Research Article

Received: 27 March 2021 / Accepted: 29 April 2021 / Published: 17 May 2021

Comparison of the Effectiveness of Four Scoring Techniques in Secondary School Multiple-Choice Test Items in Financial Accounting

Olaitan Isaac Fadipe

Department of Guidance and Counselling
Adekunle Ajasin University,
Akungba-Akoko, Ondo State, Nigeria

Dr. Yinusa A. Faremi

Department of Educational Foundations and Management
University of Eswatini, Kwaluseni, Swaziland

DOI: https://doi.org/10.36941/mjss-2021-0013

Abstract

Poor academic performance of students in senior secondary school certificate examination in some commercial subjects may be attributed to ineffective scoring techniques used by classroom teachers and examination bodies. Therefore, this study compared the effectiveness of four scoring techniques in multiple-choice financial accounting in secondary schools of Osun state, Nigeria. The study adopted a descriptive survey and Quasi-experimental research designs. The sample consisted of 420 senior secondary schools (SSSII) Financial Accounting students selected using Multi-stage sampling technique. An instrument titled "Financial Accounting Multiple-Choice Test (FAMT)" with a reliability coefficient of .80 was administered to selected students. The marks obtained through standard conventional, logical weight, corrected and confidence scoring techniques were analyzed using descriptive and inferential statistics. Findings showed that the logical weight scoring technique had the highest mean value of 40.63 when compared with other scoring techniques. It was also revealed that there was a significant difference in the performance of students in Financial Accounting multiple-choice test items using the four scoring techniques ($F(3,1676) = 31.494$, $p<.05$). Based on the findings, it was concluded that students performed better using the logical weight scoring technique in Financial Accounting multiple-choice test items and is the most effective scoring technique among others. The study recommended that a logical weight scoring technique should be encouraged and used in scoring Financial Accounting multiple-choice test items in secondary schools by examination bodies.

Keywords: *standard conventional, corrected, logical weight, confidence, scoring techniques*

1. Introduction

A test is an instrument for eliciting sample of behaviours which may be cognitive, affective and psychomotor domains about an individual or group of individuals (Kolawole, 2011). An achievement test is designed to assess how much knowledge a person has in a certain area or set of areas because of teaching. Ali (2006) viewed achievement test as an instrument administered to an individual or

group of individuals as a stimulus to elicit certain desired or expected responses, which represents his/her ability. A test is made up of questions or series of tasks designed for an individual to respond to independently and the results of which can be used to determine a quantitative academic change in individuals and for the quantitative comparison of individuals' performance or their level of achievement (Ugbamadu, et al., 2001). It may also be viewed as a summary of the evidence contained in a testee's responses to the items of a test that are related to the construct(s) being measured.

The type of test carried out in secondary school system fall under two categories: standardized achievement test and teacher-made test. A standardized achievement test is the type of test constructed by expert or specialist, administered and scored under standard and uniform testing. While the teacher made test constructed and administered by teachers for use within their classrooms (Sax, 1974). Teacher-made tests have been described as been subjective when compared to standardized tests (Alonge, 2004). The subjectivity of teacher-made test sometimes shows that the scoring lacks reliability.

Financial Accounting is the act of recording, classifying, summarizing, analyzing and interpreting financial statement used by individuals and organisation (Robert, 2009). Financial accounting is one of the essential subjects among the commercial subjects in senior secondary school that requires assessment to ascertain students' basic knowledge, skills and understandings of the concepts and the nature of Financial Accounting problems in any society. Nitks (1996) identified assessment as a process for obtaining information that is used for deciding curricula, programmes and educational policy (as cited in Kolawole, 2005). It is also a process of gathering and discussing information from multiple and diverse sources to develop a deep understanding of what students know, understand, and what they can do with their knowledge as a result of their educational experience. Assessment may be effective through the use of achievement test conducted during the teaching and learning process. Achievement test measures how much of the stated objectives of the course content achieved after completion of the course outline (Alonge, 2004).

Examiners make use of different evaluation instruments such as essay tests and objective tests. An objective test is one of the evaluation instruments used in testing or assessing students' academic achievement after a course of classroom instruction. In objective tests, such as multiple-choice questions, students are asked a question and required to pick the best possible answer(s) out of the choices from a list. Multiple-choice test items consist of a stem and a set of options. The stem is the questioning part that presents the problem to be solved, a question asked of the respondents, or an incomplete statement to be completed, as well as any other relevant information. The options are the possible answers that the examinee can choose from, with the correct option called the key and the incorrect options called distractors or foils. (Grounlund, 1976; Sax, 1974; Alonge, 2004). Multiple choice test items recognised as inevitable item format among other test formats in the assessment of the area of knowledge, aptitude and ability testing. Multiple choice consists of alternatives, one of which is correct (Key) or recognised at least the best, while the remaining alternatives are false called distractors (Socan, 2009).

