Research Article

TEST ITEM FAIRNESS FOR ASSESSMENT OF MALE AND FEMALE STUDENTS' ACHIEVEMENT IN ECONOMICS

Patrick. U. Osadebe*

Department of Guidance and Counselling, Delta State University, Abraka

A R T I C L E   I N F O

A B S T R A C T

The study was carried out to develop test item fairness for the assessment of male and female students' achievement in Economics. The study covered the senior secondary school Economics curriculum before the test administration. The study was guided with research question and hypothesis for each item. A sample of 200 (100males and 100females) was randomly selected using balloting and stratified random sampling techniques. The instrument used was a Standardized Economics Achievement Test (SEAT). The test has face and content validities and a reliability coefficient of 0.85 was obtained through Kuder Richardson formula 20 as a measure of internal consistency. It has 50 items. The test was adapted and only 40 items were fair to male and female students when it was revalidated outside the area of study. A test-retest reliability was done to ensure stability of the items. A reliability coefficient of 0.92 was obtained. The data collected were analysed in line with the research question and hypothesis for each item. Percentage was used to analyse the research questions while Z-test was used to test the hypothesis for each item at 0.05 level of significance. The result revealed that the 40 items used showed fairness for male and female students. There was no significant difference between the proportion of male and female students passing each item in Economics. It was recommended among others that any item that discriminate or not fair to testees should be removed from the test before use. The test should be used from time to time for the assessment of male and female students because of the item fairness. An item should be fair to both group of testees before use.

INTRODUCTION

Most test items are not usually fair to different groups of testees. These group include testees' sex (male and female), location (rural and urban), among others. With regard to testees' performance, a test item could be fair to male and not fair to female or same to female and not to male or both. This problem called for investigation as in this study. A test item is expected to be fair to both male and female group of testees. One of the most important aspect of test construction and administration is to ensure that each test item is fair to testees in terms of their scores or performance across sex, location, school-type, race among other groups. It has been pointed out that if the purpose of a test is to determine students' ability and improve on performance, then it should not discriminate the group of testees (Osadebe, 2015). Therefore, test item fairness should be considered during test construction and administration. It is a situation where a test item does not discriminate across testees of different groups in terms of performance. These group of testees could be male and female, rural and urban, black and white among others.

Corresponding author: **Patrick U. Osadebe**
Department of Guidance and Counselling Delta State University, Abraka

The greatest problem in test administration or psychological testing is the use of a test whose items are not fair to students based on their gender (male and female), location (rural and urban), school type, school ownership, race among others. This has often led to test items bias or differential item functioning. Therefore, there was the need to ensure that test items are fair to students before the administration. A test is made up of items. The items should be fair to the group of testees in terms of performance or scores. There should be no significant difference between the performance of male and female testees, rural and urban among others. This will justify test item fairness across the different group of testees in terms of performance. A lot is done during item analysis to ensure that each item is valid and reliable and does not discriminate the testees in terms of their performance (Osadebe, 2001). Similarly, during the administration of the test at any point in time, the test item should be fair to the testees across their groups after the scoring or measurement of their performance.

It should be noted that item analysis is only done with a small sample, and may not be from the population that will attempt the test items. However, item analysis could help to detect some items that are not fair to the testees when it is specifically analysed with the variable under investigation, such as the proportion of male and female passing an item with regards to

difficulty index. Items that discriminate or not fair to male and female are removed from the test. Item analysis helps to establish item validity.

The administration of the test items is to the sample from a given population. This cut across the testees' gender, location among others. It is necessary to remove the items that are not fair to the testees acrossthe groups in terms of performance. The final items should be composed to form a fair test item before estimating the reliability of the test.

Literature Review

Test item fairness is a situation where each test item does not discriminate testees in terms of their performance or score across sex (male and female), location (rural and urban) among others. The test Items should favour male and female testees as in this study. The test item fairness could be observed when the test mean score or measurement of testees in both groups such as male and female do not differ significantly. It is expected in a fair test items that the testees should demonstrate close or similar scores in a given test items. Therefore, there should be no significant difference between the scores of male and female testees in each test item. In ensuring test items fairness, items that discriminate are removed while items that favour both groups are retained. Hence, the valid and reliable items are composed to form a fair test. It should be noted that fair items could produce fair test. After the composition of the fair test items, the reliability the test should be determined. Therefore, fairness in a psychometric context is the extent to which a test is used in a just and equitable method. Ideally, the test developer ensures fairness in the test development process and the usage (Swerdlk, 2009).

