

Full Length Research Paper

Difficulty and discriminating indices of three-multiple choice tests using the confidence scoring procedure

M. S. Omirin

Faculty of Education, University of Ado-Ekiti, Ekiti State, Nigeria. E-mail: Sundayomirin@yahoo.com.

Accepted 8 september 2006

The study investigated the comparison of the difficulty and discrimination indices of three multiple choice tests using the confidence scoring procedure (CSP). The study was also set to determine whether or not the difficulty and discrimination indices would be improved, if the tests were scored by the confidence scoring procedure. Two null hypotheses were generated and tested at 0.05 significance level. The population consisted of all SS2 Secondary School Students in Gbonyin local government of Ekiti State in Nigeria. A sample 450 students was selected using the purposive and proportional sampling techniques. Questionnaires containing 50-items achievement test in mathematics (in three formats) were constructed and administered on the students. The three multiple-choice tests were mixed and the testees did not know that they were answering different forms of the same test. The test scripts were scored using the confidence scoring method. Data were analyzed using the student t-tests and ANOVA. The result revealed that the contribution of blind guessing to testees was not directly related to the discrimination and difficulty indices of the three multiple-choice tests used. The confidence scoring procedure improved significantly the difficulty index of multiple-choice tests but did not significantly improve the discrimination index of three-index test used. The result showed that confidence scoring procedure reward partial knowledge of testees on the multiple-choice tests. It was recommended that confidence-scoring procedure should be encouraged for scoring multiple-choice tests, it discourages guessing.

Key words: Are difficulty; discriminating; multiple; confidence; scoring; testees and alternative.

INTRODUCTION

Teaching activity is not complete until when the students taught are assessed. One major instrument for such assessment is test, which Payne, (1982) called a systematic method of gathering data for the purpose of making intra and inter comparisons between individuals, within a class or in a school or system, a test could be in easy or objective form, but in Nigerian educational system, the objective test has gained prominence, particularly the multiple-choice and true or false formats, due to the increase in the number of students enrolment and the need to often periodically assess the new national policy on education approved for use in 1997 in the handbook on continuous assessment (1985). The shortcomings of the objective test such as the partial knowledge of the students when responding to the items are not considered, thus causing a reduction in the total scores of testees.

The testees are prone to greater propensity to cheat or do blind guessing in objective tests. Cheating and blind

guessing enable testees to be credited with undeserved scores where an academically poor or test-wise student would score higher point than the knowledge he has in the subject. Hence, making it difficult to discriminate between the bright students and the poor students. The study, thus attempts to compare the difficulty and discrimination indices of 3-multiple choice test formats, using the confidence scoring procedure.

Statement of the Problem

The researcher observed that multiple choice tests have seriously disregarded testees and for being highly susceptible to blind guessing, whereby a student who prepares poorly accidentally scores higher than those who actually prepare well for such test. Hence, making it difficult to discriminate between the bright and poor students. Attempts to correct these flaws prompted experts to develop various correction formulae but the formulae proved ineffective. While some formulae under-correct, some over-correct and hence, the need to apply alternative procedures such as the confidence scoring procedure to rectify

Table 1. Comparison of item difficulty indices of format A in the three levels of confidence.

Sources of variation	SS	MS	Df	F	P
Between groups	0.577	0.289	2	21.57	P<0.05
Within groups	1.964	0.0134	147		
Total	2.541		149		

Table 2. Pair - wise Comparisons of item Difficulty indices of Format A.

Level of confidence	N	\bar{x}	S	T	P
AC	50	0.7565	0.105	4.382	P<0.05
PC	50	0.660	0.115		
AC	50	0.7565	0.105	6058	P<0.05
RG	50	0.606	0.123		
PC	50	0.660	0.115	2.268	P<0.05
RG	50	0.606	0.123		

the anomalies. Arising from the above, the following questions were raised:

1. Would confidence-scoring procedure discriminate more than the conventional number-right scoring procedure in the variants of multiple-choice tests?
2. Would the difficulty index affect the confidence scoring procedure than the conventional number-right method?