Test scores obtained from the multiple-choice questions seems to be used to assess the competence of the students. Some of the advantages of the multiple-choice questions as reported by Alonge, (2004) and Kolawole, (2011) include:

- Multiple-choice tests can be used to measure both the lower and higher levels of the cognitive domain.
- They allow teachers to ask a large number of questions that adequately cover the course content.

When students are exposed to test, the responses of the students are mandatory to be evaluated. Evaluation of students' achievement must start from the planning of a test and the scoring techniques to be adopted in order to have a reliable result. For this study, the four scoring techniques adopted are standard conventional, corrected, logical weight and confidence scoring techniques.

In a research conducted by Ajayi and Omirin (2012) on the effect of two scoring methods on multiple-choice agricultural sciences test scores. The result revealed that the logical weight scoring

method was a better method that favoured the scoring of the students in multiple-choice Agricultural Science test. In another study conducted by Gardner- Medwin (2006) confidence scoring technique is another type of assessment technique in which a student is asked to show how confident he/she feels about the correct answer selected from the options provided to a question (as cited in Salehi, et al., 2015). This method can be used to improve the scoring of different types of objective tests. It was also found that confidence assessment scoring technique can be used in both formative and summative assessment to improve learning and scoring method respectively (Salehi, et al., 2015).

2. Statement of the Problem

Most students consider Financial Accounting as a difficult subject. Students' perception of any task, especially at the beginning, affects the outcome more than anything else (Maxwell, Mergendoller & Bellisimo, 2005). The declined of students' performance in Financial Accounting in internal and external examinations attributed to factors such as; the teaching approach, ability and interest of the students. It was observed that poor performance might have been attributed to ineffective scoring technique employed at the scoring phase of financial accounting during the internal and external examinations. Most of the study previously carried out focused on scoring of multiple-choice using logical choice weight and confidence method in science subjects. Only few studies had examined the effectiveness of different scoring techniques in Financial Accounting at secondary school levels. Also, there is scarcity of literature on how effective is the combination of these four techniques in determining students' performance in Financial Accounting in providing additional knowledge and scoring skills which other studies failed to provide. Based on the various scoring techniques and the probability that they may yield different test scores under the same scoring condition, there is need to compare the effectiveness of standard conventional, corrected, logical weight and confidence scoring techniques in financial accounting multiple-choice test items.

3. Purpose of the Study

Purpose of this study is to compare the effectiveness of four scoring techniques in Financial Accounting multiple-choice items in order to identify the best technique that attempts to improve the measure of examinee's ability in determining the weight of their choices.

4. Objectives of the Study

The objectives of the study are to:

1. Find out the most effective scoring techniques among standard conventional, corrected, logical weight and confidence scoring techniques in determining the performance of students in Financial Accounting multiple-choice items.
2. Determine the difference in the performance of male and female Financial Accounting students in multiple-choice items in the four scoring techniques.
3. Examine the difference in the performance of Financial Accounting students using the standard conventional, corrected, logical weight and confidence scoring techniques in multiple-choice objective test items in Financial Accounting.

5. Research Questions

1. How effective are standard conventional, corrected, logical weight and confidence scoring techniques when compared in determining the performance of students in Financial Accounting multiple-choice choice test items?
2. Is there any difference in the mean scores of male and female students in standard conventional, corrected, logical weight and confidence scoring techniques in Financial

Accounting multiple-choice items?

6. Research Hypothesis

There is no significant difference in the performance of students in financial accounting multiple-choice test items using standard conventional, corrected, logical weight and conference scoring techniques.

7. Review of Related Literature

Test theory is concern about the construction and scoring of test item which focuses on psychometric analysis of data. Magno (2009) categorized psychometric analysis into two: Item Response Theory (IRT) and Classical test theory (CTT). Item Response theory which is sometimes referred to as the "strong true score theory" while the classical test theory is regarded as the "True score theory". Socan (2009) stated that the scoring of multiple-choice test items in the context of classical test theory received more attention in the past when compared to recent times while interest for the scoring issues is shifting towards the items response theory (IRT) in the psychometric theory. Both CTT and IRT can be used for the assessment of individual change in clinical contexts. Even though researchers are shifting towards IRT, the CTT uses a common estimate of measurement precision that assumed to be equal for all student irrespective of their levels. Jabrayilov, et al., (2016) found that it is generally better at determining a change in an individual in short test using CTT. Socan (2009) also pointed out the relevance of classical test theory by comparing the following four methods listed below in scoring multiple choice test items within the framework of classical test theory in determining the validity and reliability of the test items.

