THE INTERNATIONAL JOURNAL OF HUMANITIES & SOCIAL STUDIES

Recent Issues in the Construction and Scoring of Multiple-Choice Items in Examinations

Dr. Emaikwu, Sunday Oche
Associate Professor, Department of Educational Foundations & General Studies
College of Agricultural & Science Education, Federal University of Agriculture
Makurdi, Benue State, Nigeria

Abstract:
Multiple-choice tests are widely attractive assessment techniques that have captured the attention of many standard examination bodies all over the world. They can be used to objectively assess cognitive, analytical and other skills in addition to straightforward factual knowledge. This article discusses the rationale and technical hitches associated with multiple–choice items in the measurement of ability. The paper presents some fundamental rules for constructing multiple-choice items. The paper maintains that many traditional marking schemes, with no adjustment to counter the effects of lucky guesses give rise to artificially high marks. In this article, a number of improved scoring schemes have been proposed for multiple-choice tests, some of which may require the candidates to do more than just select one answer per question. The schemes succinctly presented in this article would arguably be an improvement over the traditional scoring method as they could lead to the detection of partial knowledge of examinee and might have the potential of providing various kinds of examinee feedbacks that may invariably enhance learning.

Keywords: *partial knowledge, test construction, test scoring, multiple-choice items, examination.*

1. Introduction

The main mission of educators consists of determining learning progress and diagnosing learning difficulty experienced by students when studying. Testing is a conventional means of evaluating students, and test scores can be adopted to observe learning outcomes. The examination process in many universities is bedeviled by multifarious obstacles ranging from missing scripts/scores to late release of results among others. Currently in many universities, there is a pressure for speedy reporting of student's examination results not minding the fact that the groups to be tested are large as occasioned by the 21st century population explosion in tertiary institution. The only panacea to late processing and release of results is the gradual adoption and use of multiple choice tests in examinations. Sequel to this, the attention of many psychometric connoisseurs world over is currently turning towards the use of multiple choice tests in the measurement of ability. Corroborating this issue, Emaikwu (2014) affirms that with the urgency with which results are being demanded by the senate of universities few days after the conduct of examination, the only remedy to address late submission of results, especially in the courses which involve many students is to incorporate multiple-choice tests in the measurement of ability to ease marking. Multiple-choice items have come to be one of the most popular forms of items used in tests and examinations in recent years and make up probably the most valuable and unbiased type of tests today.

Multiple-choice item is used in examinations, in elections to choose among multiple candidates, parties, or policies; in market research and many other areas. Frederick J. Kelly has been credited with creating multiple choice items in 1914 at the University of Kansas. One of the first uses of multiple choice questions according to him was to assess the capabilities of World War I military recruits. While many people often colloquially referred to multiple-choice item as "questions," this is a misnomer because many items sometimes are not phrased as questions. They can be presented as incomplete statements or mathematical equations (Ebel, & Frisbie, 2007). The use of multiple choice questions in certain educational fields is sometimes contested due to some of the negative aspects, whether actual or perceived, but the format remains popular due to its utility and cost effectiveness.

Several sensitive issues have come up over the years while using multiple choice test items in the measurement of ability. The fast improvement in data–processing equipment, which helps to eliminate the burdensome clerical and statistical labour necessarily associated with the scoring of multiple-choice items, has led to a substantial boost in the study of more elaborate techniques in this area. Until the middle of the twentieth century, conventionally scored multiple –choice test items were customarily administered on the examinees without sufficient instructions and information related to the scoring methods or the recommended strategy (Abu-Sayf, 2005). However, as a result of today's growing moral concerns, it is being increasingly advocated that every test ought to be preceded

by instructions informing the examinees about the strategy to follow in order to maximize their scores. In addition, when a scoring formula is used, it is also desirable to inform the examinees about it with the maximizing strategy. Unfortunately these salient aspects are often neglected in the measurement of ability and their consequence undermined (Traub, 1998). Many psychometrics experts have argued that it is expedient to provide examinees with sufficient information so that they can formulate an optimal strategy for working the test without however, supplying them with the strategy. It is worth mentioning that the scoring system which comprises the scoring formula and the accompanying instructions serve as an indivisible entity. It is thus senseless to give instruction pertaining to a certain scoring method, then compute the test statistics on the basis of a number of different scoring formulae for comparison purposes as often done in some standard examinations in Nigeria. This article discusses issues in the construction and scoring of multiple-choice items in examinations.

