

Comparative Effectiveness of Logical-choice Weight and Confidence Scoring Methods on Reliability and Validity of Chemistry Multiple-choice Test Items in Nigerian Secondary Schools

Awodele Babatunde Adedeji (M.Ed)

*Department of Educational Foundation and Counselling,
Ekiti State University, Ado-Ekiti, Nigeria,*

Adediwura Adeyemi Alaba (Ph. D)

*Department of Educational foundations and counseling,
Obafemi Awolowo university, Ile – Ife, Nigeria,*

Adetunji Akeem Adedeji (Ph.D)

*Department of Special Education and Curriculum Studies,
Faculty of Education. Obafemi Awolowo University, Ile-Ife, Nigeria,*

Bamidele Emmanuel Folorunso (Ph.D)

*Department of Special Education and Curriculum Studies,
Faculty of Education. Obafemi Awolowo University, Ile-Ife, Nigeria, and*

Ajeigbe Taiwo Oluwafemi (Ph.D)

*Department of Educational Foundation and Counselling,
Faculty of Education. Obafemi Awolowo University, Ile-Ife, Nigeria.*

Doi:10.5901/jesr.2013.v2n3p387

Abstract

This study compared the effectiveness of Confidence Scoring Method (CSM) and Logical-choice Weight Method (LWM) of scoring objective tests in Chemistry. It examined which of the two scoring methods is more effective, reliable and valid. These were with a view to enhancing the reliability and validity of Chemistry objective test for standard assessment. The population for the study comprised of Senior Secondary School two (SSS II) students in Osun State. The sample consists of 280 SSSII Chemistry Students in their intact classes from four randomly selected schools and in four randomly selected local government areas of the state. The classes were randomly assigned to the two scoring methods. The instruments used for the study were Chemistry Multiple-choice Test type A and B (CMTA and CMTB). The 40-item Chemistry multiple-choice test was administered on the students in each school. Two scoring methods were used to score the test items. Data collected were analyzed using Kuder-Richardson (KR-21) formula and Fisher z-test. Results obtained revealed that significant difference existed in the reliability coefficient of CSM and LWM of scoring ($z^2=7.82$; $p<0.05$). Also, a significant difference existed between CSM and LWM in the validity coefficient ($z^2=5.2$; $p<0.05$). It was found that CSM was better in the reliability and validity of the test scores ((CSM: $z=2.903$; LWM: $z=0.847$ (for reliability) and CSM: $z=2.568$; LWM: $z=0.239$ (for validity)). It was therefore concluded that CSM could be used to authentically assess Chemistry students' performance and able to identify students with genuine learning difficulties.

1. Introduction

Teaching activity may not be completed until the students taught are authentically assessed. One major instrument for such assessment is test, which Omirin (1999) called a systematic method of gathering data for the

purpose of making intra and inter comparisons between individuals within class or in a school system. Also, Ugbamadu et. al (2001) defined test as an instrument made up of questions or tasks designed and presented to individuals or testees to respond to independently and the results of which can be used to determine quantitative academic change in individuals and for the quantitative comparison of performance of different individuals or their level of achievement. There are many types of tests based on different parameters, however, the objective test has gained prominence particularly the multiple-choice test because it is one of the most flexible, versatile and widely applicable test item for measuring different types of knowledge effectively, it also measures different types of complex learning outcomes in the areas of application, analysis and synthesis. In addition, it can be scored quickly, accurately and with much ease by teachers and even clerks and students. Aside from these facts, due to increase in the number of students' enrolment and the need to periodically assess the students as stipulated in the new national policy on education approved for use multiple choice item tests in National examinations.

However, as much as multiple choice items had its strengths, it also has its weaknesses. 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 academically poor or test-wise students would score higher points more than the knowledge he has in the subject. Hence, this makes it difficult to discriminate between the bright students and the poor students thus making valid and reliable judgment about the performance of students to be difficult. In order to preserve the advantage of objective test in general, and that of multiple-choice test in particular and hence to sustain their continued usefulness, a number of scoring procedures have been developed. Examples include point biserial, number right scores, confidence marking, liberal multiple choice, alternative multiple choice and the conventional multiple choice test methods. Number right score has tendency to view test scores as measure of students' cognitive capacity which is derived from reliable scoring procedure. Ebel (1979) stated that the simplest scoring method in objective test is to avoid one mark to each right answer. These are specifically designed to reduce or remove the corrupting influence of the identified test score contaminants. As Reid (1977) noted, the obvious disadvantage of this method is an upward bias in scores particular for students with low ability.