Generally, assessment of students focuses on cognitive, affective and psychomotor areas of human behaviour. These areas should be assessed as recommended by Aiken (1979), Ministry of Education, Science and Technology (1985), Ukwuije (1996) and the Federal Government of Nigeria (2004). Assessment in this study was the use of a Standardised Economics Achievement Test to make judgement about test item fairness. It was a cognitive assessment which ensured item fairness for male and female students in Economics. Since the items are fair, it could be used from time to time to examine male and female students in continuous assessment. Assessment could be single or continuous. A valid and reliable test should be used in both mode of assessment as pointed out by Osadebe (2015). Assessment leads to judgement or decision derived from a given data. This is similar to the work of Osadebe and Obiunu (2005).

Test item fairness is a situation where the test item favour both group of testees such as male and female, rural and urban among others. There should be no significant difference between the group in terms of their performance. It has been pointed out that fairness includes routine analysis of test questions to establish that questions do not unfairly contribute to group differences (Educational Testing Service, 2015). This was in line with the study because items were analysed to ensure that there was no significant difference between male and female students' performance. This helped to determined test item fairness for male and female testees. According to the Code of Fair Education (2014), test developers are expected to provide test that the items are fair to all test takers based on gender, location, race among others. However, the study focussed on gender, which includes the male and female testees.

In the study of test fairness, Fishbein (1977) had found that items are fair for examinees if their achievement levels are equated statistically. The study was carried out to determine test item fairness between black and white examinees. However, the study focussed on test item fairness for male and female testees. Kunnan (2000) pointed out that the principle of justice requires that a test should be fair to all testees. This study ensured that the test items are fair to both male and female testees.In the development of test items, fairness could be achieved during items analysis or revalidation of instrument. The item analysis should be specific to the group under investigation. Instead of the proportion of testees passing an item, it should be the proportion of male and female passing an item. The same approach should apply to other specific groups such as location, race, and school-type among others. Similarly, an adapted test should be revalidated to ensure that the items are fair to the specific group under investigation as in the case of this study, where males and females were considered.Much to the knowledge of the researcher, have not been written about test item fairness but has been mentioned as recommendation for test bias or differential item functioning as pointed out by the following authors: Oche (2012), Alordiah (2014), Ebisine (2014), Obinne and Amale (2014) and Osadebe (2016), Agbure (2017), Onyehi (2014).

It should be noted that the Code of Fair Testing Practice in Education (2004) suggested Differential Item Functioning (DIF) as a useful way of examining test for their fairness. It should be noted that a fair test is made up of fair items.One of the greatest problem in testing is the use of a test whose items are not fair to the groups of testees taking the test. These groups may include Gender (male and female), Location (rural and urban), Race (black and white). It was as a result of this problem that the researcher carried out this study and provided more information on how to ensure test item fairness for the assessment of male and female testees. This is a gap that the study covered. A research question and hypothesis for item guided the study as follows:

Research Questions

What is the proportion of male and testees passing each item in the Economics Achievement Test?

Hypothesis

There is no significant difference between the proportion of male and female testees passing each item in the Economics Achievement Test.

METHOD

The study determined test item fairness between male and female students' achievement in Economic test. A sample of 200 (100 males and 100 females) was randomly selected using stratified random sampling techniques. The instrument of the study was a Standardized Economics Achievement Test (SEAT) adapted for the study. The original instrument has 50 items. But only 40 items were fair to male and female students when it was revalidated outside the area of study. The instrument has face and content validities. The original instrument has a reliability coefficient of 0.85 obtained through Kuder Richardson formula 20 as a measure of internal

consistency. The adapted instrument for the study has a reliability coefficient of 0.92 obtained through a test retest method as a measure of stability. The data collected were analysed in line with the research questions and hypotheses. The research questions determined the proportion of male and female passing each item. An item was considered fair when at least 50% of both male and female students passed it. An item was also considered fair when there was no significant difference between the proportion of male and female students passing it. Z-test was used to test the hypothesis on the proportion of male and female students passing each item at 0.05 level of significance. The proportion analysis was in line with the recommendation of Ukwuije and Orluwene (2012).

$$Z = \frac{P_1 - P_2}{SE_P}$$

P_1 and P_2 are proportion in the two samples; and SE_P is the Standard error of the difference between proportions for the independent samples.

