Journal of Education and Practice
ISSN 2222-1735 (Paper) ISSN 2222-288X (Online)
Vol.4, No.17, 2013

A Study of Differential Item Functioning in Ekiti State Unified Mathematics Examination for Senior Secondary Schools

ADEBULE, Samuel Olufemi PhD
e-mail: doctorolufemiadebule@yahoo.com
Faculty of Education, Ekiti State University, Ado-Ekiti, Nigeria

Abstract
This study was designed to find out if differentially functioning items were used in Ekiti State Unified Mathematics Examination (ESUME) and also to confirm if the test items function in different ways for different groups of test takers. A sample of 400 students selected using the stratified and combined sampling techniques was involved in the study. A 3-20 item multiple choice objective mathematics test items selected from Ekiti State Unified Mathematics Examination for 2008/2009 and 2009/2010 academic sessions were used as instrument of data collection. One research question was raised and one research hypothesis was generated and tested at 0.05 level of significance. The results show closeness in the means and standard deviations of the scores of the groups of testees indicating that the testees are of comparable ability levels. It can be concluded that the items of ESUME did not function differentially among the testees on the basis of gender, age, parental qualifications and location. It is recommended that differential item functioning procedure should be carried out on all items of the various subject examinations by experts, examination bodies and Ministry of Education.

Introduction
Mathematics is one of the cornerstones on which the much expected technological development in Nigeria is hinged. Hence it is a pre-requisite for admission into technology based and other science oriented courses in the higher institutions of learning. Ajayi & Ighoroje (2000), Faleye & Dibu-Ojerinde (2006) observed that registration of the students in science related courses at the higher institutions is declining rapidly, particularly for female students. Also, it was observed that male students seem to have advantage over their female counterparts in Mathematics and science examinations while better performance on verbal ability measures tilt towards female (Adebule, 2004).
According to Barret (2001), multiple choice tests are generally biased towards male while the female students experience more difficulties with questions involving numerical, spatial or high reasoning skills. Also, Lee (1999) observed that questions always arise concerning whether high average test scores by certain groups are due to actual achievement differences, bias in test or a combination of both.
Some groups are favoured while some are disadvantaged during promotion and admission or selections into science-based courses in the higher institutions. The disadvantaged group may be affected by some factors tagged extraneous and irrelevant variables that interfere with measurement of the underlying psychological construct, which Smith (1985) referred to as measurement disturbance. These could result from characteristics of the person that are independent of the items, interaction between the characteristics of the person and properties of the item and properties of the items which are independent of the characteristics of the person, i.e. the person's ability and item difficulty. Test bias occurs when a decision, grounded to some degree by the scores yielded from a test is unfair or has a perceived disparate impact on one group or another (Adebule, 2009).
A test is biased if the characteristics of the measures, the measurement process or the interpretation of the results of measurement lead to inaccurate inferences about the knowledge, skills and other attributes of an individual or a group of testees (Gregory, 2006). According to Stark, Chemyshenko, Chuah, Lee & Wadlington (2001) measurement equivalence exists in tests when the relationship between observed scores and the latent variable measured by the test are identical across the sub-population. The only observed difference is due to differences among the sub-population on the basis of how much of the construct being measured is possessed.
A child that scores zero in a test cannot be adjudged to have zero knowledge of that subject, there are probably some simple questions that he could have answered correctly but were not included in the test. Forty items answered correctly on a Mathematics test does not have the same meaning as forty items correctly answered in a more difficult one. The scores do not represent the same level of achievement as forty correct items on Physics, Chemistry or English Language. Measurement equivalence is only achieved when provision are made to overcome the variance and inequality of the unit of measurements. Such that a particular measure can hve the same meaning to members of different groups of testees. However, Stark et al (2001) stated that violations of measurement equivalence are evidenced by differential item functioning.
When performance of the item task is within the testee's capabilities, he or she will typically produce a correct response, but at times, test items may set demands other than those intended by the test developer leading to

different interpretations or meanings for members of different groups or sub-population within the group, such items are said to be biased (Gierl, 2004) and Hopkins (1998).