Hypotheses

Two null hypotheses were generated and tested at 0. 05 level of significance.

H01: There is no significant difference in the difficulty indices of 3-alternative, 4-alternative and 5-alternative multiple choice test items using the confidence scoring procedure.

H02: There is no significant difference in the discrimination indices of 3-alternative and 5-alternative multiple-choice test using the confidence scoring procedure.

Methodology

The research design used for this study was the survey design of the descriptive research. The study compared the difficulty and discrimination indices of three multiple choice test formats (3-alternative, 4-alternative and 5-alternative) in terms of the confidence used by the testees in responding to the items. This degree of confidence ranges from absolute confidence (AC), partial confidence (PC) to random guessing (RG). The population consisted of all SS2 Secondary School Students in Gbonyin local government area of Ekiti State in Nigeria,

All the school are mixed and had been accredited by West African Examinations Council (WAEC) of having minimum requirements for offering mathematics, in terms of personnel and facilities. A sample of four hundred and fifty (450) students was sampled by purposive and proportional sampling techniques, as the numbers of students in each of the school are not the same.

The instrument consists of three-choice test formats namely 3-alternative, 4-alternative, and 5-alternative called formats A, format B and format C respectively. The items were both adopted and adapted from past West African school certificate examination questions. The items covered the entire mathematics syllabus. Since the items had been validated by the West African Examinations Council (WAEC) and standardized, they were administered on the four hundred and fifty (450) SS2 students already selected for the study. The three formats of the MC test were mixed and administered randomly on the testees in each of the schools. The hypotheses were tested using student t-test and analysis of variance (ANOVA).

RESULTS

H01: There is no significant difference in the difficulty indices of 3-alternative, 4-alternative and 5-alternative multiple choice test items using confidence-scoring procedure. The hypothesis was tested using one-way ANOVA to compare the difficulty indices in the three levels of confidence, and separately for each objective test format.

Table 1 shows that in the 3-alternative multiple-choice format, the F-ratio is 21.57 and the table value of F-ratio is 3.07. That is, $F_c > F_t$. This shows a significant result. Hence the null hypothesis is rejected. This means that in a 3-alternative multiple choice format, the difficulty index

Table 3. Comparison of Difficulty index of format B in the three levels of confidence.

Sources of Variations	SS	MS	DF	F	P
Between groups with Groups	0.394 2.737	0.197 0.0186	2 147	10.59	P<0.05
Total	3.131		149		

Table 4. pair wise Comparison of items difficulty indices of format B.

Level of confidence	N	\bar{X}	S	T	P
AC	50	0.782	0.125	2.794	P<0.05
PC	50	0.709	0.136		
AC	50	0.782	0.125	4.635	P<0.05
RG	50	0.657	0.144		
PC	50	0.709	0.136	1.856	P>0.05
RG	50	0.657	0.144		

Table 5. Comparison of items difficulty indices of format C in the Three Levels of Confidence.

Sources of Variation	SS	MS	DF	F	P
Between groups	0.238	0.119	2	8.75	P<0.05
Within groups	1.994	0.136	147		
Total	2.232		149		

Table 6. Pair wise Comparisons of items Difficulty Indices of Format C.

Level of confidence	N	\bar{X}	S	T	P
AC	50	0.764	0.106		
PC	50	0.706	0.118	2.586	P>0.05
AC	50	0.764	0.106		
RG	50	0.666	0.112	4.244	P>0.05
PC	50	0.706	0.118		
RG	50	0.666	0.122	1.625	P>0.05

Table 7. Comparisons of the Difficulty and Discrimination indices of Formats A, B and C.

Format	Difficulty	Discrimination
A	0.606	0.324
B	0.657	0.302
C	0.666	0.378

will be affected by the confidence scoring procedure.

Table 2 shows the pair wise comparisons of the testees' responses with AC and PC; AC and RG; PC and RG. The table further reveals a significant t-calculated value of 4.382, 6.58 and 2.268 respectively at 0.05 level of significance.