- The number of the right score (NR)
- The guessing corrected score (GS)
- The first dimension obtained by homogeneity analysis (HA)
- The sum of the dummy variables weighted by the correlation weight (CW)

According to Open Psychometric Test Resources (2020) explained that for any given test there is going to be errors in test measurements and these errors are random variables that could be correlated and indexed. Through the correlation of the errors, improvement can be made to improve the quality of the test by increasing the reliability of the tests. The more reliable a test is the truer the score answers. Despite all the criticism on classical test theory is still one of the theories that predict the outcomes of psychological testing to determine the difficulty level of the item and the ability of the examinees. Classical test theory (CTT) starts from the assumption and explains the systematic effects between responses of the examinees as a result of variation in the ability of interest (Magno, 2009).

Classical tests theory (CTT) focuses on the total test score conducted. The construct of CTT operates on the summary of items, involve the sum of responses, average response or other quantification of the overall level. The fundamental features of classical test theory as explained by Tractenberg (2010) is the formulation of every observed score (X) as a function of the individual's true score (T) and random measurement error (E).

It can be expressed as $X = T + E$

Where

X = observe score

T = True score

E = measurement error

For the reasons pointed out by Magno (2009); Tractenberg (2010) and Jabrayilov, et al. (2016), the classical test theory was adopted for this study.

Standard conventional technique appears to be an old method and used in scoring students' academic performance in our senior secondary school. Under this method, only the correct options

termed key is counted for the students as his/her score in any given Financial Accounting multiple-choice tests. Using this scoring technique, it seems that it is faced with a major problem of guessing since provision is not made for a penalty. The conventional multiple-choice test is the type of test in which each question consists of a stem and options called choice. Out of these options or choices, one is the correct answer while the remaining options referred to incorrect called distractors (Jennings & Bush, 2006).

In standard conventional scoring technique candidate who knows the correct answer to the questions assumed of scoring 1 mark, while other answers may be guessed. Unlucky guesses are not penalized. The candidate may be guessing where the answers to the questions are not known. This type of scoring technique can also refer to as "number – right scoring". As a result of the fact that penalty is not made against students' blind guessing under standard conventional scoring technique, Corrected scoring techniques came into place because of its weakness at permitting guessing and refusing to penalize student where he chooses the answer to the item through guessing.

Corrected technique is when a testee is penalised by 1/N-1 mark where he chooses a wrong option. Where N represents the number of options in an item (alternatives). It is assumed that candidate either know to answer particular items correctly or not. If the candidate has the knowledge, he/she gets the correct answer, but if not he/she will guess. The incorrect response was the result of a random guess among all the options in the item. Any incorrect response attracts a penalty of 1/N-1 point while correct response attracts a point. When using the corrected scoring technique, a candidate has to make use of N option for each item in a test. If the candidate guesses in corrected technique, he/she had a 1/N chance of getting it correct due to random guessing and an N – 1/N this chance of getting it correct. If a candidate has R item correct and W items wrong in N of option in multiple-choice test. The final score of the candidate is

$$R - \frac{W}{N-1}$$

Where;

R= total correct options chosen

W= total wrong option

N=Number of alternatives (Kolawole, 2006).

Logical weight looks like a free choice test in which the candidate is allowed to choose more than one option out of the four or five options provided with each question. In free-choice, the candidate gains a fraction mark for any other options that are nearest to the correct answer (Key) that is, by selecting more than one answers (Jennings & Bush, 2006). Even though the correct answer is unknown once the candidate steal shows a partial knowledge of identifying answer that very close in each case. Kolawole (2006) found that differential weight is attached to the alternatives or the options following the degree of nearness to the correct answer. Scores are awarded to the testee according to the knowledge he/she has in the item being answered. The only student without any knowledge gets zero (0) mark when this scoring technique is used in scoring multiple-choice test. In a study conducted by Afolabi (1994) "Towards an accurate assessment of students cognitive capacity: An evaluation of three scoring procedure in objective testing". It was found that logical weight is more reliable.

The scoring of multiple-choice test under this technique requires some technicality. Suppose the test item is a five-option item, one (1) will be awarded to the correct best option, .75 to the second best option, .50 to the best third option, .25 to the best fourth option and zero (0) to the fifth-best option. But if the test item has 4 alternatives (options); the first best option will be awarded one (1), .67 to the second-best option, .33 to the third-best option and zero (0) to the last best option. Logical weight scoring method tends to be the best method due to the following reasons. Firstly, it will favour the testees because only those who have no idea about a particular item will get score zero (0). Secondly, it will force the testees to have full knowledge of the stem, key and alternative option of all the items of the test, above all the degree of nearness response (Kolawole, 2011). This method took into cognizance each of the option and the degree of closeness to the key. Also, it is paramount for

the testees to be able to have assurance on the correctness of option chosen when multiple-choice items are given to them.

