ESTABLISHING ITEM DISCRIMINATION INDICES OF MULTIPLE CHOICE TEST FOR SENIOR SECONDARY SCHOOLS' MOCK EXAMINATIONS IN CROSS RIVER STATE, NIGERIA: INSTRUMENT APPRAISAL FOR STUDENTS' ASSESSMENT AND FEEDBACK.

ANAGBOGU GERMAN EFFA Ph.D[1]

BICHENE, CAROL EBUTA[2]

DEPARTMENT OF EDUCATIONAL FOUNDATIONS, [1 & 2]
FACULTY OF EDUCATION, UNIVERSITY OF CALABAR, NIGERIA.

Abstract

The study investigated the item discrimination indices of multiple choice test instruments for senior secondary schools' mock examinations in Cross River State-Nigeria, as part of a process for quality assurance, a basis for instrument appraisal for students' assessment and feedback. This is so because the quality of an examination instrument and administration determines the quality of students output. The population consisted of 14,739 students' scripts for 2014 mathematics mock examinations. Systematic sampling technique was used in selecting 1,475 scripts for the study. The researchers did not engage in constructing any instrument as the study was aimed at validating the Mock Examinations instrument developed and administered by Cross River State Ministry of Education. Thus, the researchers visited Cross River State Ministry of Education, collected 2014 Mathematics Mock Examinations Multiple choice and marked scripts, all fully adopted for the study. The instrument consisted of 50 items, with 4 options each, one of which is the correct answer and the other three distracters. Item analysis procedure was applied to determine discrimination indices. Results of the analysis showed that; discrimination indices for most items in the instrument were inconsistent with recommendation in extant literature. From the results of the study, it was concluded that the State Ministry of Education's level of quality assurance in developing multiple choice items for senior secondary Mock Examinations was seen to be poor. The implications of this results is that the test instrument is of poor quality thus, it was recommended that test experts should always be recruited to assist in development and creation of an item bank to enable a dependable administration of Mock Examinations instruments in Cross River State.

Keywords: discrimination, test instrument, students' assessment and feedback, quality assurance.

Introduction

Quality assurance is germane to any assessment of teaching and learning outcomes. For testing situation in particular, the examinees have varying abilities, and the main task of the examiner is to establish the various levels of abilities and stratify or group them accordingly. In doing this, one of the most appropriate tools is the discrimination index (Anagbogu, 2009; Bichene, 2017). Item discrimination index indicates whether an item differentiates between test takers having varying degrees of knowledge or abilities. Items on scholastic ability test

should differentiate between students with higher grade average who should answer an item correctly and more frequently, than students with lower grades average. Also discrimination is seen as the correlation of test scores and scores on the criterion as grades would be the basis for computation. When applied to the teacher made test, such external criterion are not made available. Thus, the total score of the test is used as a criterion. The basic assumption of the discrimination indices therefore is that the test as a whole is an adequate measure of the domain.

To compute the discrimination index, the scores of an individual item (total) scores on the test may result if a student scored high on the test tends to answer the items correctly and those who score low, answer incorrectly; the item test correlation would be positive, but if there is no relation in answering them and the test scores, then the discrimination index would be zero. When a group of examinees scores are divided into two or more sub-groups on the basis of the test scores, then the possible discrimination index (d) would be;

$$d = \frac{U - L}{N}$$

Where;
- U = Number of candidates in the upper group who got the item correct
- L = Number of candidates in the lower group who got the item correct
- N = Total number of candidates in each groups.

Discrimination index or sometimes called choice of criterion, according to Anastasi & Urbina (2008) is the level to which an item distinguishes rightly among testees in the ability called for by the test. In another approach, Denga (2003) stated that the purpose of discrimination test is to distinguish as much as possible among students (Examinees) at all level of achievement. It indicates the effectiveness or power of an item in discriminating between bright and dull students. Similarly, Kelly and Linacre (2002) see discrimination as an indication of the extent of which success on an item corresponds to the success on the whole test. They point out that since all the items in a test are geared towards jointly generating an overall score any item with a negative or zero discrimination undermine the test, in other words, any item that falls short of appropriate discrimination requirement is not suitable for inclusion in a test battery. Positive item discrimination is considered productive unless it is so high that the item is merely repeating the information provided by other items.