Attempts have been made to develop other scoring procedures which would have fewer defects. For instance Reid(1977) has proposed a Zero Variance formula which produces a higher score than the simple scoring formula for the students who can select the answers which they believe are wrong, rather than selecting those they believe are right. The fundamental principal employed here is that a student's score should be proportional to the average number of incorrect alternatives that he can be eliminated. Based on the criticism of Zero Variance assumption, an alternative scoring formula was derived, which is intended to approximate the ideal score for the general case in which students know the correct answer to some items, guess randomly on some items. Yet, this scoring method has been found not to be very usable in classroom testing.

As a result, empirical and a priori methods of scoring have emerged. Two of these are the Logical-choice Weight scoring procedure which allocated to each answer a weight or score according to its judged level of correctness and Confidence Scoring procedure which level ranges from Absolute Confidence (AC) through Partial Knowledge (PK) to Random or blind Guessing (RG). This study, thus attempted to compare two of the psychometric properties of multiple-choice test formats using Logical-choice Weight and Confidence Scoring procedure. However, studies have shown that both liberal and Logical choice weight methods are better than conventional method (Bradbard, Parker and Stone, 2004). Logical choice weight was found to reward partial knowledge more generously and punish misinformed examinee more severely than conventional one. (Ben-Simon, Budesen and Nevo, 2002)

2. Statement of the Problem

Multiple-choice test should be able to discriminate between knowledge mastering and novices disposition on a particular test. Cheating and blind guessing have been noticed as great disadvantage of multiple-choice tests whereby a student who prepares poorly, accidentally scores higher than those who actually prepared well for such tests. Hence, making it difficult to discriminate between the bright and the poor students. Attempts to correct these flaws prompted experts to develop various corrections formulae on scoring techniques like Confidence Scoring and Logical-choice Weight which were found to be effective as authentic assessment in

some subjects. This study made use of Confidence Scoring and Logical-choice Weight to determine their efficacies removing guessing and cheating; with a view to having a correct curriculum evaluation as far as the performance objectives are concerned.

3. Purpose of Study

The study was designed to compare the effectiveness of Logical-choice Weight and Confident Scoring Methods on the reliability and validity of Chemistry Multiple-choice test. This was with a view to providing an efficient testing and scoring procedure for authentic assessment of Chemistry students on the achievement of performance objectives in the curriculum. The objectives of the study were to:

1. determine the reliability of Chemistry Multiple-choice test when Logical-choice Weight and Confidence Scoring Method are used, and
2. investigate the relative validity of Chemistry Multiple-choice test when Logical-choice Weight and Confidence Scoring Methods are used in the test construction and scoring.

4. Research Hypotheses

1. There is no significant difference in the reliability of Chemistry Multiple-choice test using Logical-choice Weight and Confidence Scoring Methods of assessment.
2. There is no significant difference in the Validity of Chemistry Multiple-choice test using Logical-choice Weight and Confidence Scoring Methods of assessment.

5. Research Methodology

The population for the study consisted of all Senior Secondary School II (SSS II) students in Osun State. 280 students were randomly selected from four randomly selected Local Government areas of the state. Research instruments used for the study were Chemistry Multiple-choice, Test type A and B tagged CMTA and CMTB. These contained 4 options. Items were all adopted from past West African Certificate Examination. The items covered Chemistry Curriculum for SSS I and SSS II. The questions used were standard and validated WAEC questions.

5.1 Data Collection

The 40-item Chemistry Multiple-Choice tests were administered on the students with the assistance of the subject teacher in each school. The items were presented to the students as mid-term test; answers were indicated on the question sheets. The time limit was liberal and the total answer scripts collected were 280. The responses were scored using Logical-choice Weight and Confidence Scoring Methods.

5.2 Data Analyses

The internal consistency reliability of the test was found using a modified region of K-R₂₁ (Cronbach, 1984).

$$K - r_{21} = \frac{n}{n-1} \left[1 - \frac{\bar{x}(n-\bar{x})}{ns^2} \right]$$

Where

- | | | |
|-----------|---|---------------------------------------|
| n | - | Number of items in the test |
| \bar{x} | - | Means of the test scores. |
| s | - | Standard deviation of the test score. |

The reliability values obtained from the scoring procedures, that is Logical-choice Weight and Confidence Scoring now compare using Fisher z test.

$$z^2 = \frac{(z_1 - z_2)^2}{N_1 - \frac{1}{3} + N_2 - \frac{1}{3}}$$

Where z_1 and z_2 are Fisher z-transformation of r-values.

The two validity co-efficients obtained by using Logical-choice Weight and Confidence Scoring Methods were compared using Fisher z-test.

6. Results

The hypothesis 1 stated there is no significant difference in the reliability of Logical-choice Weight and Confidence Scoring Methods on Chemistry Multiple-choice tests. To test the hypothesis, the r-value of the two scoring methods were compared using higher z-test. The result is presented in Table 1.