Table 3 shows that in the 4-alternative multiple choice format, the calculated F-ratio of 10.59 is greater than the table value of 3.07. This shows a significant result. Hence, the null hypothesis rejected. That is the varying degree of examinees' confidence in responding to a 4-alternative multiply choice test items has a significant eff-

ect on its difficulty index.

Table 4 shows the pair wise comparisons of the testees responses with AC and PC, AC and RG, PC and RG. The result shows significant values of 2.794 and 4.635 for AC and PC, AC and RG respectively, and a non-significant value of 1.856 for PC and RG.

Table 5 reveals that in a 5-alternative format, the calculated f-ratio of 8.75 is greater than the table value of 3.07. This shows a significant result. Hence, the null hypothesis is rejected. This means that the varying degree of confidence of the examinees in responding to a 5-alternative multiple choice test items has significant effect on the difficulty index.

Table 6 shows significant difficulty indices by pair – wise comparisons for AC and PC, AC and RG as 2.586 and 4.244 respectively and a non-significant value of 1.625 for PC and RG.

Table 7 shows that format Discrimination least with a mean discrimination of 0.302 while format C has the highest mean discrimination value of multiple choice test was found to be proportional to its number of options.

DISCUSSION

Confidence scoring procedure has a significant effect on the difficulty indices of multiple choice test items decreases as the level of testees' confidence decreases. For instance, in format A, the level of difficulty of responses with AC, PC and RG are 0.7565, 0.660 and 0.0606 respectively: the same trend was observable in formats B and C. This result contradicts the finding of Afolabi, (1990) and Boyinbode, (1986) when they investigated the effect of confidence level on the psychometric properties of true-false test answer, and found that there was no significant difficulty in the difficulty level of true – false and multiple choice test. Multiple –choice test become more and more difficult when responses are to be made with increased confidence level (Odeyemi, 2003).

Discrimination indices obtained at different confidence levels reveals that the means increase as the level of confidence of the examinees increase. In a 3-alternative multiple – choice test, the means obtained are 0.34, 0.348 and 0.374 for random guessing partial confidence and absolute confidence respectively. This trend is same for the 4-alternative and 5-alternative multiple –choice tests. This agrees with the finding of Boyinbode, (1986)

who found that discrimination value increases with confidence level, though the mean values were not significantly different. Confidence scoring procedure improves the discrimination indices of multiple choice tests. The discrimination indices of 3-alternative, 4-alternative and 5-alternative multiple choice tests were found to be statistically insignificant at 0.05 level of significance as a result of the confidence scoring procedure (Omirin, 1999). The varying degree of confidence has a significant effect on the difficulty indices of the various formats used in this study.

RECOMMENDATIONS

Based on the findings of this study, the following recommendations were made.

1. The confidence scoring procedure should be encouraged and used in schools as it has been found to be effective in reducing the contribution of random guessing to testees' total score and in rewarding the partial knowledge of testees' on multiple choice tests.
2. Confidence scoring procedure considerably reduces the 'craze' for a do or die affair to pass examination at all cost, hence should be used in all schools.

REFERENCES

- Afolabi ERI (1990). The reliability and validity of the three confidence scoring methods in multiple choice test unpublished manuscript. Faculty of Education, Obafemi Awolowo University, Ile-Ife.
- Boyinbode ERI (1986). Effects of Confidence level on some properties of true –false test answers. Nigeria Journal of Educational Evaluation 1 (1)
- Odeyemi JO (2003). Comparison of the Psychometric properties of three multiple choice test, using the confidence scoring procedure. An unpublished M. E. D. thesis. University of Ado – Ekiti.
- Omirin MS (1999). Construction and validation of a science – Oriented Attitudinal Scale for Nigerian Schools. An unpublished PhD thesis. University of Ado – Ekiti.
- Payne DA (1982). Measurement in Education. In: Mitzed HE (ed) Encyclopedia of Educational Research. New York, Macmillan Publishers.