RESULTS

The data collected were analysed in line with the research question and hypothesis for each test item.

Research Question 1: What is the proportion of male and testees passing each item in the Economics Achievement Test?

Table 1 Percentage Analysis of male and female students on test item fairness

Item	Proportion of male passing item (N=100)	Proportion of female passing item (N=100)	Percentage of male passing item %	Percentage of male passing item %	Decision
1	0.90	0.90	90	90	Fair item
2	0.90	0.85	90	85	Fair item
3	0.81	0.81	81	81	Fair item
4	0.80	0.81	80	81	Fair item
5	0.90	0.85	90	85	Fair item
6	0.85	0.85	85	85	Fair item
7	0.88	0.88	88	88	Fair item
8	0.86	0.85	86	85	Fair item
9	0.85	0.85	85	85	Fair item
10	0.87	0.86	87	86	Fair item
11	0.81	0.82	81	82	Fair item
12	0.83	0.82	83	82	Fair item
13	0.83	0.80	83	80	Fair item
14	0.80	0.84	80	84	Fair item
15	0.80	0.80	80	80	Fair item
16	0.79	0.77	79	77	Fair item
17	0.75	0.75	75	75	Fair item
18	0.75	0.75	75	75	Fair item
19	0.72	0.71	72	71	Fair item
20	0.71	0.71	71	71	Fair item
21	0.70	0.69	70	69	Fair item
22	0.68	0.68	68	68	Fair item
23	0.65	0.66	65	66	Fair item
24	0.64	0.60	64	60	Fair item
25	0.61	0.61	61	61	Fair item
26	0.55	0.60	55	60	Fair item
27	0.58	0.58	58	58	Fair item
28	0.56	0.56	56	56	Fair item

Item	Proportion of male passing item (N=100)	Proportion of female passing item (N=100)	Percentage of male passing item %	Percentage of male passing item %	Decision
29	0.55	0.54	55	54	Fair item
30	0.55	0.53	55	53	Fair item
31	0.52	0.51	52	51	Fair item
32	0.52	0.52	52	52	Fair item
33	0.51	0.51	51	51	Fair item
34	0.50	0.51	50	51	Fair item
35	0.52	0.50	52	50	Fair item
36	0.50	0.50	50	50	Fair item
37	0.51	0.50	51	50	Fair item
38	0.50	0.51	50	51	Fair item
39	0.51	0.51	51	51	Fair item
40	0.50	050	50	50	Fair item

The table 1 indicates the test item fairness for both male and female students. An item was considered fair when at least a proportion of both male and female students passed it. The test items were composed of achievement test in Economics. The students have been taught the content areas. The items in ascending order of difficulty. The simple items came first followed by the difficult items. The result showed that each item was fair to both male and female students. It should be noted that each item was well analysed before being used for assessment.

Hypothesis 1: There is no significant difference between the proportion of male and female testees passing each item in the Economics Achievement Test.

Table 2 Z-test Analysis on the Proportion of male and female students passing each item