In a research conducted by Ajayi (2014) compares logical choice weight and confidence scoring methods on multiple-choice Agricultural science test scores revealed that logical choice weight scoring method was best in scoring of the students' response in multiple-choice agricultural science test. Also, there was a significant relationship between the logical choice weight and confidence scoring method. In another research conducted by Awodele, et al. (2013) on the comparative effectiveness of logical-choice weight (LWM) and confidence scoring technique (CST) on reliability and validity of chemistry multiple-choice test items in Nigerian secondary school. The result revealed that confidence scoring method is better in reliability and validity of test scores when compared with Logical weight.

The examiners in **confidence scoring** assessment proposed different types of marking schemes. The difference in the number of certainty levels that examinees are chosen is marked for a correct and wrong answer. In most of the marking schemes used for scoring has three certainty levels. The three certainty levels identified by Gardner-Medwin (2006) as high, middle and low which is rated as C = 1, C = 2 and C = 3. The 1 point stand for low certainty levels, 2 points stand for mid certainty level, while 3 points stand for high certainty level (as cited in Salehi et al., 2015). The penalties for the wrong answer are not the same, it depends on the examiner that design the marking scheme.

The multiple-choice questions confidence scheme will be fully discussed in this study. Gardner-Medwin (2006) proposed a marking scheme for multiple-choice questions in which the probability of responding to the question by chance and getting the answer right in less than 50 percent (as cited in Salehi et al., 2015). The scheme specifies that candidate that answer the questions correctly at C = 1, C = 2 and C = 3 receives 1, 2 and 3 points respectively. When candidate confidence to his/her answer less than 50 percent choose C = 1, when candidate confidence falls between 50 and 75 percent choose C = 2 and above 75 percent choose C = 3. While a wrong answer at C = 1, C = 2, and C = 3 a candidate receive 0, - 1, and -4 points respectively. In another confidential scheme proposed by Davies (2002) as cited in Salehi et al (2015), the wrong answer at C= 1, C = 2 C = 3, a candidate will receive -1, -2 and -3 points respectively. Salehi et al (2015) found that confidence assessment shows that a test produces more reliable results than a number–right scoring (standard conventional).

For Soderquist (1936) proposed three levels of confidence for testees to response to (as cited in Kolawole, 2006). The levels range from absolute confidence through partial knowledge to random or blind guessing. Examinees responses would then be stored thrice with the exclusion of the answer made by random or blind guessing and with only the answer made with absolute confidence. Confidence level in multiple-choice tests scored with absolute confidence has greater reliability and validity than other levels of confidence scoring (Abu-Sayf & Diamond, 1976 as cited in Kolawole, 2006). In a research carried out by Awodele, et al. (2013) on Difficulty and Discrimination Indices of Logical Weight and Confidence Scoring Methods (CSM) on Chemistry Multiple Choice Test in Nigeria Secondary Schools. The CSM was favoured to assess Chemistry students' performance.

Awaluddin and Sari (2017) investigated the effect of multiple-choice scoring methods Number Right Elimination Testing (NRET) and Formula Scoring (FS) toward Chemistry learning outcomes by considering students' risk taking attitude. The study was conducted in SMAN 13 Kota Bekasi. The result showed that there were differences in Chemistry learning outcomes between students who were tested with multiple-choice tests using Number Right Elimination Testing (NRET) scoring method and Formula Scoring (FS) method. There was an interaction effect between scoring methods on multiple-choice tests and risk taking attitude toward Chemistry learning outcomes. Awaluddin and Sari (2017) also revealed that male students perform higher than the female student in an objective test. Gafoor and Shilna (2014) found gender influence on the academic performance of male and female student in an achievement test. The difference may be as a result of high guessing tendency of male student in multiple-choice test items.

Machin and Pekkarinen (2009) investigated the phenomenon of 'higher variance' in boys'

educational performance. The investigation was to know whether the phenomenon of 'higher variance' is an accurate characterisation of boys' educational performance relative to girls, using data from the 'Organisation for Economic Co-operation and Development (OECD) Programme for International Student Assessment (PISA). A survey of 15-year-olds enrolled in full or part-time education in 41 industrialised countries. The researchers analyse students test scores in Mathematics and reading by country, focusing on differences in the mean and variance of the scores. For reading, they find that the boy-girl mean difference is negative in all 41 countries, indicating that girls generally outscore boys. In 35 out of 41 countries, the boy-girl variance ratio indicates that boys' scores have greater variance than girls' scores. In a study carried out by Attah and Ita (2017) on gender as a predictor of academic achievement in English showed that the mean value of female students was slightly higher than the students.