2. The Concept and Rationale for Multiple–choice Item in the Measurement of Ability

A multiple choice item is designed for objective measurement and contains a stem and response options, one of which is the correct answer (Murayama, 2009). It is a kind of test item in which some options are given and the examinee is expected to pick the correct one out of those options provided. Multiple-choice items consist of a stem and a set of options. The stem is the beginning part of the item that presents the item as a problem to be solved, or a question asked of the examinee, 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 answer called the key and the incorrect answers called distracters. A multiple choice item is expected to be as long as necessary to ensure maximum validity and authenticity to the problem at hand. The stem ends with a lead-in question which describes what the examination taker must do. The stem is expected to be expressed clearly and concisely, avoiding poor grammar, complex syntax, ambiguity and double negatives. Negative statements are not characteristic of normal thought processes, and consequently may place the candidate who is attempting to decipher the item at a disadvantage (Emaikwu, 2012). If a negative question is used, it is expected to be emphasized with italics or underlining.

The options are divided into two and they include the key and the distracters. The key is the correct option, while the distracters are the options which appear to the examinees to be the correct answers but are not correct in the actual term. The correct answer should be clearly correct to the informed. The other distracters should also seem plausible to the candidates who have partial, incomplete or inappropriate knowledge. The distracters may be considered as the logical misconceptions of the correct answer. Multiple-choice test has long been the most widely used among the objective test formats

Examinees have less chance of guessing the correct answer to a multiple-choice test question if the distracters are plausible (Gronlund, 1988). In a five-choice item, a student who knows nothing about the subject matter has a pseudo-chance parameter of one-out-of-five chances of choosing the correct answer by random guessing. Multiple-choice items have continued to dominate educational testing owing to their ability to effectively and efficiently measure constructs such as ability and achievement. Measurement experts and testing organizations prefer the multiple-choice item format to others (e.g., short-answer, essay, and constructed-response) for obvious reasons.

If item writers are well trained and items are quality assured, this can be a very effective item format. First of all, if students are instructed on the way in which the item format works and the myths surrounding the assessment type are destroyed, students are found to perform better on multiple choice tests than any form of objective tests. On many assessments, reliability has been shown to improve with larger numbers of items on a test, and with good sampling and care over case specificity, overall test reliability of multiple-choice item can be further increased (Murayama, 2009)).

Multiple choice tests often require less time to administer for a given amount of material than would tests requiring written responses, meaning that more questions can be given in the assessment without increasing the time needed; this results in a more comprehensive evaluation of the candidate's extent of knowledge. Even greater efficiency can be created by use of online examination delivery software. Multiple choice tests lend themselves to the development of objective assessment items. Because this style of test does not require a teacher to mark the given answers arbitrarily, test-takers are graded purely on their selections, creating a lower likelihood of teacher-student bias in the results. Scoring is objective as it cannot be influenced by factor such as poor handwriting of examinee. Factors irrelevant to the assessed material, such as handwriting and clarity of presentation, do not come into play in a multiple choice assessment, and so the candidate is marked purely on their knowledge of the topic. If test-takers are aware of how to use mark sheets and/or online examination tick boxes, their responses can be relied upon with clarity.

However, multiple choice tests do have some technical hitches and drawbacks. A free response test allows the test taker to make an argument for their viewpoint and potentially receive credit. Even if a student has some knowledge of a question, they receive no credit for knowing that information if they select the wrong answer. However, free response questions may allow a test taker to demonstrate their understanding of the subject and receive partial credit. More so, test takers may be able to rule out answers due to infeasibility. In some cases they may even test each answer individually, especially when dealing with mathematics, thereby increasing the chance of providing a correct answer without actually knowing the subject matter. On the other hand, especially on mathematics tests, some answers are included to actually encourage the test taker to logically rule out responses. An example would be giving the equation $4x^2 + bx = 3$ and asking what b^2 equals. The test taker should be able to eliminate all answers that are a *negative number*. It should be noted that in some cases the candidate receives partial credit for certain incorrect, yet somewhat plausible, selections (and might be penalized for other, less suitable ones), but this is rare and is still not a solution that is as optimal as that offered in a written examination.

Another drawback of multiple choice examinations is that a student who is incapable of answering a particular question can simply select a random answer and still have a chance of receiving a mark for it. It is common practice for students with no time left to give all remaining questions random answers in the hope that they will get at least some of them right. Some examinations have systems in place to negate this, in this case by making it more beneficial to not give an answer than to give a wrong one. This is usually not a great issue, however, since the odds of a student receiving significant marks by guessing are very low when four or more selections are available.