Item	Proportion of male passing item (N=100)	Proportion of female passing item (N=100)	$Z = \frac{P_1 - P_2}{SE_P}$ Z-cal	Z-critical at 0.05 level of Significance	Decision
1	0.90	0.90	0.00	1.96	Accepted as a fair item
2	0.90	0.85	1.06	1.96	Accepted as a fair item
3	0.81	0.81	0.00	1.96	Accepted as a fair item
4	0.80	0.81	0.17	1.96	Accepted as a fair item
Item	Proportion of male passing item (N=100)	Proportion of female passing item (N=100)	$Z = \frac{P_1 - P_2}{SE_P}$ Z-cal	Z-critical at 0.05 level of Significance	Decision
5	0.90	0.85	1.06	1.96	Accepted as a fair item
6	0.85	0.85	0.00	1.96	Accepted as a fair item
7	0.88	0.88	0.00	1.96	Accepted as a fair item
8	0.86	0.85	0.25	1.96	Accepted as a fair item
9	0.85	0.85	0.00	1.96	Accepted as a fair item
10	0.87	0.86	0.20	1.96	Accepted as a fair item
11	0.81	0.82	0.20	1.96	Accepted as a fair item
12	0.83	0.82	0.20	1.96	Accepted as a fair item
13	0.83	0.80	0.20	1.96	Accepted as a fair item
14	0.80	0.84	0.70	1.96	Accepted as a fair item
15	0.80	0.80	0.00	1.96	Accepted as a fair item
16	0.79	0.77	0.33	1.96	Accepted as a fair item
17	0.75	0.75	0.00	1.96	Accepted as a fair item
18	0.75	0.75	0.17	1.96	Accepted as a fair item
19	0.72	0.71	0.17	1.96	Accepted as a fair item
20	0.71	0.71	0.00	1.96	Accepted as a fair item
21	0.70	0.69	0.17	1.96	Accepted as a fair item
22	0.68	0.68	0.00	1.96	Accepted as a fair item

Item	Proportion of male passing item (N=100)	Proportion of female passing item (N=100)	$Z= \frac{P_1-P_2}{SE_P}$ Z-cal	Z-critical at 0.05 level of Significance	Decision
23	0.65	0.66	0.14	1.96	Accepted as a fair item
24	0.64	0.60	0.57	1.96	Accepted as a fair item
25	0.61	0.61	0.00	1.96	Accepted as a fair item
26	0.55	0.60	0.63	1.96	Accepted as a fair item
27	0.58	0.58	0.00	1.96	Accepted as a fair item
28	0.56	0.56	0.00	1.96	Accepted as a fair item
29	0.55	0.54	0.14	1.96	Accepted as a fair item
30	0.55	0.53	0.20	1.96	Accepted as a fair item
31	0.52	0.51	0.14	1.96	Accepted as a fair item
32	0.52	0.52	0.00	1.96	Accepted as a fair item
33	0.51	0.51	0.00	1.96	Accepted as a fair item
34	0.50	0.51	0.14	1.96	Accepted as a fair item
35	0.52	0.50	0.29	1.96	Accepted as a fair item
36	0.50	0.50	0.00	1.96	Accepted as a fair item
37	0.51	0.50	0.14	1.96	Accepted as a fair item
38	0.50	0.51	0.14	1.96	Accepted as a fair item
39	0.51	0.51	0.00	1.96	Accepted as a fair item
40	0.50	050	0.00	1.96	Accepted as a fair item

$P < 0.05$

The table 2 shows the hypothesis for each item in terms of the proportion of male and female students passing each item of the Economics Achievement test. Each item was accepted at 0.05 level of significance. It revealed that there was no significant difference between the proportion of male and female students passing each item. This justified the test item fairness for both male and female students in Economics Achievement Test (EAT).

DISCUSSION

The result showed that each test item favoured both the male and female students in their achievement in Economics. This was an indication of the test item fairness for both the male and female testees. There was no significant difference between the proportion of male and female students passing each item. The result was similar to the recommendation of Educational Testing Service (2015) that test questions should not unfairly contribute to group difference. The result was also similar to the works of Kannan (2000) that test items should be fair to test items takers. Fishbein (1977) found that items are fair for testees if their achievement level is equated. Moreover, Code of Fair Testing Practices in Education (2004) supported that test developers are expected to develop test items that should be fair to the testees based on gender. However, the result was at variance with Osadebe (2014) on test construction and validation.

CONCLUSION

The study was carried out to determine test item fairness for male and female students in an adapted Standardized Economics Achievement Test (SEAT). The test items were revalidated to ensure fairness to the group of male and female testees in their achievement in Economics. Test items that discriminated or not fair to male and female testees were removed from the test.

The result revealed that both the male and female testees' performance in each item was fair. There was no significant difference between the proportion of male and female testees passing each item. Therefore, test items should be well validated using the specific group under investigation such as gender (male and female) used in this study.