In another situation, item discrimination is usually investigated against total scores on a test itself, this is as a result of the increase need on construct validation, as it makes total scores an appropriate criterion, for its selection. On the other hand, differences may exist in procedure and assumptions, most item discrimination index provide closely similar results. The numerical values of the indices might differ in the items that have been retained and those that are rejected based on the different discrimination indices are largely the same. He further analyzed that variations in item discrimination data from sample to sample is generally greater than that among different methods. Denga (2003), seem to be in support of this position when he argued that item discrimination indices could be affected by some of the following students and examination factors; The

previous learning experience of the test taker; the appropriateness of the stem to structure the question for the examinee; the extent of ambiguity in the item; the attractiveness of foils (attractiveness) to fools those who do not know the correct answer; the difficulty of the item and the presentation of the best foil which will appeal to the upper group. Despite the influences of these students and examination factors on discrimination, some scholars argue that a test with a high average discrimination index is always better indexed. They added that, despite the index, the former test will always produce more reliable scores than the later one.

In another study, Hotni (2006) found a definite correlation between level of difficulty and the discrimination index, as level of difficulty increases so does the discrimination index, which is not expected, but the result indicated that there is a maximum degree of difficulty beyond which the discrimination index starts to fall. This means that at this point, the test items become too difficult for both higher and low scorers to answer, so they no longer discriminate effectively. This implies that there are too extreme questions that are too easy with a small difficulty value and those that are too hard with a very high difficulty value

In a similar study carried out by Hostter and Kaky as reported by Thompson (2013) , using 300 students as a sample, found that the correlation between the degree of difficulty and discrimination showed that as the difficulty increased the average discrimination increased, but there was a critical level of difficulty beyond which the discrimination decreased. The desirable index of item difficulty for each item is asserted to have maximal potentials measurability. A test with some very few easy and few very difficult items are highly recommend, for besides providing diagnostic information, enables discrimination of both ends of the students' ability and thus prevent floor or ceiling effects. In another empirical finding by WAEC (1985), the option format in terms of better discrimination between high and low achievers, revealed that between subject officers, item writers and teachers, a greater percentage of teachers representing 60.50 per cent and item writers 66.7 per cent asserted that the 5-option format discriminates relatively better than 3 option or 4 option formats. The result appears to suggest that the 5-option format is preferred followed by the 4-option format and 3-option in terms of better discrimination between high and low achievers.

In another study on the influence of psychological variables on classical test item parameters, person by item interaction analysis by Ekpenyong (2001) with 600 students randomly selected from the population, it was found that, the classical test item parameters, test item discrimination and option distraction indices do not differ significantly across examinees with different level of their tendency to guess in examinations, but the parameters – item difficulty in Chemistry, differ significantly across examinees with different level of tendency to guess in examination.

The study focused on the appraisal of item discrimination levels in multiple choice test for senior secondary schools' mock examinations in Cross River State. This is done to enable the researcher to establish the quality of the items that are used in decision making in the state. Specifically, the study was focused on determining the extent to which the items that make up the test instrument truly

represent the expected discrimination levels

Research question
To what extent does Cross River State Mock Examinations multiple choice test items meet the acceptable range of discrimination indices?

Methodology
The study adopted instrumentation design; instrumentation research design involves the development and establishment of suitability of an instrument for the measurement of a particular construct. The population of this study consists of all Senior Secondary Two (SS2) students scripts for the state Mock Examinations in 2014 numbering fourteen thousand, seven hundred and thirty nine (14,739). The students already marked answer scripts for mathematics were used for the study. Stratified and systematic sampling techniques were adopted in selecting a sample of 1,475 for the study.

