INCIDENCE OF SOCIO-ECONOMIC STATUS DIFFERENTIAL ITEM FUNCTIONING ITEMS IN MATHEMATICS TEST: A RASCH MODEL APPROACH

By

ALORDIAH, Caroline Ochuko PhD
Department of Educational Psychology
College of Education, Agbor
E-mail: carolinealordiah@gmail.com
Phone Number: 08035504867

And

OHORIEMU, Blessing Okeoghene Msc
Department of Mathematics
College of Education, Agbor
E-mail: oohoriemu@gmail.com
Phone Number: 07035365447

Paper Presented at Counselling Association of Nigeria, Delta State Chapter conference at Delta State University Abraka (March 6th-9th 2018)

Abstract

This study attempted to detect differential item functioning items in Mathematics Achievement Test (MAT) using the Rasch Model approach in relation to socio-economic status (SES). The researchers adopted the instrumental design for this study. The population of the study comprised of all senior secondary school class three students in Edo and Delta States, while a sample of 1896 students were selected for the study using cluster random sampling method. The result showed that the items in MAT do not significantly function differentially for examinees from low and high SES. It was recommended that test developers should take into cognizance the disparities that exist between low and high SES examinees when writing test items.

Key words: Differential Item Functioning, Item Response Theory, Rasch Model and socio-economic status.

Introduction

Mathematics is a subject that helps the individual to reason logically and sequentially when faced with everyday problems. Just as English language forms the bedrock of liberal art, so also does the study and mastery of mathematics form the life wire of technological development and advancement of a nation. Most students' identify mathematics as their least favourite subject. It has become a barrier to some students' success in the school and in gaining admission into university (Alordiah, 2015). Some of the reasons why mathematics subject is difficult to learn is that the concept in mathematics are abstract; mathematics relied greatly on deductive methods, axiomatic structure, and a wide range of unfamiliar symbols. When mathematics knowledge does not relate directly to concrete or real objects and is filled with signs and symbols representing abstract relations, structures and patterns can lead to students from different groups giving different interpretation to concept and mathematics test items. This can lead to differential item functioning among students.

The extent to which a test measure what it purports to measure is of paramount importance to psychometricians. A fair test is one that is considerably valid for all groups and individuals' and that affords all examinees an equal opportunity to demonstrate the skills and knowledge, which they have acquired and which are relevant to the test purpose (Roever, 2005). According to Crane, Gibbons, Narasimhalu, Lai, and Cella (2007) differential item functioning (DIF) could be defined as the different probability of giving the right answer to a test item by two individual with the same ability but from different groups. Thus, Differential item functioning implies that even after controlling for ability, an item appears to be more difficult for examinees from one group, as compared to examinees in the other group. There are two main types of DIF, namely uniform DIF and non-uniform DIF. Uniform DIF is said to occur when difference in correct response probability are found across all ability level for a particular item. On the other hand, non-uniform DIF occurs when there is an interaction between the ability and group membership such that an item may seem difficult for those

at the higher level in one group and after a particular point, it becomes more difficult for those at the lower level in the other group (Alordiah, 2013).

An item is said to flag DIF if:

1. It contains language or content that is unequally difficult for different subgroups of test takers.
2. The test item, item stem, test instruction or distracter is not good enough or/and can be understood in more than one way by the test takers.
3. There are no equal learning opportunities so much that one group is more exposed to the information being tested than the other group.
4. There are no equal access to relevant textbooks, instruments, equipment, laboratories and workshops.
5. There is no equal scoring format for the test takers.
6. A topic is of greater importance to one group than the other group (Alordiah, 2013).

Test items that are constructed in which test developers do not put into consideration variables in the society that create different subgroups, can result in examinees with the same ability but from different subgroups likely having different probabilities of getting an item in a test right, not because they do not have the ability to get the item correct but because the items contains traits that are alien to the subject matter being tested, which tend to favour one group over the other. A variable such as socio-economic status (SES) is capable of dividing examinees into such subgroup. This is in line with Odili (2003) findings when he carried out a study on the detection of DIF items in WAEC/SSCE biology paper 2 for 1990 and 2001; there was evidence of the presence of SES differential item functioning items.

Socio-economic status is the way economic and social characteristics divides individuals in the society into groups. Evans (2004) repeatedly discovered that low SES children are less cognitively stimulated than high SES children because of reading less and experiencing less

complex communications with parents involving more limited vocabulary. Hart (2014) said that students with a lower SES often face additional challenges which include unavailability of learning resources, difficult learning conditions, and poor motivation that negatively affect their academic performance. There is need for mathematics teachers and test developers to focus on the integration of SES into their teaching and learning process as well as putting it into consideration during evaluation. This implies that our test items should be free of DIF items. There are several methods of detecting DIF items in test. These methods are available for the mathematics teacher and test developers to use.

