ANALYSIS OF PSYCHOMETRIC PROPERTIES OF KANO STATE SENIOR SECONDARY CERTIFICATE QUALIFYING EXAMINATION MATHEMATICS IN MINJIBIR ZONAL EDUCATION DIRECTORATE, KANO STATE, NIGERIA

HAMISU DAUDA ABUBAKAR
Department of Educational Psychology and Counselling, Federal University Dutsin-ma, Katsina State.
hdabubakar@fudutsinma.edu.ng

Abstract

The study analyzed psychometric properties of Kano State Secondary Certificate Qualifying Examination Mathematics multiple-choice test in Minjibir Zonal Education Directorate, Kano State, Nigeria. Ex-post facto design is used for collection, analysis and interpretation of data, the population of the study consisted of 8,184 students, out of which 500 were selected as the sample through multi-stage sampling technique. The analysis was performed on responses of students on SSCQE Mathematics Multiple-Choice test consisting of forty (40) items. The research questions were answered by IRT-based item analysis and the hypotheses were tested by t-test for independent sample at a 0.05 level of significance. The findings of the study revealed that, 32.5% of items falls within the very high (≥ 1.7) discrimination value, 47.5% are easy items, 32.5% are difficult items while 20.0% are items that do not fall within the range of -3 to +3 respectively. It was concluded that SSCQE Mathematics items in Minjibir Zonal Directorate have averagely difficult items and discriminate poorly. Therefore, the researcher recommends that item analysis should be carried out before the final version of the examination to ensure equity (fairness on the test) so that the items can discriminate well among the examinees.

Keywords: Item analysis**,** Discrimination on value, Difficulty level

Introduction

Item Analysis is a systematic procedure designed to obtain specific information about each item of a test. It is designed primarily for use with objective test (especially Multiple-Choice). In Item analysis, the test constructor is concerned with item difficulty level, the discriminative power of the item and effectiveness of the distracters (Ukwuije, 2003). Item analysis is a process of assessing responses to each item in order to judge the quality or worth of the test. Item analysis is the process of collecting, summarizing, and using information from students' responses to assess the quality of the test items and allows one to observe the characteristics of a particular item which can be used to ensure that items are of an appropriate standard for inclusion in a test or need an improvement Denga (2003).

The item analysis is an important phase in the development of an exam program. In this phase statistical methods are used to identify any test items that are not working well. If an item is too easy, too difficult, failing to show a difference between skilled and unskilled examinees, or even scored incorrectly, an item analysis will reveal it. The two most common statistics reported in an item analysis are the item difficulty, which measure the proportion of examinees who responded to an item correctly, and the item

discrimination, which measure how well the item discriminates between examinees who are knowledgeable in the content area and those who are not. Item analysis helps to observe the item characteristics, and to improve the quality of the test (Gronlund, 1998).
Moreover, item difficulty describes where the item functions along the ability scale. An easy item functions among the low-ability examinees and a difficult item functions among the high-ability examinees (Baker, 2001). Therefore, the difficulty is a location index along z-axis, i.e. how far to the right or left the curve is displaced. The index of an item's location is the point on the z-axis at which the curve crosses 0.5 probability value on the y-axis. In other words, item difficulty is called item location, item threshold parameter or the b – parameter. This is the point of inflection on the ability scale, a point wherein examinees have 50% probability of correctly answering the item. Item difficulty is expressed in terms of trait level.

