Journal of Education and Practice
ISSN 2222-1735 (Paper) ISSN 2222-288X (Online)
Vol.4, No.6, 2013

Difficulty and Discrimination Indices of Logical-Choice Weight and Confidence Scoring Methods on Chemistry Multiple Choice Test in Nigeria Secondary Schools.

Babatunde Adedeji Awodele [1], Dr.Yinusa Akintoye Faremi [2], Dr. Akeem Adedeji Adetunji[3] and Dr. Emmanuel FolorunsoBamidele[3]

1. Department of Guidance and Counselling, Faculty of Education, Ekiti State University, Ado-Ekiti, Nigeria.
2. Department of Educational Foundations and Management, Osun State College of Education, Ila-Orangun, Nigeria.
3. Department of Special Education and Curriculum Studies, Faculty of Education, ObafemiAwolowo University, Ile-Ife, Nigeria.
*Email: fayisod@yahoo.com

Abstract
This study determined which of the Confidence Scoring Method (CSM) and Logical-choice Weight Method (LWM) of scoring Chemistry objective tests is more reliable and valid. This was with a view in enhancing the Discrimination and Difficulty indices of Chemistry objective test for authentic assessment. The population comprised all Senior Secondary School Students in Osun State. A sample of 280 SSSII Chemistry Students was selected from 14 schools that were selected from the 30 local governments including Modakeke-Ife Area Council in Osun State using multistage sampling technique. The instruments used were Chemistry Multiple-choice Test types A and B (CMTA and CMTB). The 40-items Chemistry multiple-choice test was administered on the students in each school, on two scoring methods group. The students in CSM and LWM were told how they would be scored following their responses to the test items and were accordingly marked and scored. Data collected were analyzed using t-test statistics. Result of the analyses revealed that a significant difference existed between CSM and LWM in the discrimination indices ($t = 6.18$, $p<0.05$) with CSM ($S = 0.598$) better than LWM ($S = 0.496$). However, there was no significant difference between CSM and LWM in difficulty indices ($t =0.06$, $p>0.05$) with CSM ($S = 0.768$) and LWM ($S =0.870$). It was concluded that CSM could be used to authentically assess Chemistry student's performance with a view to identify students with genuine learning difficulties, thereby improving curriculum evaluation and instruction.
Key Words: Discrimination Indices, Difficulty indices, Objective test, Learning Difficulties, Confidence Scoring and Logical-choice Weight.

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. The objective test has gained prominence particularly the multiple-choice test 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 since 1997 in the hand book continuous assessment.

The testees are prone to greater propensity to cheat or do blind guessing in objective tests. Cheating and blind guessing enable testes to be credited with underserved scores where academically poor or test-wise subject. Hence, making it difficult to discriminate between the bright and poor students. In order to preserve the advantage of objective tests 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 including point biseral, number right score, among others. Number right scores is 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 list is to award one mark for each right answer. There are specifically designed to reduce or remove the corrupting influence of the identified test score contaminants. As Reid noted, the obvious disadvantage of this method is an upward bias in scores particular for students with low ability.

Attempts have thus 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 normal classroom situation.

As a result, empirical and a priori methods of scoring have emerged. Two of these are the Logical-choice Weight scoring procedure and Confidence Scoring procedure. In confidence scoring, candidates have to assign a confidence level to their best answer on each question. It starts from Absolute Confidence (AC), to Fairly Confidence (FC) and Not Confidence (NC). For instance: A mole of neon is (A) Diatomic (B) Ionic (C) Positively Charged (D) Triatomic (E) Monatomic (a student must be able to select one out of these options: AC, FC & NC with the correct option from A-E listed under each question). For each question candidates will be awarded +5 for picking Absolute Confidence (AC) and the right answer, +3 for Fairly Confidence (FC) and the right answer, and +1 Not Confidence (NC) and the right answer. Candidates are awarded −2 for picking Absolute Confidence (AC) and one of the incorrect options, −1 for Fairly Confidence (FC) and one of the incorrect options, and zero for Not Confidence (NC) and one of the incorrect options (Davies, 2002).

In logical choice weight, the candidates' scores depend on the degree of nearest to the correct option. According to Kolawole (2006) in a five option items a candidate with correct option obtains 1 mark, second best option 0.75, third best option 0.50, fourth best option 0.25 and fifth best option zero. The grading of the options should be done by experts in the subject concern. This study, thus attempted to compare two of the psychometric properties of multiple-choice test formats using Logical-choice Weight and Confidence Scoring procedures.