The researchers visited Cross River State Ministry of Education, extracted data from 2014 Mathematics Mock Examinations Multiple choice question papers, mathematics multiple choice marked answer scripts and marking guide, and mathematics curriculum/scheme of work, all of which were fully adopted for the study. The instrument consisted of 50 multiple choice items, with 4 options each, one of which is the correct answer and the other three were distracters. To be sure that the instrument used for this study was valid to be so used, the researchers presented the 50 items test instrument obtained from the Cross River State Ministry of Education; to three staff of the Department of Examinations and Records in the Cross River State Ministry of Education, three mathematics teachers in different public secondary schools in the state, and three students who took the State mock examinations in 2014, they all verified and affirmed that; the instrument was the one administered by the Cross River State Ministry of Education in 2014 Senior Secondary Schools Mock Examinations, and that no alteration has been made on it.

The sampled answer scripts were coded with numerals from 0001 to 1,475. Students' scores and response pattern were extracted from the answer scripts and arranged in a table called person-item matrix. The total score per item was recorded (showing how each examinee responded to the items), the person by 50 items matrix thus served as data bank, from where data were extracted for analysis to determine the discrimination indices of the items.

Results
Table 1:
Guideline for evaluating item discrimination

Index of discrimination	Item evaluation
0.40 and above	Very good items
0.30 to 0.39	Reasonably good (but possibly subject to improvement).
0.20 to 0.29	Marginal items (subject to improvement).
0.19 and below	Poor items (to be rejected or revised).

Adapted from Anagbogu, 2009

Table 2: Item discrimination indices of Cross River State 2014 Mock Examinations, mathematics multiple choice items.

Item	Bright	Dull	Total	d-values
1	418	378	796	0.08
2	386	284	670	0.20
3	299	333	632	-0.69***
4	369	244	613	0.25
5	436	329	765	0.21
6	302	371	673	-0.14***
7	414	299	713	0.23
8	377	315	692	0.12
9	421	267	688	0.31
10	285	393	678	-0.21***
11	126	159	285	-0.06***
12	335	211	546	0.25
13	307	260	567	0.09
14	408	314	722	0.19
15	425	301	726	0.25
16	311	368	679	-0.11***
17	480	349	829	0.26
18	488	373	861	0.23
19	474	332	806	0.28
20	482	377	859	0.21
21	161	105	266	0.11
22	312	218	530	0.19
23	376	379	755	-0.00***
24	255	179	431	0.16
25	343	250	593	0.18
26	410	180	590	0.46
27	386	206	592	0.36
28	392	241	633	0.30
29	412	261	673	0.30
30	480	174	654	0.62
31	359	181	540	0.36
32	317	219	536	0.19
33	472	340	812	0.26
34	311	155	466	0.31
35	278	463	741	-0.38***
36	294	166	460	0.26
37	187	193	380	-0.12***
38	339	218	557	0.24
39	144	188	332	-0.09***
40	301	177	478	0.25
41	118	73	191	0.09
42	421	190	611	0.46
43	315	116	431	0.40
44	299	184	483	0.23
45	211	225	436	-0.03***
46	318	196	514	0.24
47	444	204	648	0.48
48	470	218	688	0.51
49	407	210	617	0.40
50	466	251	717	0.43

*** = negative d-values

Table 3
Group range of d- values of 50 items for Cross River State 2014 Mock Examinations, mathematics multiple choice items.

High d-values	Moderate d-values	Low d-values	Negative d- values
0.62 (30)	0.36 (27)	0.28 (19)	-0.69 (3)
0.51 (48)	0.36 (31)	0.26 (17)	-0.14 (6)
0.48 (47)	0.31 (9)	0.26 (33)	-0.21 (10
0.46 (26)	0.31 (34)	0.26 (36)	-0.06 (11)
0.46 (42)	0.30 (28)	0.25 (4)	-0.11 (16)
0.43 (50)	0.30 (29)	0.25 (12)	-0.02(23)
0.40 (43)		0.25 (15)	-0.38 (35)
0.40 (49)		0.25 (40)	-0.12 (37)
		0.24 (38)	-0.09 (39)
		0.24 (46)	-0.03 (45)
		0.23 (7)	
		0.23 (18)	
		0.23 (44)	
		0.21 (5)	
		0.21 (20)	
		0.20 (2)	
		0.19 (14)	
		0.19 (22)	
		0.19 (32)	
		0.18 (25)	
		0.16 (24)	
		0.12 (8)	
		0.11 (21)	
		0.09 (13	
		0.09 (41)	
		0.08 (1)	

