Validity and Reliability Indices of Three-Multiple Choice Tests Using the Confidence Scoring Procedure

M.S. Omirin

Faculty of Education, University of Ado-Ekiti, Ado-Ekiti, Nigeria

Abstract: The study investigated the validity and reliability indices of three multiple choice tests, using the Confidence Scoring Procedure (CSP). The study also sought to determine whether or not the validity and reliability indices would be improved, if the tests were scored using the confidence scoring method. The population consisted of all SS2 secondary school students in Gbonyin Local Government Area of Ekiti State, Nigeria. The sample consisted 450 SS2 mathematics students, selected by the purposive and proportional sampling techniques. The instrument was a 50-item achievement test in mathematics of three formats each, the instrument was administered on the students as a class test. The three multiple choice tests were mixed and the testees did not know that they were answering different forms of the same tests. The responses were scored using the confidence scoring method. Two null hypothesis were generated and tested at 0.05 level of significance, using Kuder-Richardson's formula 20, Z-ratio and correlation. The result showed that the contribution of blind guessing to testees was not directly related to the validity of the three multiple choice tests used. The confidence scoring procedure improved the reliability index of multiple choice of the tests used, except the validity of the 3-alternative choice. The result also showed that confidence scoring procedure rewards partial knowledge of testees on multiple choice tests. Based on the findings above, it was recommended that the confidence scoring method should be encouraged, since it improves validity and reliability indices of tests.

Key words: Purposive, Ekiti State, hypothesis, reliability, CSP

INTRODUCTION

Learning is defined as a relative permanent change in behaviour. This change in behaviour is sequel to teaching. The assessment and subsequent evaluation of students' progress in mastering the subject matter has been emphasized as basic in the art of teaching. Crow and Crow[1] explained that the teacher whose main business is to effect positive changes in students' behaviour, whether intentionally or not, will want to asses the extent or magnitude of the changes. Good and Brophy[2] also believed that students' progress need to be evaluated in order to provide a basis for grading, gather feedback about the effectiveness of their teaching, identify the topics and individual students that will need additional instruction and provide guidance for planning new instruction is test, which could either be essay or objective in nature.

In Nigeria educational system, the objective tests has gained prominence, particularly the multiple choice and true or false formats as a means of assessing students' performance in a specified task. The increase in the number of students' enrolment and the need to often periodically assess the students' performance could be adduced for the prominence of objective test in our educational system. Objective test is now widely being severely criticized by many test experts for being proned to greater propensity to cheats or do blind guessing. Cheating and blind guessing enable testee to be credited with underserved scores, where an academically poor or a test-wise student would score higher than the knowledge he has got in the subject. More so, that test experts[3] found out that the greater the contribution that blind guessing makes to the determination of students' scores, the lower the reliability of such scores.

Cheating in objective test could be prevented through administration, while blind guessing is only tackled through scoring strategies or procedures. The fact that the conventional number-right scoring procedure neither gives way to minimising blind guessing nor reward partial knowledge, test experts developed some formulae to correct for these deficiencies. The result of these in-adequacies found in the correction formula brings about the need for alternative scoring procedures like the confidence scoring procedure; logical choice-weighted scoring procedure and Empirical scoring procedure. This study, thus attempts to compare the validity and reliability indices of 3-multiple-choice tests, using confidence scoring procedure.

Statement of the problem: Objective tests, particularly the multiple-choice format have been observed to seriously disregard testees' partial knowledge and for being highly susceptible to blind guessing, whereby a student who prepares poorly, may score higher mark than he would get considering his response to each question or item. Attempts to correct these flaws prompted experts to develop, various correction formulae, but these formulae either under-correct or over-correct, hence, the need to apply alternative procedures like the confidence scoring procedure to rectify the anomalies arising from the above, the following questions were raised.

- Would Confidence Scoring Procedures (CSP) be more valid than the Conventional Number Right (CNR) scoring procedure in the variants of Multiple-Choice (MC) tests?
- Would the score procedure by Confidence Scoring Procedure (CSP) be more reliable than the scores of the Conventional Number-Right (CNR) scoring procedure in the variants of three multiple-choice tests?

Hypotheses: The following hypotheses were generated and tested at 0.05 level of significance.

- There is no significant difference in the validity coefficients of 3-alternaltive, 4-alternative and 5-alternative multiple-choice tests due to confidence scoring procedure.
- There is no significant difference in the reliability coefficients of 3-alternative, 4-alternative and 5-alternative multiple-choice tests due to confidence scoring procedure.

MATERIALS AND METHODS

The survey design of descriptive research was used in the study. This enable a systematic description of the facts, qualities of characteristics of interest as accurately as possible. The study compares the validity coefficients and the reliability coefficients of three multiple-choice test formats (3, 4 and 5-alternative).