Specifically, an item difficulty is defined as the trait level required for participants to have a .50 probability of correctly answering the item. The item difficulty identifies the proficiency at which about 50% of the examinees are expected to answer the item correctly (Baker, 2001). Items with high value of b are difficult items, and low ability examinees have a probability of correctly responding to these items. Items with low values of b are easy items with most examinees, even those with low ability values, having at least a moderate probability of answering these items correctly (Harris, 1989 as cited in Ojerinde, Popoola, Ojo & Onyeneho, (2012). Also, item discrimination indicates how well an item separates respondents with abilities below (to the left of) the item location from those with abilities above (to the right of) the item location. Discrimination is shown by the steepness of the curve in its middle section- the steeper the curve, the better the discrimination of the item and the flatter the curve, the less the item is able to discriminate (Baker, 2001). However, the Senior Secondary School Certificate Qualifying Examination (SSCQE) is an examination designed and administered by Kano State Educational Resource Department (KERD) to students in SS11 in order to select the best candidates for state government sponsorship to write SSCE. Nevertheless, mass failure has been observed over years in the outcome of SSCQE Mathematics. The poor performance of students in the Examination became a serious concern to parents, stakeholders, teachers, educational planers, and other members of the society. Therefore, this study analyzed the psychometric properties of 2008 SSCQE in Minjibir Zonal Education Directorate.

Objectives of the Study

1. To determine the discrimination parameter estimate of the items in SSCQE Mathematics among senior secondary students in Minjibir Zonal Education Directorate.
2. To determine the difficulty parameter estimate of the items in SSCQE Mathematics among senior secondary student in Minjibir Zonal Education Directorate.
3. To determine the effect of gender on mean discrimination parameter of items in SSCQE Mathematics among senior secondary student of Minjibir Education Zonal Directorate.
4. To determine the effect of gender on mean difficulty parameter of items in SSCQE Mathematics among senior secondary student of Minjibir Zonal Education Directorate.

Research Questions

The study answered the following questions

1. What are the discrimination parameter estimates of items in SSCQE Multiple-choice Mathematics Examination?
2. What are the difficulty parameter estimates of items in SSCQE Mathematics Multiple-choice Examination?

Research Hypotheses

The study tested the following hypotheses

Ho_1: There is no significant difference in the difficulty mean parameter of SSCQE Mathematics between male and female candidates from Minjibir Zonal Education Directorate.

Ho_2: There is no significance difference in the discrimination mean parameter of SSCQE Mathematics between male and female candidates from Minjibir Zonal Education Directorate.

Methodology

Ex-post facto design is used in this research as it seeks to find out the factors associated with certain occurrences, outcomes, conditions or type of behavior by analysis of past event or already existing conditions. It can also be defined as an attempt to determine the causes for, or consequences of, and differences that already exists in group or individuals (Anikweze, 2012). The population of the study comprises all students in public senior secondary schools who sat for 2018 SSCQE Mathematics in Minjibir Zonal Education Directorate, Kano State. However, the study employed two forms of population such as population as a unit of observation and population as unit of analysis. According to Awotunde, Ugodolunwa and Ozoji (1997) as cited in Abdullahi, (2015), "a unit of analysis refers to the person or thing under study" while observation unit is "the element from which information is collected ". According to the data obtained from Kano Educational Resource Department (KERD, 2019), there were total of eight thousand one hundred and eighty four (8,184) students who sat for 2018 SSCQE Mathematics in Minjibir Zonal Education Directorate and this population is considered as an observation unit. While the population as unit of analysis are the scores of 2018 SSCQE Mathematics Examinations of the 40 items.The study used five hundred (500) as the sample size, this is because adequate sample size is a major factor in obtaining stable parameters for Rasch and IRT models and is especially critical for the IRT 2-3 parameter models as choosing a sample size sufficient to make the standard error small enough for the researcher's purpose is suggested by (Samejima as cited in Abba 2020). The justification of using a large sample size can be supported by Stone and Yumota (2004) where they reported that "reliability indices differ with sample size".