Types and Purpose

The major function of classroom testing is to measure students' performance and thus contribute to the evaluation of educational progress and attainments. Ebel (1979) opined that test act as extrinsic motivator of leaning efforts. They do give direction and make students think carefully on the goals of instructions and hence on instructional process. Ojerinde (1988) suggested that achievement test serve three general functions:

- to survey students overall performance,
- to assess relative strength and weakness (diagnosis), and
- to determine whether students are ready to learn new material or one of a higher level of complexity.

To these, Gronlund (1981) added the monitoring of pupil's learning progress and the provision of ongoing feedback to pupils and teacher. The most commonly used types of test are the essay (discussion) type, the objective (or short-answer) type, the Mathematical problem type and the oral examination type.

Tests are valuable for the following reasons:

- Act as baseline measurement for behaviours modification therapy, then a test is necessary in Language Development.
- For placement of students into different classes or placement of individual to certain positions in their organization.
- For prediction of future performance i.e. Aptitude tests of client.
- To measure the level of what students have known i.e. knowledge will be recalled
- For selection of candidate for jobs, career, etc. selection into various schools. This selection is based on the scores on a particular examination.
- For classification: In Schools, pupils are classified according to age, Intelligence Quotient (IQ), performance, identification of pupils with superior talents and those with low ability.
- Tests are very important in counseling, especially students with behavioural problems; use test before therapy.
- Adaptation: Test data are used in this sense to assist teachers and other personnel within the school system to know students and his their particular characteristic in order to be of assistance to them. Data form will help to identify the special problem the students have in order to mount programmes that will be beneficial to them.
- Development and revision of instructional practices to meet the needs and characteristics of students or employees in general. The data generated from testing the students could be used for the development of a curriculum or to revise the existing one. In this case, tests are used for research form which results into new curricula.

Discrimination

If the test and a single item measure the same thing, one would expect people who do well on that test to answer that item correctly, and those who do poorly to answer the item incorrectly. A good item discriminates between those who do well on, the test and those who do poorly. Pyrczak (1973) defined discrimination as the proportion of the group who do not answer the item correctly. The chief purpose of most tests of educational achievements is to rank examinees as accurately as possible in order of their attainments. The discriminating ability of an item also depends upon how well it is written, but even the best of writing will not convert some ideas into discriminating test items.

To study item discrimination, two extreme groups of examinees on a test are compared. The percentages suggested for use by test experts for items discrimination ranges from 25 through 27 to 33. Henrysson (1971) suggested taking the upper and lower N/3 examinees. It is obvious that the smaller the percent used for the lower and upper groups, the greater will be the differentiation; and however, the smaller the extreme groups, the less reliable are the resulting discrimination values. Kelly (1989) showed that the optimum point at which those two conditions balance is reached when upper and lower 27 percent values are used. Nunnaly (1972) suggested using 25%.

In this study therefore, 27 percent upper and lower values were used. This is given by D=U – L/N

Where U=Number of testees in the 27% that got the item correct

L = Number of testees in the lower 27% that got the item correct

N= Number of testees in the upper or lower 27% of the group

Difficulty

Item difficulty is simply the percentage of students taking the test who answered the item correctly. The larger the percentage of getting an item right, the easier the item. The higher the difficulty index, the easier the items is (Wood 1960). Marson (1969); Sax Reade, (1964) opined that positive things that we know about difficult tests is that they tend to make pupils study harder.

The concept of difficulty or decision for how difficult the test should be depends upon a variety of factors. (1) The purpose of test, (2) the ability level of the students, and (3) the age or grade level of the students. These are the same factors that must be considered in planning the number of items as well as the item format to be used. To compute the item difficult, the total number of people answering the item divides the number of people that answers an item correctly. An item answered correctly by 85% of the examinees would have an item difficulty, or p value of 85

Rather than defining difficulty in terms of some intrinsic characteristics of the difficulty, it is defined in terms of relative frequency with which those taking the test choose the correct response (Thorndike, 1971). For tests intended for selection, they recommended that the bulk of items must have a difficulty level appropriate for those at the ability level where selection will be made. However, in selection of tests, a few items should be included to provide encouragement to the poorer students to attempt the test.

In a free response item where there is a large range of possible response by the examinees, it is commonly assumed that guessing by students plays a little or no role. In a multiple-choice (MC) test on the other hand, there is likelihood that certain percentage of students correctly answers the item by random guessing. Because of these guessing tendencies in multiple-choice test, Lord (1952) was of the opinion that average item difficulty before correlation should be half way between the chance probability of success and 1.00, whereas it should be 0.50 for a free response test, in 4-option MC, the chance of probability of success is 0.20, the difficulty level midway would then be 0.60. According to Henrysson (1971) a more satisfactory index of item difficulty for more purposes is z, which transforms the item properties into standard scores of difficulty.