According to Braimoh (2011), Adebule (2009), Sheppard (1984) and Echernacht (1974), item bias is the invalidity or systematic error in how a test item measures a construct for the members of a particular group. Plake (1981) defined bias in test items as a situation when items in an achievement test are found to favour one group of the sub-population over another for reasons not explainable by differences in achievement level between groups. Le (1999) submits that in order to investigate bias at the item level, developers of large scale assessment usually conduct a differential item functioning (DIF) analysis. This is the statistical detection of systematic performance differences between individuals with the same underlying true ability from different groups.

Differential Item Functioning (DIF) also referred to as measurement bias, occurs when people from different groups (commonly gender or ethnicity) with the same latent traits (ability/skill) have a different probability of giving a certain response on a test or questionnaire. An item does not display DIF if people from different groups have a different probability to give a certain response, it displays DIF if and only if people from different groups with the same underlying true ability have a different probability of giving a certain response (Wikipedia.org/wiki).

Research Question

The following question was raised to assist the study:

1. Will the testees in the various groups be of comparable ability levels?

Research Hypothesis

A null hypothesis was generated and tested at 0.05 level of significance:

1. There is no significant difference between the performance of the different groups of testees in Ekiti State Unified Mathematics Examination.

Research Method

A descriptive research design of the survey type was adopted. The population consists of all 2011/2012 senior secondary school III students in Ekiti State. A sample of 400 students selected using the stratified random sampling technique was involved in the study. The instrument used to gather data was a-3-20-item multiple choice objective Mathematics test item selected from Ekiti State Unified Examination for 2008/2009, 2009/2010 and 2010/2011 academic sessions totaling 60 items in all for each testee. The items were already subjected to the processes of validation and standardisation by the examination unit of Ekiti State Ministry of Education. Thus the items were okay in terms of subject matter contents and instructional objectives. However, some experts in Mathematics and psychometrics at Ekiti State University, Ado-Ekiti certified that the items have face, content and construct validity.

The instrument was administered on the students by the researcher with the assistance of their Mathematics teachers and research assistants already trained for the purpose. Data collected were analysed descriptively and inferentially using mean, standard deviation and student t-test.

Results

The results of the analysis are presented below.

Research question 1: Will the testees in the various groups be of comparable ability levels?

Table 1: Summary of distribution of scores on Ekiti State Unified Mathematics Examination (ESUME) mathematics test

Scores	*Gender*		*Age*		*Parents' Qualification*		*Location*		*Total*
	M	*F*	*15yrs & Below*	*Above 15yrs*	*Below O/L*	*Above O/L*	*Rural*	*Urban*	
1—20	53	26	35	44	49	30	40	39	79
21—40	154	127	100	181	160	121	117	164	281
41—60	23	11	14	20	21	13	16	18	34
Above 60	4	2	3	3	2	4	1	5	6
Total	234	166	152	248	232	168	174	226	400
Mean score	27.95	28.92	27.8	28.43	27.84	28.69	27.41	28.81	28.2
Standard deviation	10.94	9.39	11.98	10.56	11.13	11.05	10.73	11.13	11.10

A cursory look at table 1 shows that there are closeness in the mean score of the testees in the various groups. This is an indication that the testees are of comparable ability levels.

Research hypothesis 1: There is no significant difference between the performance of the different groups of testees in Ekiti State Unified Mathematics Examination (ESUME)

Table 2: Summary of t-test statistics on the various sub-groups total performance in ESUME

	Variables	***N***	$\overline{X}$	***SD***	***df***	t_{cal}	t_{table}
Gender	Male Female	234 166	27.95 28.92	10.94 9.39	388	0.951	1.960
Age	15yrs & below Above 15yrs	152 248	27.8 28.43	11.94 10.56	388	0.5344	1.960
Parents' qualification	O/L Above O/L	232 168	27.84 28.69	11.13 11.05	388	0.757	1.960
Location	Rural Urban	174 226	27.41 28.81	10.73 11.13	388	1.10	1.960

$\rho > 0.05$ (Results not significant)

Table 2 shows the sample sizes for each of the subgroups, their mean scores and standard deviation. The table also reveals the test of significance for equality of means as well as the ρ-values. The $t_{calculated}$ values are 0.951, 0.5344, 0.757 and 1.10 for gender, age, parents' qualification and location of students respectively. The difference between the mean scores of all the variables (subgroup) is not significant. Hence there is no significant difference in the performance of all the examinees based on gender, age, parents' qualification and location.