The results shows that the discrimination index ranges from –0.02 for item 23 to 0.62 for item 30, indicating a wide gap of discrimination indices. This implies that some items were good at pulling apart the bright examinees from the dull ones, while some items were not. Also, from the findings it can be seen that out of the 50 items used for the test, 8 (16%) items had high positive discrimination indices, 6 (12%) items had moderate discrimination indices, 26 (52%) items had low discrimination indices, though positive, while 10 (20%) items discriminated negatively. Joshua (2005) suggested that indices of 0.20 and above are preferred; Anagbogu (2009) suggested that discrimination indices of 0.40 and above are better. From the foregoing it can be seen that the lowest index of discrimination recommended among test experts is 0.20.

It can be observed from the findings that 20 out of the 50 items used for the test had discrimination indices below 0.20, which is the minimum recommended in extant literature. As if that was not enough, out of the 20 items, 10 items were seen to have

negative discrimination indices (more examinees from the lower ability group got the items right than those from higher ability group). To this end, such items are not fit for the test. This result is similar to that of Abimbola (2004) who in a study of item characteristics and students performance in Kogi State Junior Secondary three examinations, reported that of the 50 multiple choice items in mathematics, 13 had discrimination indices of less than 0.20, while 11 more discriminated negatively. However, the results disagree with most other studies carried out on item discrimination, which reported high positive discrimination for almost all 60 test items studied.

The findings of this study has serious implications on the testing phenomena, as the general principle is that a good test should be calibrated in such a way that the items are able to distinguish between the more knowledgeable and the less knowledgeable examinees (Nenty and Umoinyang, 2004; Joshua, 2005; Olatunji and Onofeghara, 2008). Similarly, in the opinion of other test scholars, a test with many poor items will give a false impression of the learning situation, and that items that discriminate negatively are bad, and if possible be discarded and or replaced with better items from the item bank. Thus, this findings call to question the quality of the test used by Cross River State Ministry of Education for the state mock examinations.

Conclusion and Recommendations

Cross River State Ministry of Education's level of adherence to test psychometric properties in developing multiple choice test items for Senior Secondary Schools Mock Examinations is seen to be below expectation, test items are not of appropriate discrimination values. Thus, it was recommended that Cross River State Ministry of Education should recruit or consult test experts to help develop item bank and correspondingly assist in the administration of Mock Examinations instruments in Cross River State.

References

Anagbogu, G.E. (2009)*Analysis of psychometric properties of WAEC and NECO examinations and students ability parameters in Cross River State, Nigeria* .Unpublished Ph.D thesis, University of Calabar, Nigeria.

Anastasi, A. & Urbina, S. (2008). *Psychological testing* (7th ed). New Jersey: prentice-hall

Bichene, C. E. (2017)*Assessment of adherence to psychometric properties of multiple choice items in Cross River State mock examinations for improved test quality*. Unpublished M.Ed Thesis, University of Calabar.

Denga, D. I. (2003). *Educational measurement, continuous assessment and psychological testing*. Calabar: Rapid educational publishers.

Kelley, T.L. & Linace, T.R. (2002). *Interpretation of educational measurement,* Yonkers-on- Hudson, New York and Chicago, Llinois: World Book Company.

Nenty, H.J. and Umoinyang, I. E. (2004). *Principles of test construction*. Calabar: Helimo Associates.

Olatunji, S.O., and Onofeghara, N. (2008). *Psychological testing: foreign and indigenous psychological tests*. Lagos: San Solid Publishers.

Thompson, M.E. (2013). *Analysis of psychometric properties and students' academic performance in 2010 and 2011 secondary mathematics examinations in Akwa Ibom State* Unpublished M.Edthesis, University of Calabar, Nigeria.