Ebel (1979) defined item difficulty as the proportion of students that do not answer the item correctly. Symbolically, the index of difficulty is written as P = NW/NT, where NW is the number of testes that do not answer the item correctly, and NT the total number of testees that took the test. Theoretical analysis and experimental studies demonstrates quite convincingly that in most situations questions that are neither very difficult nor very easy are best. Richardson (1936) found that a test composed of items of 50 percent difficulty has a general validity which is higher than tests composed of any other degree of difficulty. Gulliksen (1945) concluded on the basis of theoretical analysis that in order to maximize the reliability and variance of a test, the items should have high intercorrelations, all items should be of the same difficulty level, and the level should be as near 50 percent as possible.

Statement of the Problem

Multiple-choice tests 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 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 good in removing guessing and cheating; with a view to have a correct curriculum evaluation as far as the performance objectives are concerned in Chemistry.

Purpose of Study

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

Objectives of the Study

The objectives of the study were to:

1. determine the discrimination index of Chemistry Multiple-choice test when Logical-choice Weight and Confidence Scoring Methods are used and
2. investigate the difficulty index of Chemistry Multiple-choice test when Logical-choice Weight and Confidence Scoring Methods are used in Chemistry test construction.

Research Hypotheses

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

Research Methodology

The population for the study consisted of all Senior Secondary School students in the 30 local governments including Modakeke-Ife Area Council in Osun State. The sample for the study consisted of 280 SSSII students selected using multistage sampling technique. In stage one; the 30 local governments were divided into three districts using Stratified sampling technique. In stage two; out of the 10 Local Government per district, two local governments were selected using simple random technique, while Modakeke-Ife Area Council in the State was selected using purposive sampling technique to make the total Local Government selected for the study to be seven. In stage three; two schools were selected from each Local Government including Modakeke-Ife Area Council using simple random technique to make a total of 14 Senior Secondary Schools selected for the study. In stage four; 20 SSSII students were selected from each school using simple random technique to make a total of 280 students.

The research instruments were Chemistry Multiple-choice Test types 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 were standard WAEC questions that were already validated.

Data Collection

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

Data Analysis

The difficulty level of each items of the multiple-choice test should be found for each of the two scoring procedures. The t-test is used to determine whether a significant difference exists between the mean difficulty values. The discrimination index of each item of the test was found using scores obtained from the two scoring procedures. Thereafter, t-test was applied to determine whether the two mean discrimination indices were significantly different. Each of the statistical tests was carried out at the 0.05 level of significance.

Result

Hypothesis 1

The hypothesis stated that there is no significant difference between the difficulty indices of 4-option multiple-choice test scores using Confidence Scoring and Logical-choice Weight Methods. To test the hypothesis, the

mean difficulty values of the two scoring methods were compared using t-test analysis. The result is presented in Table 1below.

Table :1t-test Analysis of LWM and CSM on test Difficulty Indices

	N	$\bar{X}$	S	T	df	P
Confidence Scoring	**40**	**1.290**	**0.768**			
Logical-choice weight	**40**	**1.300**	**0.870**	**0.06**	**39**	**>0.05**

N= number of item (40)
Number of student = 280
From the table, the number of test item is 40, and the mean difficulty indices of CSM and LWM are 1.29 and 1.30 respectively. A t-test comparison of the mean value yielded a t-value of 0.06 which is not significant (P>0.05). Thus the null hypothesis is accepted.

Hypothesis 2
The hypothesis stated that there is no significant difference between the discrimination indices of Logical-choice Weight and Confidence Scoring Methods. To test the hypothesis, the mean discrimination indices of the two scoring procedures were computed and then compared, using t-test analysis. The result is presented in Table 2.

Table 2: Effect of Logical choice weight and confidence scoring on Discrimination Index.

Scoring Method	N	$\bar{x}$	S	T	df	P
Confidence Scoring	**40**	**1.740**	**0.598**			
Logical-choice weight	**40**	**0.834**	**0.496**	**6.18**	**39**	**<0.05**

N= number of item (40)
Number of student = 280
From the Table, the mean discrimination values of Confidence Scoring and Logical- choice Weight Methods are 1.74 and 0.83 respectively. A comparison of the two mean scores yielded a t-statistic of t= 6.18 which is significant at the 0.05 level. Thus, the null hypothesis is rejected with Confidence Scoring Method having a significant effect on the discrimination index of four-option multiple-choice test.