3. Characteristics of Item Writers and Basic Rules for Constructing Multiple-choice Items

Writing multiple choice items is a matter of precision, perhaps more akin to computer programming than to writing prose. A multiple choice item must focus on the attention of the examinee on the principle or construct upon which the item is based. Ideally, students who answer a test item incorrectly will do so because their mastery of the principle or construct in focus was inadequate or incomplete. Any characteristic of a test item which distracts the examinee from the major point of focus of the item reduces the effectiveness of that item. Any item answered correctly or incorrectly because of extraneous factors in the item, results in misleading feedback to both examinee and examiner. Given a task of precision of communication, there are several attributes or mind sets that are characteristics of a proficient item writer.

One of the essential characteristics of multiple choice item writers is knowledge and understanding of the material being tested. At the University level, the depth and complexity of the material on which students are tested necessitates that only faculty members fully trained in a particular discipline can write concise, unambiguous test items in that discipline.

Another essential characteristic of multiple choice item writers is continuous awareness of the objectives. A test must reflect the purposes of the instruction it is intended to assess. This quality of a test, referred to as content validity, is assured by specifying the nature and/or number of items prior to selecting and writing the multiple choice items. Instructors sometimes develop a chart or test blueprint to help guide the selection of items. Such a chart may consider the modules or blocks of content as well as the nature of the skills a test is expected to assess.

In the case of criterion-referenced instruction, content validity is obtained by selecting a sample of criteria to be assessed. For content-oriented instruction, a balance may be achieved by selecting items in proportion to the amount of instructional time allotted to various blocks of material using test blueprint

Furthermore, continuous awareness of instructional model is another essential characteristic of multiple choice item writers. Different instructional models require items of quite different characteristics for adequate assessment. Ideally, item discrimination (the degree to which an item differentiates between students with high test scores and students with low test scores) should be minimal in a mastery-model situation. We would like to have all students obtain high scores. In the normative-model, item discrimination should be as high as possible in order that the total test differentiates among students to the maximum degree.

In addition, understanding of the students for whom the items are intended is essential characteristic of multiple choice item writers. Item difficulty and discrimination are determined as much by the level of ability and range of ability of the examinees as they are by the characteristics of the items. Normative-model items must be written so that they provide the maximum intellectual challenge without posing a psychological barrier to students' learning through excessive difficulty. In either the normative or mastery models, item difficulty must not be so low as to provide no challenge whatever to any examinee in a class.

It is generally easier to adjust the difficulty than to adjust the discrimination of an item. Item discrimination depends to a degree on the range of examinee ability as well as on the difficulty of the item. It can be difficult to write mastery-model items which do not discriminate when the range of abilities among examinees is wide. Likewise, homogeneous abilities make it more difficult to write normative-model items with acceptably high discriminations. No matter what the instructional model or the range of abilities in a class, the only way to identify appropriate items is to select them on the basis of subjective judgment, administer them, and analyze the results. It is only items of appropriate difficulty and discrimination that may be retained for future use.

There are many helpful hints and lists of pitfalls to avoid which may be helpful to the item writer. This is an area where measurement specialists may be particularly helpful.

1. Design each item to measure an important learning outcome.
2. Present a single clearly formulated problem in the stem of the item.
3. Put much of the wording as possible in the stem.
4. Eliminate unnecessary or redundant words from multiple-choice items
5. Avoid negatively worded stems, but where it is unavoidable, emphasize it in the stem of an item.
6. Ensure that the intended answer is correct or clearly the best.
7. Do not have linked or clued items. A *stem clue* occurs when an identical or similar term appears in both the stem and an option,
8. All options should be plausible and relatively homogeneous.
9. Put repeated words in the stem, not in the options
10. Make all options grammatically consistent with the stem of the item.
11. List options logically and if possible, vertically.
12. Ensure that each item is independent of the other items in the test.
13. Vary the position of the correct options in a random manner.
14. Make the distracters plausible and attractive to the uninformed
15. Use the option "all of the above" and "none of the above" sparingly. Knowing that two options are correct, could lead a clever candidate to choose "all of the above" without knowing the importance or correctness of the remaining responses

To make our higher-level multiple-choice questions more effective, we can use justification to assess reasons behind an answer. We can ask students to specify why they choose their answers and this allows them to demonstrate knowledge at the higher order level.

4. Alternative Scoring Schemes for Multiple-choice Test Items

A number of alternative marking schemes for multiple choice tests have been proposed and used over the years for written tests which are an improvement over the traditional scheme. In some cases these require candidates to do more than just select one answer per question.

