Sciences Journals of Innovative Education*. 3(1), 73-91.
- Smith, L., Sinclair, K.E. & Chapman, E.S. (2002). Students’ goals, self-efficacy, self handicapping, and negative affective responses: An Australian senior school student study. *Contemporary Education Psychology*, 27, 471-485.
- Ugwuja, J. & Igbokwe, U. L. (2009). Classroom assessment. In U. L. Igbokwe & U. N. Eze (Eds.). *Classroom management for curriculum implementation: Applying psychological principles* (210-228). Enugu: Timex Publisher.

- Walker, C. (2011). What's the DIF? Why differential item functioning analyses are an important part of instrument development and validation. *Journal of Psychoeducational Assessment*, 29, 364-376
- Zampetakis, L.A., Bakatsaki, M., Litos, C., Kafetsios, K.G. & Moustakis, V. (2017). Gender-based differential item functioning in the application of the theory of planned behavior for the study of entrepreneurial intentions. Retrieved from <https://pubmed.ncbi.nlm.nih.gov/28386244>