In another research conducted by Gwarjiko (2015) on the effect of mixed- gender streaming on students' performance in English Language revealed that the female students performed better than the male ones despite the performance of the two genders were generally poor. Okafor and Egbon (2011) found that male Accounting undergraduates performed better than their female counterpart.

8. Research Methodology

The descriptive survey and quasi-experimental research designs were adopted for this study. The survey research design enables the researchers to capture a larger number of students among the population of Senior Secondary School (SSS II) Accounting students. While the Quasi-experimental designs help to evaluate specific intervention and outcome to demonstrate causality between an intervention and an outcome (Harris, et al., 2006). The scoring techniques used in this study categorized the research as quasi-experimental research because the scoring method serves as a treatment in which the researchers measures the answers to the test items without a control group called posttest design, in which each of the scoring method and outcome represents one group. The population for this study comprises of all the senior secondary school Financial Accounting students SSS II in Osun State, Nigeria of three senatorial districts, in which two local government were selected from each district, 21 secondary schools and 420 students from public senior secondary schools were selected for the study using a multi-stage sampling technique.

An instrument titled Financial Accounting Multiple-choice Test (FAMT) of sixty multiple-choice objectives items of four options drawn from West African Examination Council (WAEC) 2017 and 2018 senior secondary school syllabuses in Financial Accounting. The test was administered on selected students after establishing the difficulty power and discriminating index of the items which reduced the items from 100 to 60 items. The instrument was also validated in term of content and face by an expert in the field of Financial Accounting in a secondary school and later subjected to Kuder-Richardson 21 reliability with a coefficient value of .80. The testees are expected to pick the correct option from the options provided. Three sub-options (confidence levels) were also provided for the testee to establish his/her level of confidence in picking an option.

After administering the instrument with the permission of the principals and consent of the students to be used for the study, the four scoring techniques were used in scoring the responses of the examinees: In standard conventional scoring technique, the correct answer to each question will be awarded 1mark, while incorrect answer will be awarded omark. In logical weight scoring technique, the weights were designed based on the nearness of the option to the key. Correct option (key) =1mark, second nearest option to the key =.66, third nearest option to the key =.33, last nearest option = 0. Confidence scoring technique used three certainty levels for both correct and incorrect answers. For correct answer; Low level =1, mid-level = 2 and high level =3. For incorrect answer; lower level =0, mid-level =-1 and high level =-4. While no reply for both correct and incorrect answer =0. A corrected scoring technique used R - $\frac{W}{N-1}$

The research questions were answered by presenting the frequency count, mean and standard

deviation values in a chart. While the hypothesis was tested at .05 level of significance using analysis of variance (ANOVA) and Scheffe Posthoc analysis in other to know where there is a difference among the students' performance in the four scoring techniques in Financial Accounting multiple-choice test items.

9. Results

9.1 Research question one

How effective are standard conventional, corrected, logical weight and confidence scoring techniques when compared in determining the performance of students in Financial Accounting multiple-choice choice test item?

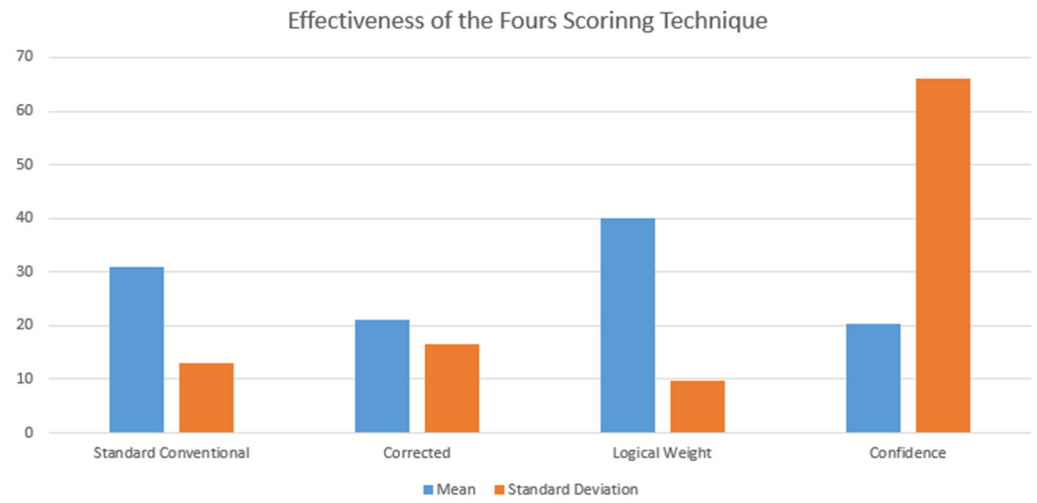

Figure 1: Effectiveness of the four scoring techniques in Financial Accounting multiple-choice test items.