In this study, item analysis and conventional (descriptive and inferential) analyses were performed. The descriptive statistics performed via SPSS with an attempt to summarize the responses of each item on gender and locations of the respondents using frequency count and percentage as well as mean and standard deviation. All the three (3) research questions asked on item difficulty, discrimination on SSCQE Mathematics multiple-choice items were answered by analyzing the person-by-item matrix using IRTPRO package. This type of analysis is called item analysis, in which the decisions were guided by a (discrimination) and b (difficulty) parameters as guide lines put forward by Baker (2001) and Demars (2010). However, inferential statistics using t-tests for independent sample were performed in testing the hypotheses at 0.05 level of significance, via SPSS statistical package respectively. The t-test for independent sample was considered as the most appropriate statistical technique because all hypotheses entail on determining whether the mean scores from two groups significantly differ at a selected probability level (Bichi, 2004; Louise, Manion, & Morrison 2010; Gay, Mills & Airasian 2009). In data analysis two types of analyses were performed. The first analysis was Item Analysis which was done via IRTPRO software in determining the indices for the discrimination and discrimination parameters. In the other hand, two types of conventional statistics were performed; the first one was descriptive statistics particularly frequency counts and percentages. Thus, the said descriptive statistics were done in answering the research questions. The second type of conventional statistics performed was inferential statistics particularly t-test for independent sample for testing the hypotheses and were tested at .05 levels of significance.

Research Question One: What are the discrimination parameter estimates of items of SSCQE Multiplechoice Mathematics Examination?

Table 1: Item Parameter discrimination values of SSCQE Mathematics Examination Based on 2PL Logistic Model.

Item Number	Estimate	Item Number	Estimate
1	7.25	**21**	-0.92
2	1.37	**22**	-0.15
3	1.64	**23**	0.13
4	2.56	**24**	0.80
5	0.73	**25**	0.45
6	0.47	**26**	0.09
7	-0.06	**27**	6.16
8	-0.06	**28**	0.38
9	0.40	**29**	2.09
10	2.46	**30**	2.80
11	1.45	**31**	6.52
12	-0.08	**32**	0.46
13	3.14	**33**	0.60
14	4.70	**34**	2.05
15	0.50	**35**	-0.22
16	2.40	**36**	-0.19
17	-0.16	**37**	-0.19
18	0.38	**38**	9.73
19	2.06	**39**	0.77
20	1.62	**40**	1.24

The parameter discrimination was estimated by running the person-by-item-matrix on SSCQE Multiplechoice Mathematics Examination via IRTPRO package. The output was extracted and presented in the table 1 above. However the decision was guided by the parameter guidelines presented in table 2 below as suggested by Baker (2001) Demars (2010).

Table 2: Labels for discrimination Parameter Values

	Category	Number of Items	%	
	None ≤ 0	9.0	22.5	-.34
5.0	Very Low .01			2.0
	Low .35			-.64
	Moderate .65			8.0
20.0	High 1.35			
	Very High			-1.34
		40.0	**100.0**	4.0
	10.0			
	-1.69	4.0	10.0	
	≥ 1.7	13.0	32.5	
	Total			

The table 2 above portrays the summary of the labels for a parameter values. From the table it can be seen that 9 items (7, 8, 12, 17, 21, 22, 35, 36, 37) fall under none (≤ 0) value accounting for 22.5%. The items that fall under very low (.01-.34) value are 2 (23, 26) accounting for 5.0%, 8 items account for low (.35-.64) value representing 20.0% and these items are: 6, 9, 15, 18, 25, 28, 32, and 33. The items that falls under moderate (.65-1.34) value are four (5, 24, 39, 40) representing 10.0% of the 40 items. Similarly, 4 items fall under high (1.35-1.69) value and the items are: 2, 3, 11 and 20 representing 10.0% and 13 items are within the very high (≥ 1.7) value. Thus, these items are: 1, 4, 10, 13, 14, 16, 19, 27, 29, 30, 31, 34 and 38 respectively.

Research Question Two: What are the difficulty parameter estimates of items of SSCQE Multiplechoice Mathematics Examination?

The parameter difficulty was estimated by running the person-by-item- matrix via IRTPRO package. The output was extracted and presented in the table 3 below. However the decision is based on the values within the range of -3 to +3 as suggested by Baker (2001) and Demars (2010).