The population consisted of all the SS2 secondary school students in Gbonyin local Government Area of Ekiti State, Nigeria. A sample of 450 SS2 mathematics students were purposively and proportionally selected.

The instrument consists of 50-items each, of three multiple choice test formats. That is, 3-alternative MC (format) 4-alternative MC (format B) and 5-alternative MC (format C). The items were both adopted and adapted from WAEC past questions, covering all the sections in mathematics syllabus.

Test-retest reliability method was used to ascertain the reliability of the instrument and a reliability coefficient of 0.82 was obtained, which was considered high enough. The content validity of the instrument was ensured by giving the draft to experienced mathematics teachers, who offered useful suggestions that were effected before administration.

- AC represents Absolute Confidence
- PC represents Partial Confidence
- RG represents Random Guessing

RESULTS

Hypothesis 1: States that there is no significant difference in the validity coefficients of 3, 4 and 5-alternative multiple-choice tests due to confidence scoring procedure.

Table 1 shows the coefficients of validity as 0.842, 0.885 and 0.691 for AC, PC and RG respectively. Pair-wise a comparisons of AC and PC; AC and RG; PC and RG, gave a calculated z-ratios of 1.443, 3.117 and 4.460, respectively (Table 2).

Whereas, z-table gives 1.96. This shows that the comparison of PC and RG; AC and RG gave significant results whiles that of AC ad PC is not significant-ratios for testing multiple correlation coefficients was also calculated to be 21.729 as against the table value of 5.991. This shows that the result is significant. Hence, the stated null hypothesis is rejected.

Table 3 shows the Coefficients of Correlations Obtained in 4-alternative MC test for AC, PC and RG are 0.728, 0.814 and 0.725, respectively.

The pair wise comparison of AC and PC; Ac and RG; PC and RG gave 1.78, 0.053 and 1.833, respectively (Table 4), as against the table value of 1.96. This shows

Table 1: validity and reliability indices of format A with AC, PC and RG

Basis of answer	Right	Wrong	K-R20	Val. coeff.
AC	3668	11332	0.993	0.842
PC	5100	9900	0.983	0.886
RG	5926	9027	0.972	0.691

Table 2: Pair wise comparisons of validity coefficients of format A

Level of confidence	N	R	Z	Z-ratio	P
AC	139	0.842	1.228	1.443	p>0.05
PC	139	0.885	1.403		
AC	139	0.842	1.226	3.117	p<0.05
RG	139	0.691	0.850		
PC	139	0.886	1.403	4.560	p<0.05
RG	139	0.691	0.850		

Table 3: Validity and reliability indices of format B with AC, PC and RG

Basis of answer	Right	Wrong	K-R20	Val. coeff.
AC	3306	11694	0.989	0.728
PC	4350	10650	0.978	0.814
RG	5172	9828	0.968	0.725

Table 4: Pair-wise comparisons of validity coefficients of format B

Level of confidence	N	R	Z	Z-ratio	P
AC	141	0.728	0.924	1.78	p>0.05
PC	141	0.814	1.139		
AC	141	0.728	0.924	0.053	p>0.05
RG	141	0.725	0.918		
PC	141	0.814	1.139	1.833	p>0.05
RG	141	0.725	0.918		

Table 5: Validity indices of format C with AC, PC and RG

Basics of answer	Right	Wrong	K-R20	Val. coeff.
AC	3532	11468	0.993	0.836
PC	4410	10590	0.987	0.861
RG	5014	9986	0.983	0.836

Table 6: Pair wise comparisons of validity coefficients of format C

Level of confidence	N	R	Z	Z-ratio	P
AC	142	0.836	1.208	0.742	p>0.05
PC	142	0.861	1.297		
AC	142	0.836	1.208	0.00	p>0.05
RG	142	0.836	1.208		
PC	142	0.861	1.297	0.742	p>0.05
RG	142	0.836	1.208		

Table 7: Pair wise comparison of reliability coefficients of format A

Level of confidence	N	R	Z	Z-ratio	P
AC	150	0.993	2.826	3.823	p>0.05
PC	150	0.983	2.380		
AC	150	0.993	2.826	5.993	p>0.05
RG	150	0.972	2.127		
PC	150	0.983	2.380	2.169	p>0.05
RG	150	0.972	2.127		

Table 8: Pair wise comparison of reliability coefficients of format B.