Figure 1 shows that the logical weight scoring technique is the most effective scoring technique with the highest mean score of 40.63 and standard deviation of 9.678 in Financial Accounting multiple-choice test items when compared with standard conventional, corrected and confidence scoring techniques as presented in the figure above. This was closely followed by standard conventional with mean scores of 31.08 and standard deviation of 12.939 and corrected technique with mean scores of 21.09 and standard deviation of 16.546. While the confidence scoring technique had the least mean score of 20.33 and standard deviation of 66.033 in scoring Financial Accounting multiple-choice items test.

9.2 Research Question 2

Is there any difference in the mean scores of male and female students in standard conventional, corrected, logical weight and confidence scoring techniques in Financial Accounting multiple-choice items?

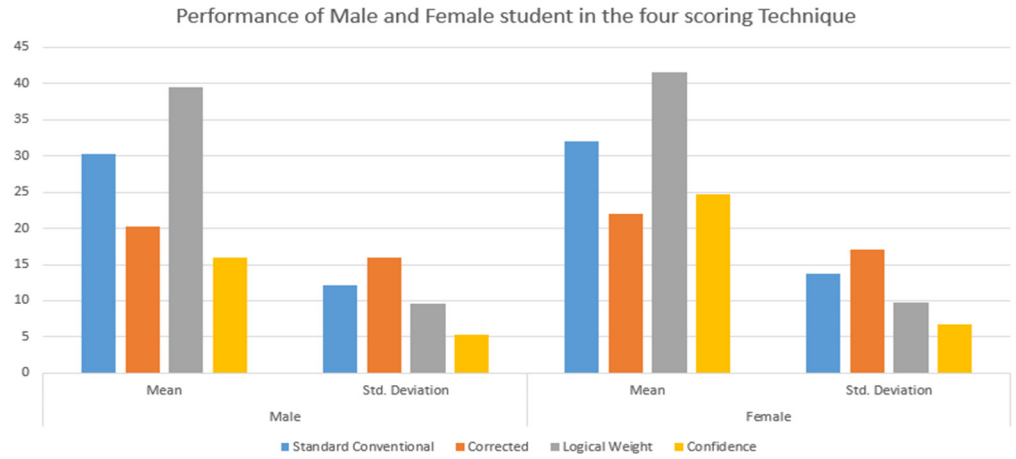

Figure 2: Performance of male and female students in the four scoring techniques

The result in Figure 2 shows the performance of female students (mean=41.60) score using the logical weight method in Financial Accounting multiple-choice items test which outweighed their male counterparts (mean=39.54). The performance of female students (mean=31.95) scores using a standard conventional scoring technique which had a slightly higher mean score than their male counterparts (mean=30.22). The performance of female students (mean=24.73) score using a confidence scoring technique in Financial Accounting multiple-choice items test outweighed their male counterparts (mean=15.93). The performance of female students (mean=21.96) score using the corrected technique in Financial Accounting multiple-choice items test outweighed their male counterparts (mean=20.23). The logical weight which is the most effective scoring technique, the female students perform better than their male counterpart in Financial Accounting multiple-choice test items.

9.3 Hypothesis

There is no significant difference in the performance of students in financial accounting multiple-choice test items using standard conventional, corrected, logical weight and conference scoring techniques.

Table 1: ANOVA showing the performance of students in financial accounting multiple-choice test items using the four scoring techniques.

Source	SS	Df	MS	F	P
Between Groups	115628.435	3	38542.812	31. 494	.000
Within Groups	2051097.993	1676	1223.805		
Total	2166726.428	1679			

*p<.05

Table 1 showed that the computed F-value ($F_{(3,1676)}$ =31.494, p<.05) was significant at p<0.05 level. The null hypothesis was rejected. This implies that there was a significant difference in the performance of students in financial accounting multiple-choice test items using standard conventional, corrected, logical weight and conference scoring techniques.

Table 2: Scheffe Posthoc analysis showing where differences occurred in the four scoring techniques.

Scoring Technique	1	2	3	4	N	Mean	SD
1. Standard conventional (1)		*	*	*	420	31.08	12.939
2. Corrected (2)			*		420	21,09	16.546
3. Logical Weight (3)				*	420	40.63	9.678
4. Confidence (4)					420	20.33	66.033

*p<0.05

Table 2 indicates a significant difference between the performance of Financial Accounting students in the standard conventional and corrected scoring techniques. Similarly, there was a significant difference between the mean of standard conventional and logical weight, standard conventional and confidence, corrected and logical weight, logical weight and confidence at 0.05 level in each case. While there was no significant difference exist between corrected and confidence scoring techniques.