Table 3: Item Parameter difficulty values of SSCQE Mathematics Examination Based on 2PL Logistic Model.

Item Number	Estimate	Item Number	Estimate
1	0.08	**21**	1.27
2	1.36	**22**	2.50
3	2.99	**23**	-3.77
4	-1.11	**24**	-3.08
5	-1.46	**25**	-4.93
6	-1.34	**26**	-5.08
7	0.21	**27**	-0.72
8	0.27	**28**	-1.25
9	-0.36	**29**	-0.49
10	-0.64	**30**	-0.61
11	-0.26	**31**	-0.56
12	-2.12	**32**	-2.82

13	1.38	**33**	-1.80
14	2.33	**34**	-0.07
15	-0.77	**35**	1.40
16	2.43	**36**	-4.41
17	-3.94	**37**	1.18
18	-3.25	**38**	-0.63
19	-0.15	**39**	-3.84
20	1.01	**40**	-0.39

The table above presents item parameter difficulty values of the 2018 SSCQE Multiple-choice Mathematics Examination items based on two-parameter logistic (2PL) model. From the table it can be seen that nineteen (19) items that is items 4, 5, 6, 9, 10, 11, 12, 15, 19, 27, 28, 29, 30, 31, 32, 33, 34, 38 and 40 are within the b-value range of -3 to +3 and had negative difficult estimates while thirteen (13) items that is items 1, 2, 3, 7, 8, 13, 14, 16, 20, 21, 22, 35 and 37 are within the b-value range of -3 to +3 and had positive difficulty estimates. The negative estimates imply that 19 items are easy while 13 items are difficult. However, eight (8) of the items which are: 17, 18, 23, 24, 25, 26, 36 and 39 were rejected in terms of difficulty levels of the items based on two-parameter logistic (2PL) model.

Table 4: Labels for difficulty Parameter Values

Category	Number of Items	%
Difficult Items	13.0	32.5
Easy Items	19.0	47.5
Rejected Items	8.0	20.0
Total	**40**	**100.0**

As per as the labels of the difficulty parameter is concerned, the above table revealed that nineteen (19) items are easy items representing 47.5%, thirteen (13) items are difficult representing 32.5% while eight (8) items accounting for 20.0% of the items were rejected.

Hypotheses Testing

The two (2) hypotheses formulated were tested by t-test for independent sample and were tested at 0.05 significance levels.

HO_1: There is no significant difference in the difficulty mean parameter of the SSCQE Mathematics between male and female candidates from Minjibir Zonal Education Directorate.

Table 5: Result of t-test for difficulty mean Parameter Values by Gender in SSCQE Mathematics Examination

Gender	N	Mean	S.D	Std. Error Mean	t-value	DF	*P*value	R
Male	40	-1.69	2.58	.4075				
Female	40	1.14	3.08	.4869	4.45	78	.00	Sig.

An independent sample t-test was performed in examining the differences in the difficulty mean parameter between male and female that responded to the forty (40) multiple choice item in SSCQE Mathematics Examination in Minjibir Zonal Education Directorate. From the table above, male candidates were having a difficulty mean parameter value of -1.69 while female candidates were having a difficulty mean parameter value of 1.14 respectively. Computed t value was 4.445, p= .000, $p< .05$ revealed that the obtained mean of male significantly differed from that of the female candidates. Based on the obtained result, the stated null hypothesis was therefore rejected. The result revealed that a statistically significant difference exists in the mean parameter difficulty between male and female candidates.

HO_2: There is no significant difference in the discrimination mean parameter of the SSCQE Mathematics between male and female candidates from Minjibir Zonal Education Directorate.