Discussion

Ekiti State Unified Mathematics Examination (ESUME) is a novel policy of the government designed to bring about a better performance at both internal and external examinations. This lofty idea is geared towards evaluating the achievement of the students through valid and reliable test items. However, if test items are biased, the learners would be prevented from demonstrating appropriate knowledge, skills and attitude they had acquired. This study found out if Ekiti State Unified Mathematics Examination (ESUME) test items functioned differentially among different subgroups of testees. To be able to confirm this, the researcher in line with Zumbo (1999) found out that the testees in the various groups (gender, age, parents' qualification and location) are of comparable ability levels. The testees irrespective of their numbers in respective score intervals are close across the sub-populations indicating comparable ability levels.

The only research hypothesis generated was to find out if there exists any significant difference in the performance of each of the subgroup of testees. It was found out that the differences between the mean scores of all the subgroups are not significant. Thus it could be confirmed that no differentially functioning items were used in Ekiti State Unified Mathematics Examination for the period under study. However, the finding is not in line with Braimah (2011) that reported that some differentially functioning items were used in the 2005—2007 Edo State Unified Mathematics Examination.

Conclusion

It can be concluded that the items of Ekiti State Unified Mathematics Examination did not function differentially among the testees on the basis of the subgroups of gender, age, parents' educational qualification and location.

Recommendations

Since it is inappropriate to judge the adequacy of any test item without subjecting it to the process of item analysis, it is recommended that differential item functioning procedure for declaring item bias should be carried out on all items of the various subject examinations. All examination bodies, experts in Ministry of Education and people saddled with the process of developing, validation and administration of test need to carry out measurement bias at all times.

REFERENCES

Adebule, S. O. (2004). Gender difference on a locally standardised rating scale in Mathematics. *Nigerian Journal of Counselling and Applied Psychology*. 2(1): 22—29.

Adebule, S. O. (2009). Identification of biased items in school certificate Mathematics achievement test. *International Journal of Research in Education*. 1(2): 69—73.

Ajayi, O. B. & Ighoroje, A. D. (2000). Female enrolment for information technology training in Nigeria. Available online at www.wigsat.org. Retrieved on 20/5/2013.

Barret, S. (2001). Differential item functioning: A case study of first year Economics students. *International Education Journal*. 2(3).
Braimah, M. (2011). A study of differential item functioning in Mathematics examination on selected secondary school students in Edo State. An unpublished M.Ed. proposal, University of Ado-Ekiti.
Echernacht, G. (1974). A quick method of determining test bias. *Educational and Psychological Measurement*. 34(2): 271—280.
Faleye, B. A. & Dibu-Ojerinde, O. O. (2006). A review of the enrolment and performance of male and female students in Education Economics programme Obafemi Awolowo University, Ile-Ife. *Journal of Social Sciences*. 12(2): 143—146.
Gierl, M. J. (2004). Using a multidimensionality-based framework to identify and interpret construct related criterion that elicit group differences. *AERA*. USA.
Gregory, R. J. (2006). Psychological testing: History, principles and applications. 4th edition. India: Pearson Education Inc.
Hopkins, K. (1998). Educational and psychological measurement and evaluation. Boston: Allyn and Bacon.
Le, V. (1999). Identifying differential item functioning on the NELS history achievement test. Available online at www.cse.ucla.edu. Retrieved on 24/4/2013.
Plake, B. (1981). An ANOVA methodology to identify biased test item that take instructional level into account. *Educational and Psychological Measurement*. 41: 365—368.
Sheppard, L. A. (1984). Accounting for statistical artifacts in item bias research. *Journal of Educational Statistics*. 9(2): 93—128.
Smith, R. M. (1985). A comparison of Rasch person analysis and robust estimators. *Educational and Psychological Measurement*. 45: 433—444.
Stark, S.; Chemyshenko, S.; Chuan, D.; Lee, V. & Wadlington, P. (2001). Detection of differential item/test functioning using IRT. Available online at www.umu.se. Retrieved on 20/4/2013.
Wikipedia. Available online at wikipedia.org/wiki. Retrieved on 20/05/2013.