10. Discussion

The outcome of the study revealed that the logical weight scoring is the most effective scoring techniques out of the four in determining the performance of students in Financial Accounting multiple-choice test item. This showed that students perform better when logical weight scoring technique was employed in scoring students' response in Financial Accounting. Followed by standard conventional and corrected techniques respectively. While students had the least scores when confidence scoring technique was used. The result is in line with the finding of Ajayi and Omirin (2012) which indicated that the logical weight scoring technique was a better method that favoured the scoring of the students in multiple-choice Agricultural Science test. While the result of this study contradicts the findings of Awodele et al. (2013) that the confidence scoring method provides good reliability and validity of test scores when compared with Logical weight.

A glaring difference in the performance of male and female students was revealed when logical weight, standard conventional, confidence and corrected scoring techniques were used in scoring students in Financial Accounting multiple-choice items. The female students' performance in the four scoring techniques outweighed their male counterpart. The result complements the finding of Machin and Pekkarinen (2009) who found gender influence on the educational achievement of students. It is also in line with the finding of Gafoor and Shilna (2014) who found gender influence on the academic performance of male and female students in achievement test. The difference may be as a result of high guessing tendency of male student in multiple-choice test items. However, the finding of Okafor and Egbon (2011) contradict the findings of this study and found that male Accounting students perform better than female students. The findings of the study also contradict that of Awaluddin and Sari (2017) that male is better than female academically.

The result in the hypothesis formulated indicated that there was a significant difference in the performance of students in Financial Accounting multiple-choice test items using standard conventional, corrected scoring, logical weight and confidence scoring techniques. The significant difference occurred between the standard conventional and corrected scoring techniques, conventional and logical weight, conventional and confidence, corrected and logical weight, logical weight and confidence in each case. In a related study conducted by Ajayi (2014) it was revealed that there was a significant relationship between the logical choice weight and confidence scoring method which contradict the findings of this study.

11. Conclusion and Recommendations

The logical weight had the highest mean value when compared to other scoring methods in scoring Financial Accounting. This implies that logical weight scoring is more effective in determining the

performance of students in Financial Accounting. Also, the four scoring methods when used are not yielding the same result. Since the logical weight had the highest mean value, it was recommended among the four scoring techniques analyzed in this study for use in schools, public examination bodies such as West African Examination Council (WAEC), the National Examination Council (NECO), Joint Admission and Matriculation Board (JAMB) in scoring students' performance in Finance Accounting. Seminars should be organized to expose secondary school teachers who are already on the job to the appropriate scoring methods for multiple-choice test items.

Reference

Afolabi, E. R. I. (1994). Towards an accurate assessment of students cognitive capacity: An evaluation of three scoring procedure in objective testing. *OSUA Journal of Education*, 1, 42-50.

Ajayi, B. K., & Omirin, M. S. (2012). The effect of two scoring methods on multiple choice Agricultural Science test scores, *Review of European Studies, 4*(1), 21-25. http://dx.doi.org/10.5539/res.v4n1p255

Ajayi, B. K. (2014). Comparison of logical choice weight and confidence scoring methods on multiple choice agricultural science test scores. *Journal of Education and Practice* 5(26), 25-30. https://www.iiste.org/Journals/index.php/JEP/article/view/ 15942/16280

Ali, A. (2006). *Conducting research in education and the social sciences*. Enugu: Tashiwa Networks Ltd.

Alonge, M. F. (2004*). Measurement and evaluation*. Ado-Ekiti: Adedayo Printing (Nig) Limited.

Attah, R. F. & Ita, P. M. (2017). Gender as predictor of academic achievement in English among senior secondary school two students in Calabar Metropolis, Cross River State. *Global Journal Of Educational Research,* 16, 149-153. DOI: http://dx.doi.org/10.4314/gjedr.v16i2.9.

Awaluddin, T., & Sari, F. (2017). The effect of multiple choice scoring methods and risk taking attitude toward Chemistry learning outcomes. *Journal of e-Learning and Higher Education*. https://doi: 10.5171/2017.496122.

Awodele, B. A., Adediwura, A. A., Bamidele, E. F., & Ajeigbe, T. O. (2013), Comparative effectiveness of logical-choice weight and confidence scoring methods on reliability and validity of Chemistry multiple-choice test items in Nigerian secondary schools. *Journal of Educational and Social Research*, *3* (2), 9-13.

Awodele, B.A., Faremi, Y.A., Adetunji, A. A & Bamidele, E. F. (2013). Difficulty and discrimination indices of logical-choice weight and confidence scoring methods on Chemistry multiple choice test in Nigeria secondary schools. *Journal of Education and Practice, 4*(6), 17-20.