DISCUSSION
The results of this study suggest that Confidence Scoring Method of multiple- choice test has a significant effect on discrimination index, but not on the difficulty index on the same test. 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-test. Also, Odeyemi (2003) found that multiple-choice tests become more and more difficult when responses are to be made with increased in confidence level.
Discrimination indices at different confidence levels revealed that the means increases as the level of confidence of examinees increase (Boyinbode, 1986). The study revealed that Logical-choice Weight scoring Method has a significance effect on the difficulty index of multiple-choice tests.Ebel (1979) disclosed that tests that have items mostly of middle difficulty are of good discrimination, that is, not too difficult or not too easy can possibly reveal good discrimination when the upper and lower difference index is used. This agrees with the fact that most of the items in the present study have difficulty level behind the middle value. Paradoxically, a typical teacher-made test is expected to exhibits this characteristic that is to present items having difficulty below the middle-value than above, in order to avail every student adequate opportunity to demonstrate his/her cognitive capacity.
The utilization value of confidence scoring and logical choice weight scoring method, apart from the basic psychometric attribute considered above, also depends on a large extent. In the ability of classroom teachers to develop test items that are suitable for both logical-choice weight and confidence scoring procedure.

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

Recommendations

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 their multiple-choice tests.
4. Classroom teachers should be encouraged to develop skills on 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, ObafemiAwolowo University, Ile-Ife.

Boyinbode, L.R. (1986). Effects of confidence level on some psychometric proper of true-fault test answer. *Nigeria journal of education psychology.*

Davies, P. (2002). "there is no confidence in multiple-choice testing" proceeding of CAA confidence Loughborough. Loughborough University. 119-130

Ebel, R.L. (1979). *Essentials of Education Measurement.*New Jersey, practice hall.

Gronlund N.E. (1985). A comparison several item discrimination indices.*Journal of Educational Measurement 2, 69, 76.*

Gulliksep H. (1945). The Relation of item Difficulty and Inter-item correlation to test Variance and Reliability Psychometric 10.

Henrysson S. (1971). The effect of differential option weighting on multiple-choice objective tests.*Journal of Education Measurement 8, 62, 68.*

Kely T.L. (1930). The selection of upper and lower groups for the validy of tests items.*Journal of Educational Psychology 632 – 640.*

Kolawole, E. B. (2006). Statistical methods. Lagos: Academic Publishing Consultants.

Mansun R.N. (1969). Test difficult and students achievement .*American Educational Research Journal 6.*

Odeyemi (2003).*Comparison of the psychometric properties of true multiple-choice test, using confidence scoring procedure.*Unpublished M.Ed Thesis, Faculty of education, Ekiti State university, Ado- Ekiti, Nigeria.

Ojerinde D. (1988).*Educational Tests and Measurement.*Ibadan Codat Audio-visual services.

Omirin, M.S. (1999). Construction and Validation of a Science-oriented Attitudinal Schools.

Purezone F. (1973). Validity of the discrimination index as measure of item quality.*Journal of Educational Measurement IV, 227 – 231.*

Reid, R.A. and Stem, J.A. (1977).An Alternative scoring formula for Multiple-choice and true-false tests.*Journal of Educational Research.*

Richardson M.W. (1936).The relation between the difficulty and differential validity of a test psychometrical 1, 33 – 49.

Sax G and Reade, M. 1964. Achievement Washington D.C. America Council of Educational Research 1, 22 - 25.

Thorndike R.L. (1971). *Eduational Measurement.*Washington D.C. America Council of Education.

Wood D.A. (1960). *Test Construction Development and Interpretation of achievement tests.* Charleses Merrill Booms, Inc.

Word F.M (1952). The relation of the reliability of multiple choice tests to the distribution of item difficulties psychometrical 17.

This academic article was published by The International Institute for Science, Technology and Education (IISTE). The IISTE is a pioneer in the Open Access Publishing service based in the U.S. and Europe. The aim of the institute is Accelerating Global Knowledge Sharing.

More information about the publisher can be found in the IISTE's homepage: http://www.iiste.org

CALL FOR PAPERS

The IISTE is currently hosting more than 30 peer-reviewed academic journals and collaborating with academic institutions around the world. There's no deadline for submission. **Prospective authors of IISTE journals can find the submission instruction on the following page:** http://www.iiste.org/Journals/

The IISTE editorial team promises to the review and publish all the qualified submissions in a **fast** manner. All the journals articles are available online to the readers all over the world without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. Printed version of the journals is also available upon request of readers and authors.

IISTE Knowledge Sharing Partners

EBSCO, Index Copernicus, Ulrich's Periodicals Directory, JournalTOCS, PKP Open Archives Harvester, Bielefeld Academic Search Engine, Elektronische Zeitschriftenbibliothek EZB, Open J-Gate, OCLC WorldCat, Universe Digtial Library , NewJour, Google Scholar