Level of confidence	N	K-R20	Z	Z-ratio	P
AC	150	0.989	2.599	3.001	p>0.05
PC	150	0.978	2.249		
AC	150	0.989	2.599	4.621	p>0.05
RG	150	968	2.050		
PC	150	0.978	2.249	1.620	p>0.05
RG	150	0.968	2.050		

Table 9: Pair wise comparison of reliability coefficients of format C

Level of confidence	N	R	Z	Z-ratio	P
AC	150	0.993	2.826	2.666	p>0.05
PC	150	0.987	2.515		
AC	150	0.993	2.826	3.824	p>0.05
RG	150	0.983	2.380		
PC	150	0.987	2.515	1.157	p>0.05
RG	150	0.983	2.380		

that the results are not significant. Hence the null hypothesis is upheld. That is, there is no significant difference in the validity indices of 4-alternative multiple choice test due to confidence scoring.

Table 5 Shows that the validity coefficients obtained are 0.836 and 0.836 for AC, AC and RG; and RG gave 0.742;0.00 and respectively (Table 6), against a table value of 1.96 at 0.05 level of confidence. This shows that the results are not significant. Hence, the null hypothesis is upheld. That is, there is no significant difference in the validity indices of 5-alternative multiple-choice test due to confidence scoring.

Hypothesis 2: States that the there significant difference in the reliability coefficients of 3-alternative and 5-alternative multiple-choice test due to confidence scoring procedure.

In 3-alternative MC, the reliability coefficients as on Table 1 are 0.993, 0.983 and 0.972 for AC, PC and RG, respectively. It could be observed that the coefficients, through very high, increase as the level of examines confidence increases.

Table 7 Shows Z-ratio between AC and PC; AC and RG of 3.823, 5.993 and 2.169, respectively whereas Z-table gives a value of 1.96, which is less that each of the calculated Z-ratios. That is, the result is significant, therefore the null hypothesis is rejected.

Table 8 Shows that in 4-alternative MC test, the calculated reliability coefficients are 0.989 and 0.968, respectively for AC, PC and RG, Pair wise comparisons of AC and PC; AC and RG are 3.001 and 4.621, respectively, which are significant at 0.05 significant level but the comparison of reliability coefficient of PC and RG shows a non-significant value of 1.620, Z-ratio for testing

multiple correlation coefficients was further used to find whether or not the difference in the three coefficients of reliability (0.989, 0.978, 0.968) are significant. Since Z-ratio = 21.311> table value of 5.991, Hence, the null hypothesis is rejected.

Table 9 Shows that in a 5-alternative MC test, Z-ratios between AC and PC; AC and RG; PC and RG are 2.666, 3.824 and 1.157, respectively. Multiple correlation analysis shows that the calculated Z-ratio = 14.622 which is greater than the table Z-ratio value of 5; 991; which means a significant result. Hence, the null hypothesis is rejected.

DISCUSSION

The validity and reliability coefficients obtained in the three formats of MC test used are very high. This mat be attributed the fact that the instrument used was an adopted and adapted WAEC/GCE standardized items, which had been validated for the level of students used for the study. The high validity coefficients show that teachers rating of students' papers in mathematics in all the schools used were about the same.

In all the variants of MC tests used, it can be observed that the test is most valid when testes responses by random guessing are excluded. This finding is in line with the works of Boyinbode[3] and Afolabi[4].

In the three formats of MC test-considered in the study, the reliability coefficients are direct proportional to the confidence testes have in their test answer; that is, the reliability coefficients decrease as the examined' confidence level decrease.

Boyinbode[3] and Abu-sarf and Diamond[5] obtained similar results with true-false and multiple-choice tests.

CONCLUSION

Confidence-Scoring Procedure has significant effect on the validity of 3-alternative MC but its effect on the validity of 4-alternative and 5-alternative MC test was not significant. The effect of scoring on the reliability of all the three formats was significant at 0.05 level.

Based on the above, the following recommendations were made:

- The idea of correct answer and score full mark or miss the answer and score zero should be given a second thought, as it produced the least valid result.
- Partial knowledge of the testes should be taken into consideration when responding to objective test items.

REFERENCES

1. Crown, L. and L.D. Crow, 1963. Educational Psychology. New York Van Nostrand Reinhold Company.
2. Good, T.J. and E.J. Brophy, 1980. Educational Psychology. A Realistic Approach: New York, Holt Reinhart and Winston.
3. Boyinbode, E.R.I., 1986. Effects of Confidence Level on some properties of True-false test answers. Nigerians J. Edu. Psychol., 1: 1.
4. Afolabi, E.R.I., 1990. The reliability and Validity of the three confidence scoring methods in multiple-choice test. Unpublished Manuscript. Faculty of Education, O.A.U. Ile Ife.
5. Abu-Sarf, F.K. and 1. Diamond, 1976. Effect of confidence level in multiple-choice test answers on reliability abd validity of scores. J. Edu. Res., 7: 2.