Betts, L. R., Elder, T. J., Hartley, J. & Trueman, M. (2009). Does correction for guessing reduce students' performance on multiple-choice examinations? Yes? No? Sometime? *Assessment & Evaluation in Higher Education*, *34*(1), 1–15. https://doi.org/10.1080/02602930701773091

Gafoor, K. A & Shilna, V. (2014). Gender fair efficacy of concept mapping tests in identifying students' difficulties in high school organic chemistry. *Collaborative International conference on "Learning Environment for Excellence in Education*". https://files.eric.ed.gov/fulltext/ED545404.pdf.

Gronlund, N. E. (1976). Measurement and evaluation in teaching (3rd ed.). New York: Macmillan Publishing Co., Inc.

Gwarjiko, U. I. (2015). Effect of mixed-gender streaming on students' performance in English language: a case study of English as a second language classroom in Niger State, Nigeria. *International Journal of Academic Research and Reflection, 3*(5), 12-18. https://www.idpublications.org/wp-content/uploads/2015

Harris, A. D., McGregor, J. C., Perencevich, E. N., Furuno, J.P., Zhu, J., Peterson, D.E., & Finkelstein, J. (2006).The use and interpretation of quasi-experimental studies in medical informatics. *Journal of the American Medical* Informatics Association. *13*(1), 16–23. https://doi: 10.1197/jamia.M1749

Jabrayilov, R., Emons, W. H., & Sijtsma, K. (2016). Comparison of classical test theory and item response theory in individual change assessment. *Applied Psychological Measurement, 40*(8), 559-572. https://doi: 10.1177/0146621616664046.

Jennings, S & Bush, M. (2006). A comparison of conventional and liberal (Free-choice) multiple-choice tests. *Practical Assessment Research & Evaluation 11*(8), 1-5. www.researchgate.net/publication/253352134Acomparisonof_ conventional

Kolawole, E. B. (2011). "*Test and measurement*", (Revised edition). Ikeja: Bolabay Publication

Kolawole, E. B. (2006). Principle *of test construction and administration*. Ikeja: Bolabay Publication.

Kolawole, E. B. (2005). Measurement and assessment in education. Ikeja: Bolabay Publication.

Machin, S., & Pekkarinen, T. (2009). Assessment of global sex differences in test score variability. *Journal of Science Education, 322*(5906), 1331-2. https://doi:10.1126/ science.1162573.

Magno, C. (2009). Demonstrating the difference between classical test theory and item response theory using derived test data. *International Journal of Educational and Psychological Assessment 1*(1), 1-11. https://files.eric. ed.gov/ fulltext/ED506058. pdf.

Maxwell, N. L., Bellisimo, Y., & Mergendoller,J. (2005). The effectiveness of problem-based instruction: A comparative study of instructional methods and student characteristics. *Interdisciplinary Journal of Problem-Based Learning 1*(2). https://doi: 10.7771/1541-5015.1026

Machin & Pekkarinen. (2009). Assessment of global sex differences in test score variability. *Journal of Science Education, 322*(5906):1331-2. https://doi: 10.1126/science.1162573.

Okafor, C. A. & Egbon, O. (2011). Academic performance of male versus female accounting undergraduate students: Evidence from Nigeria. *Higher Education Studies, 1* (1). https://doi:10.5539/hes.v1n1p9

Open Psychometric Test Resource (2020). Classical test theory. *Psychometric Tests.* https://www.psychometrictest.org.uk/classic-test-theory/

Robert O. I. (2009), *Financial Accounting made simple*, Isolo: ROI Publishers.

Salehi, M., Sadighi, F., & Bagheri, M. S. (2015). Comparing Confidence-based and Conventional Scoring Methods: The Case of an English Grammar Class. *The Journal of Teaching Language Skills (JTLS) 6* (4), 123-152. http://jtls.shirazu.ac. ir/article_3105.html.

Sax, G. (1974). *Principles of educational measurement and evaluation.* California: Wadsworth Publishing Conpany, Inc.

Sočan, G. (2009). Scoring of multiple choice items by means of internal linear weighting. *Review of Psychology* 16,(2), 77-85. Retrieved from http://mjesec.ffzg.hr/revija.psi/.

Tractenberg, R. E. (2010). Classical and modern measurement theories, patient reports, and clinical outcomes. *Contemp clin trials 31*(1): 1–3. https://doi: 10.1016/S1551-7144(09)00212-2

Ugbamadu, K. A., Onwuegbu, O. C., & Osunde, A.U. (2001). *Measurement and evaluation in education*. Benin-City: World of books Publishers.