However, some of these methods are based on Classical Test Theory (CTT) while some others are based on Item Response Theory (IRT). Ugodulunwa (2014) emphasises that for there to be quality assurance in assessment in Nigeria there should be a shift of emphasis from CTT to IRT because of the promises it holds in solving most of the test design problems. This is in line with the findings of Mordi, Odili, and Alordiah (2015) which showed that the DIF detection methods based on IRT was a better method that can be used to detect DIF items in a test because it does not depend on p-value but on individual student latent ability. The Rash Model is based on IRT. The Rasch Model transforms raw scores into equal interval level data to construct a measure and calibrate item estimate (Hall, 2009). There are two important parameters in the Rasch Model, namely item difficulty and examinees ability. The item difficulty is calculated from the number of examinees who succeed in that item while the examinees abilities is the estimate of the examinee's underlying ability based on performance on a set of items. An item is considered to flag DIF if the |t-value| is more than two and the probability of Δ_b (DIF contrast) is less than 0.05.

As an offshoot of this discussion, the present study is aimed at using the Rasch Model method to detect differential item functioning items in Mathematics Achievement Test (MAT) for individuals with the same ability but from different socio-economic status.

Purpose of the Study

The main purpose of the study is to detect SES differential item functioning items in MAT. The specific purposes are to determine:

1. Whether there are DIF items in MAT items among examinees from high and low SES.
2. Which group (high/low SES) the DIF items favoured.
3. The magnitude of the DIF items.

Research Questions

1. Are there DIF items in MAT items among examinees from high and low SES?
2. Which group (high/low SES) does the DIF items favoured?
3. What is the magnitude of the DIF items?

Hypothesis

Items in the MAT do not significantly function differentially among examinees from high and low SES.

Method

This study adopted an instrumental research design, which is aimed at identifying socio-economic status DIF items in MAT. The population comprised of all senior secondary class III students in Delta and Edo states. The cluster random sampling procedure was adopted to sample out 1897 students. The instruments for this study are the Mathematics Achievement Test (MAT) and the socio-economic status questionnaire (SESQ). The SESQ is a scale used by Alordiah, Akpadaka and Oviogbodu (2015). It contains 20 items on a 3-point scale; most favourable (3), favourable (2) and least favourable (1). According to them, the content validity for SESQ was established by making sure that the instrument contains items that measure the yardstick (such as family's income, parents' educational level, parent's occupation and social status) for classification of individuals into high and low socio-economic status. Using the test-retest method of establishing reliability the SESQ yielded a reliability coefficient of 0.70. The score of the SESQ ranges from 60 to 20 if all

items are responded to. For the purpose of this study the students with more than 40 were grouped as high SES while he students with 40 and below were grouped as low SES.

According to Mordi, Odili and Alordiah (2015) a table of specifications and the mathematics syllabus of West African Examination Council were used to construct the 50 items in the MAT. The instrument was dichotomously scored (i.e. 1 for correct answer or 0 for wrong answer). The test-retest teaching technique was used to determine the reliability of the instrument, which yielded a coefficient of 0.07. In any analysis involving IRT, there are two basic assumptions that must be verified. They are the model fit and unidimensionality. Mordi et al (2015) reported that the assumption of unidimensionality was sufficiently satisfied and all the items fit the model.

The data was analysed using the WINSTEPS Rasch Model software (version 3.64). The criterion for decision rule under the Rasch Model was that an item is said to revel DIF if the t-value is more than 2.00 and the probability is less than 0.05. The independent chi-square test was used to test the hypothesis at the 0.05 level of significance because the resulting data are in nominal scale.

Results

Research Question one

Are there DIF items in MAT items among examinees from high and low socio-economic status?