Table 1: Comparison of the reliability of Confidence Scoring and Logical Choice Weight Methods.

Scoring Methods	n	r	z	z^2	p
Confidence Scoring	40	0.994	2.903	7.82	<0.05
Logical-choice Weight	40	0.698	0.847		

From the table 1, the numbers of test items are 40, and r-values of Confidence Scoring and Logical-choice Weight Methods are 0.994 and 0.698 respectively. Using the z-table, these r-values when transformed become $z = 3.903$, and $z = 0.847$ respectively. Comparing these z values using Fisher z-test yielded a z-value of 7.82 which is significant at the 0.05 level of significance ($p < 0.05$). The null hypothesis is therefore rejected. This means that there is a significant difference between the internal consistency reliability of Confidence Scoring and Logical-choice Weight Methods. However, Chemistry Multiple-choice test yielded a greater reliability index when scored with Confidence Scoring Method than with Logical-choice Weight Method.

Table 2: Comparison of the validity of Confidence Scoring and Logical-choice Methods.

Scoring Methods	n	r	z	z^2	p
Confidence Scoring	280	0.527	2.568	5.21	<0.05
Logical-choice Weight	280	0.231	0.239		

From Table 2, the numbers of items are 40 and r-value for Confidence Scoring and Logical-choice Weight Methods are 0.527 and 0.231 respectively, when the r-values were transformed and z scores compared, using Fisher z-test, a z^2 statistic of 5.21 was obtained with $p < 0.05$. Therefore, the null hypothesis 2 postulated in the study could not be sustained. There is significant difference in Confidence Scoring and Logical-choice Weight Methods on the concurrent validity of four-option Multiple-choice tests. Confidence Scoring Method has a greater concurrent validity than Logical-choice Weight Method.

7. Discussion

The result showed that Confidence Scoring Method of Multiple-choice test has a significant effect on reliability and validity. These findings agreed with Afolabi (1990) and Boyinbode (1986) when they investigated the effect of confidence level on the psychometric properties of the true-false answer on multiple-choice tests.

8. Conclusion and Recommendations

The Confidence Scoring Method is more adequate to capture student's cognitive status in multiple-choice tests. Also it is less complex of improving the ability of a test to reflect the degree of knowledge students have on the

items. Finally, Confidence Scoring Method is relatively less laborious and safe time unlike Logical-Choice Weight.

From the findings, the following recommendations were made:

1. The Confidence Scoring Method should be encouraged and used in schools as it has been found to be effective in reducing contribution of random guessing of the testees on Multiple-choice tests.
2. Confidence Scoring Method considerably reduces the 'craze' for a do or die affair to pass examination at all cost, hence should be used in all schools.
3. Public Examination Council such as West African Examination Council, the National Examination Council (NECO), Joint Admission and Matriculation Board (JAMB) should apply Confidence Scoring Methods in setting and scoring Chemistry Multiple-choice tests.
4. Classroom teachers should be encouraged to develop skills to how to conduct and score Chemistry Multiple-choice tests using Confidence Scoring Methods. This can be achieved through organizing seminars for them on Testing Techniques.

References

- Afolabi, E.R.I. (1990). *Towards an accurate assessment of student cognitive capacity. An evaluation of three scoring procedures in objective testing*. Unpublished manuscript, faculty of education, Obafemi Awolowo University, Ile-Ife.
- Ben-Simon, A, Budesen, D. V., and Nevo, B. (2002). A Comparative Study of Measures of Partial Knowledge in Multiple Choice test. *Applied psychology Measurement Vol 21. P 105*.
- Boyinbode, L.R. (1986). Effects of confidence level on some psychometric proper of true-fault test answer. *Nigerian journal of education psychology*.
- Bradbard, D. A., Parker D. F. and Stone, G. L (2004). An Alternative Multiple-Choice Scoring Procedure in a Macroeconomics course. *Decision Science Journal of Innovative Education Vol. 2, pp 11 -13*.
- Cronbach, L.J. (1984). *Essentials of psychological testing 4th ed*. New York. Herper and row.
- Ebel, R.L. (1979). *Essentials of Education Measurement*. New Jersey, practice hall.
- Omirin, M.S. (1999). Comparison of the psychometric properties of three multiple-choice tests, using confidence scoring procedure. *An unpublished thesis, faculty of education, Ekiti State University, Ado-Ekiti*.
- Reid, R.A. and Stem, J. A. (1977). An Alternative scoring formula for Multiple-choice and true-false tests. *Journal of Educational Research*.
- Ugbamadu, K. A., Onwuegbu, O.C. & Osunde, A.U. (2001). *Measurement and Evaluation in Education*. Benin-City, World of books Publishers.