Table 6: Result of t-test for discrimination mean Parameter Values by Gender in SSCQE Mathematics Examination

Gender	N	Mean	S.D	Std. Error Mean	t-value	Df	*P* value	R
Male	40	1.75	2.904 4.082	.45921	1.8		.06	
Female	40	.25		.64547	86	78	3	NS

An independent sample t-test was performed in examining the differences in the discrimination mean parameter between male and female candidates that responded to the forty (40) items SCQE Mathematics Examination in Minjibir Zonal Education Directorate. From the table above, male were having a mean parameter discrimination value of 1.75 while female were having discrimination mean parameter a value of .25 respectively. Computed t value was 1.886, p= .063, $p> .05$ revealed that the obtained mean of male do not significantly differed from that of the female. Based on the obtained result, the stated null hypothesis was therefore retained. The result revealed that no statistically significant difference exists in the discrimination mean parameter between male and female candidates that responded to the forty (40) items SSCQE Mathematics Examination in Minjibir Zonal Education Directorate

Conclusions

From the findings of the study, it is concluded that the SSCQE Mathematics items in Minjibir Zonal Education Directorate have averagely difficult items, discriminate poorly. Therefore, it is concluded that the allege massive failure on Kano State Certificate Qualifying Examinations over the years were not on the quality of the SSCQE items, rather on other variables which are not psychometric properties of the tests might be responsible.

Recommendations

Since the items on the SSCQE Mathematics items in Minjibir Zonal Education Directorates discriminate poorly, item analysis should be encourage to be carried out before the final version of the examination to ensure equity (fairness on the test) so that the items can discriminate well among the examinees.

References

Abba, B. (2020). Analysis of Differential Item Functioning of 2016 General Studies Programmes (GSP) Examinations of Bayero University Kano, Nigeria. Unpublished PhD. thesis, BUK.

Abdullahi, A. (2015). Analysis of Psychometric Properties of BUK Post-UME 2014. Unpublished Ph.D. thesis, Bayero University, Kano. Nigeria.

Anikweze, C.M (2012) Measurement and Evaluation for Teacher Education (3rd Ed.) Nigeria. Constellation Nigeria Publishers.

Baker, F.B (2001). Basic of Item Response Theory (2nd Ed). Eric Clearinghouse on Assessment and Evaluation.

Bichi, M.Y. (2004), Introduction to research method and statistics. Kano debis –co press & publishing company Ltd.

Cohen L, Manion l, & Marrison K. (2003) Research Methods in Education (6th Ed). Boston: Allyn and Bacon

DeMars, C. (2010). Item Response Theory, Understanding Statistic Measurements. New York: Oxford Press

Demars, C. (2010). Item Response Theory: Series in Understanding Statistics. New York: Oxford University Press.

Denga, D. I. (2003). Educational Measurement, continuous assessment and psychological testing. Calabar: Glad tidings press limited.

Gronlund, N. E., & Linn, R. L. (1998). Measurement and evaluation in teaching (6th Ed.). New York: MacMillan.

Kano Education Resources Department (KERD 2019). 2018 SSCQE Mathematics Binary Scores. Kano State

Lord, F. M. (1980) Application for Item Response Theory to Practical Testing Problems: Hillsdale, N.J Lawrence Elrbaum.

Masheshwari, V. K (2013). Item Analysis in Achievement Tests. Retrieved August 16th, 2019.

Ojerinde, D., Popoola, K., Ojo, F., & Onyeneho, P. (2012). Item Response Theory: Parameter Models, Estimation and Application. Abuja. Marvelous Mike Press ltd.

Sidhu, K. S (2005). New Approaches to Measurement and Evaluation. New Delhi: Sterling Publishers Private Limited.

SPSS. (1999). Item analysis. Spss.com. Chicago: Statistical Package for the Social Sciences.

Stone, M. and Yumoto, F. (2004), The Effect of Sample Size for Estimating Rasch/IRT Parameters with Dichotomous Items Article in Journal of applied measurement https://www.researchgate.net/publication/8890461

Ukwuije, R. P. I (2003). Test and Measurement for teachers. Portharcourt; CELWIL Publishers.