Table 1: Rasch Model Analysis of DIF for high and low SES examinees in MAT

ITEM	bL	bH	DIF CONTRAST Δb	t-value	Prob.	DIF Index	Favoured Group	DIF Magnitude
1	-1.42	-1.51	0.08	0.82	0.41	1		
2	-1.29	-1.24	-0.05	-0.51	0.61	1		
3	-0.16	-0.44	-0.28	2.67	0.01	2	L	N
4	-0.66	-0.66	0.00	0.00	1.00	1		
5	-0.54	-0.79	0.25	2.50	0.01	2	L	N
6	-0.38	-0.29	-0.10	-0.91	0.36	1		
7	-0.23	-0.17	-0.07	-0.65	0.52	1		
8	-0.04	-0.12	0.08	0.76	0.45	1		
9	0.02	-0.25	0.27	2.46	0.01	2	L	N
10	-0.78	-0.83	0.05	0.48	0.63	1		
11	0.13	-0.20	0.34	3.08	0.00	2	L	N
12	-0.54	-0.47	-0.07	-0.68	0.49	1		
13	0.70	0.70	0.00	0.00	1.00	1		
14	-0.14	-0.18	0.05	0.45	0.65	1		
15	-0.82	-0.74	-0.08	-0.83	0.40	1		
16	-0.24	-0.24	0.00	0.00	1.00	1		
17	0.08	0.79	-0.70	-6.00	0.00	2	H	LA
18	0.16	0.01	0.15	1.37	0.17	1		
19	0.10	0.10	0.00	0.00	1.00	1		
20	-0.21	-0.39	0.18	1.72	0.09	1		
21	0.06	-0.16	0.22	2.01	0.04	2	L	N
22	-0.46	-0.29	-0.17	-1.61	0.12	1		
23	0.09	0.09	0.00	0.00	1.00	1		
24	0.13	0.13	-0.18	-1.63	0.10	1		
25	-0.55	-0.55	0.07	0.67	0.50	1		
26	-0.02	-0.02	0.00	0.00	1.00	1		
27	-0.43	-0.43	0.00	0.00	1.00	1		
28	-0.45	-0.45	-0.37	-3.49	0.00	2	H	N
29	0.49	0.49	0.17	1.45	0.16	1		
30	0.16	-0.04	0.20	1.84	0.07	1		
31	0.06	0.18	-0.13	-1.15	0.25	1		
32	0.44	0.17	0.27	2.34	0.02	2	L	N
33	0.19	0.09	0.10	0.87	0.38	1		
34	1.13	0.74	0.40	2.97	0.01	2	L	N
35	0.19	0.26	-0.08	-0.67	0.50	1		
36	-0.14	0.13	-0.26	-2.43	0.02	2	H	N
37	0.13	-0.03	0.16	1.47	0.14	1		
38	0.52	0.52	0.00	0.00	1.00	1		
39	-0.06	-0.12	0.05	0.50	0.62	1		
40	-0.19	-0.19	0.00	0.00	1.00	1		
41	0.46	0.82	-0.36	-2.95	0.00	2	H	N
42	-0.17	-0.02	0.20	1.77	0.08	1		
43	-0.50	-0.57	0.07	0.67	0.50	1		
44	0.79	0.63	0.15	1.22	0.22	1		
45	0.49	0.49	0.00	0.00	1.00	1		
46	1.06	2.00	-0.94	-6.15	0.00	2	H	LA
47	1.09	1.40	-0.32	-2.24	0.03	2	H	N
48	0.05	0.21	-0.16	-1.45	0.15	1		
49	1.31	1.66	-0.35	-2.28	0.02	2	H	N
50	-0.22	0.00	-0.23	-2.13	0.03	2	H	N

b_L- Measure of low SES b_H- Measure of high SES Δ_b- DIF contrast **H**- high SES **L**- low SES **LA**- large
N- negligible **1**- non DIF item **2**- DIF item

Table 1 showed the DIF statistics of the Rasch model method for each of the 50 items for SES. An item is said to revealed DIF if t-value is greater than 2 and the probability of Δb is less than 0.05. The Rasch Model flagged 15 items at the 0.05 level of significance. That is 30% of MAT items functioned differentially for examinees from high and low SES. The DIF items are 3, 5, 9, 11, 17, 21, 28, 32, 34, 36, 41, 46, 47, 49, and 50.

Research Question Two
Which group (High/Low SES) does the DIF items favoured?

When Δ_b is positive, it indicates DIF in favour of the focal group. The focal group is the group of interest; it is the group you are studying to see whether it will differ from the reference group. In this study the focal group comprises of the low SES examinees. When Δ_b is negative, it indicates DIF in favour of the reference group, this is the group used for the comparison, in this study the reference group comprises of the high SES examinees. Hence, from table 1, there are 7 DIF items in favour of low SES examinees namely: 3, 5, 9, 11, 21, 32, and 34. There are 8 DIF items in favour of high SES examinees namely: 17, 28, 36, 41, 46, 47, 49, and 50.

Research Question Three
What is the magnitude of the DIF item?

The magnitude of DIF is determine by the following condition

If $|\text{DIF contrast}| \geq 0.65$ logit, it is large

$0.42 \text{ logit} \leq |\text{DIF contrast}| < 0.65$ logit, it is moderate

$|\text{DIF contrast}| < 0.42$ logit, it is negligible

Hence, from table 1, the DIF in items 17 and 46 are large, while the DIF in items 3, 5, 9, 11, 21, 28, 32, 34, 36, 41, 47, 49, and 50 are negligible.