Recommendations

The following recommendations were derived from the study

1. Test items should be developed to be fair to each group of testees as in the case of male and female testees
2. Any test item that discriminate or not fair to any group of testees should be removed from the test to ensure fairness
3. There should be no significant difference between male and female testees or any group of testees as the case may be to ensure test item fairness
4. The items should be revalidated to select fair test items for the group of testees if the test is adapted.
5. During item analysis or revalidation of instrument, any item that discriminate between any group should be removed to ensure item fairness.
6. Item analysis should be specific to the group of testees under investigation to ensure that the item does not discriminate or favour one group more than the other.
7. A test from a different culture or content areas should be revalidated to ensure test item fairness.

References

1. Agbure, B. (2017). Differential item functioning in social studies multiple choice questions in Basic Education Certificate Examination in Delta state. Unpublished M.Ed. Dissertation, Delta State University, Abraka.
2. Aiken, L.W. (1979). *Psychological testing and assessment.* Boston: Ally and Bacon.
3. Alordiah, C.O., (2014). Comparison of index of differential item functioning under the methods of item response theory and classical test theory in Mathematics. Unpublished Ph.D. Thesis.
4. Code of Fair Test Practices in Education (1988). Fair testing practice Washington, DC.
5. Code of Fair Testing Practice in Education (2004). Test fairness Washington DC Joint Committee on Test Practice.
6. Ebisine, S.S., (2014). Influence of culture on differential items functioning in National Examinations Council Senior School Certificate Mathematical multiple choice test in Nigeria. Unpublished Ph.D. Thesis Delta State University Abraka.
7. Educational Testing Service (2015). Test fairness and validity. Retrieved on 29/7/17 from https://www.ets.org/gre/institution/about/fairness.

8. Federal Government of Nigeria. (2004). National Policy on education (4th Ed.) Lagos:
9. Nigerian Educational research and Development Council Press.
10. Federal Ministry of Education, Science and Technology (1985). A handbook on Continuous assessment. Lagos: Heinemann Educational Books Nigeria Ltd.
11. Fishbein, R.L. (1977). The fairness of test items for statistically equated groups. Retrieved.
12. Kunnan, A.J. (2000). *Fairness and Validation in Language assessment.* Cambridge: University Press.
13. Obinne, A.D.E. & Amali, A.O. (2004). Differential item function the implication for educational testing in Nigeria. *International Review of social science and humanities, 7*(1), 52-65.
14. Oche, E.S. (2012). Issues in test item Bias in Public examinations in Nigeria and implication for testing. *International Journal of Academic Research in Progressive Education and Development, 1*(1), 175-187.
15. Onyehi, A.K. (2014). Differential item function on gender and location in Economics Achievement Test. A Doctoral Seminar presented at the Delta State University Abraka.
16. Osadebe, P.U. & Obiunu, J. (2005). Evaluation of gender participation in Economic development. *The Counsellor,* 21, 87-94.
17. Osadebe, P.U. (2001). Construction and standardization of Economics Achievement Test. Unpublished Ph.D. Thesis, University of Port-Harcourt, Port-Harcourt.
18. Osadebe, P.U. (2014). Construction of Economics Achievement Test for assessment of students. *World Journal of Education, 4*(2), 58-64.
19. Osadebe, P.U. (2015). Comparative analysis of senior secondary school students' performance with school-based assessment scores of continuous assessment and promotion examination in Delta State. *International Journal of Education, 7*(1), 154-164.
20. Osadebe, P.U. (2016). Test bias in testing A seminar Paper, Presented in Delta State University, Abraka, Nigeria.
21. Swerdik, C. (2009). Psychological testing and assessment: An introduction to test and measurement (7th Ed) Untied States: McGraw-Hill. Retrieved on 19th September, 2017 from http://www.primisonline.com.
22. Ukwuije, R.P.I. & Orluwene, G.W. (2012). *Peanut Educational Statistics.* Port-Harcourt: Chadik Printing Press.
23. Ukwuije, R.P.I. (1996). *Test and Measurement.* Port-Harcourt: Abe Publishers.

****