Hypothesis One
Items in the MAT do not significantly function differentially among examinees from high and low SES

Table 2: Chi-Square Analysis on SES Differential Item Functioning in MAT

		No of Non-DIF items	NO of DIF items	Total	Chi-Square	df	Sig.	Decision
SES	Low	25	7	32	2.79	1	0.09	Accept Ho
	High	10	8	18				
	Total	35	15	50				

The chi-square calculated value is 2.79 was found significant at $p = 0.09$ ($df = 1$) but not significant at 0.05 since $p>0.05$. The hypothesis is accepted. In other words items in the MAT do not significantly function differentially among examinees from high and low SES.

Discussion

The result showed that there are items in MAT that measured differentially among examinees from low and high SES. This result is in agreement with similar result reported by Odili (2003). His study showed evidence of the presence of SES differential item functioning items in WAEC/SSCE biology paper 2 for 1999-2001.

The result based on research question two showed that seven of the DIF items were in favour of low SES examinees while eight of the DIF items were in favour of the high SES examinees. This probably means SES can play a dominant role in the academic performance of students. It also implies that the concept from which these DIF items were drafted from were in-depth in the activities and life style of the SES of the favoured group.

Results on research question three showed that the magnitude of DIF in two items were large while the magnitude of DIF in 13 items was negligible. It was also observed that the two items with large DIF are items that favoured examinees from high SES. Roever (2005) advised that for test construction, items with negligible DIF can still be used, items with moderate DIF can be used where there are not enough good items and/or due to test specifications requirement, but items with large DIF should be done away with or be re-written.

The result of the analysis based on the hypothesis showed that the null hypothesis was accepted. This is so because the number of items that favoured high SES almost counselled out the

number of items that favoured low SES since seven items favoured low SES examinees while eight items favoured high SES examinees. Therefore, despite the presence of SES differential item functioning items in MAT, it was not enough to say that SES differential item functioning was actually present in the test because when the DIF items that favoured each group is almost equal in number; this implies that both groups were put into consideration when constructing the test. Another reason for accepting the null hypothesis could be because the magnitude of most of the DIF items was negligible.

Conclusion

The findings of this study showed that there are SES differential item functioning items in MAT. The finding also, indicated that this is not enough to conclude that SES differential item functioning items are present in MAT. The number of SES differential item functioning items that favoured the various groups as well as magnitude of DIF in the detected items must be taken into consideration. Hence, the study further showed that items in MAT do not significantly function differentially among examinees from low and high SES.

Recommendations

On the bases of the findings and conclusion of this study, it is hereby recommended that:

1. Test writes should be sensitive to the disparities that exist between low and high SES examinees.
2. Test writers should write mathematics items that would not unduly favour students from one group over the other.
3. Test writers should be trained on how to identify DIF items and on how to write DIF-free items.
4. Test writers should construct items that are free from writing errors such as offensive, controversial and demeaning terms.
5. Test writers should not quickly draw conclusion on whether a test contain DIF items or not, Rather they should first find out how many of the DIF items favoured the various groups; the magnitude of such DIF items should be established.

References

Alordiah, C. O., Akpadaka, G., & Oviogbodu, C. O. (2015). Influence of Gender, School location and Socio-economic status on students' Academic achievement in Mathematics. *Journal of Education and Practice*, 6(17), 130-136.

Alordiah, C. O. (2013). A theoretical comparison of selected Differential item functioning detection procedures: A challenge for Psychometricians. *African Journal of Studies in Education*, 9(1), 99-104.

Crane, P. K., Gibbons, L. E., Narasimhalu, K., Lai, J. S.,& Cella, D. (2007). Rapid detection of differential item functioning in assessments of health-related quality of life: The functional assessment of cancer therapy. *Quality of life Research, 16, 101-114.*

Evans, G. N. (2004). The environment of childhood poverty. *American Psychologist, 59, 77-92.*

Hart, L. (2014). *The effect of socioeconomic status on students achievement.* Everyday life menu.Everyday.globalpost.com/effect-socioeconomic-status-student-achievement-16898.html-3k

Hall, P. L. (2009). Evaluating Head Start Program quality: an objective measurement approach. Theses and Dessertations. Paper 1068.

Mordi, C. E., Odili, J. N., Alordiah, C. O. (2015). Enhancing precision in Differential item functioning detection: A sine qua non for Educational Development in 21st century. *DELSU Journal of Educational Research and Development*, 12(2), 13-21.

Odili, J. N. (2003). Effect of language manipulation on differential item functioning of Biology multiple choice test. *An unpublished PhD thesis.* University of Nigeria Nsukka.

Roever,C. (2005) *"that's not fair" fairness bias, and differential item functioning in language testing*. http:/ /www2.hawaii.edu/~roever/brownbag.pdf

Ugodulunwa, C. A. (2014). Quality Assurance in Research, Assessment and Evaluation in Nigeria. A lead paper presented at 16th National Conference of the Association of Educational Researchers and Evaluators of Nigeria (ASSEREN) at Calaber on 16th July.