The multiple choice test can be scored manually or mechanically using optical mark readers (OMR). In scoring for multiple choice tests, an examinee's score is simply the number of items answered correctly. In order to discourage guessing, standard examination bodies in Nigeria like JAMB, WAEC, NECO use a correction formula, based on the assumption that the examinee who does not know the answer, will resort to a random guess. The scores after correction will represent real attainment which does not include lucky guesses. Though the examinee's success in guessing right after thoughtful consideration is usually a good measure of his achievement, the argument that scores uncorrected for guessing will be too high carries some weight. The logical objective of most guessing correction procedures is to eliminate the expected advantage of the examinee who guesses blindly in preference to omitting an item. This can be done by subtracting a fraction of the number of wrong answers from the number of right answers, using the formula

$$S = R - \frac{W}{K-1}$$ Where,

S is the score corrected for guessing,
R is the number of right answers,
W is the number of wrong answers, and
K is the number of choices available to the examinee in each item.

4.1. Permutational Questions

The likelihood of gaining marks through sheer guesswork can be drastically reduced by having questions with multiple correct answers, where the candidate has to select the correct permutation of answers to get the question right. Faced with a four-answer question for example, a candidate would have to choose correctly amongst 16 (i.e. 2^4) alternative responses, assuming the number of correct answers is unknown. This is a rather exacting test method, but the marking scheme can be relaxed in some way to reward students whose selections are nearly correct (Farthing, 1998).

4.2. Normalization and Negative Marking

If a traditional one-tick-per-question multiple-choice test is preferred, some form of mark adjustment can be applied to counteract the effect of lucky guesses. One approach is normalization, as depicted in Figure 1. For a test with four-answer questions the formula can be written as x→(x–25)*4/3 (where X is the unadjusted test mark) and would be used.

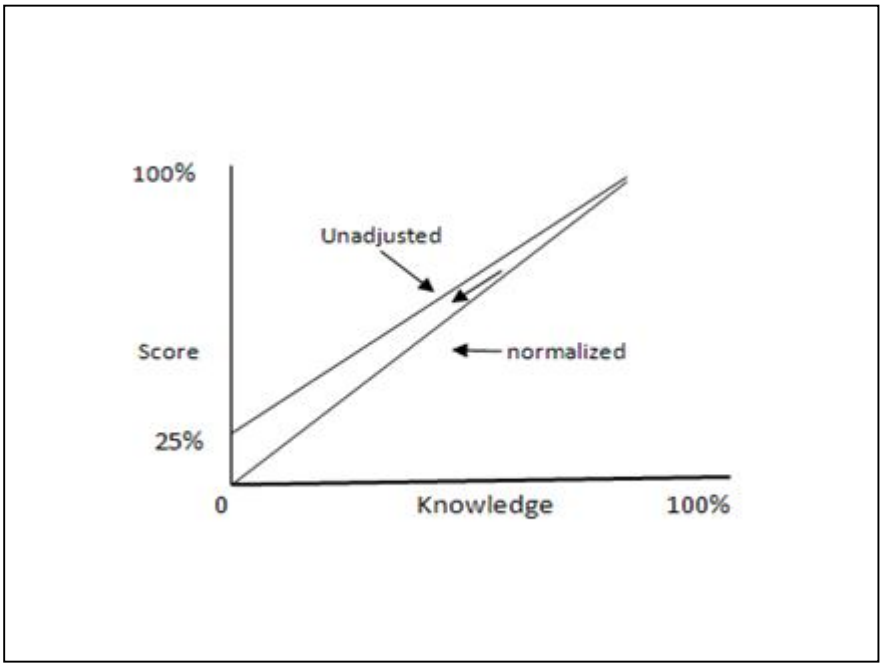

Figure 1: Normalization of marks (assuming four answers per question)

An alternative approach is to use negative marking as follows; award n–1 marks for every correct selection (where n is the number of answers per question) but subtract one mark for every incorrect selection. This has roughly the same effect as normalization, although it can result in a negative overall mark if a candidate is unlucky. Both of these approaches are well known and quite widely used.

4.3. Weighting

An alternative to the conventional scoring procedures has been tried through the weighting technique. The increase in test statistics resulting from this procedure is however, not big enough to justify the great deal of labour involved in it. This technique known as confidence weighting consists of having examinees cross out the alternatives they know to be incorrect in a four-choice item, giving them one unit of credit for each and subtracting 3 units (c-1) for crossing out the right alternative. Other techniques basically isomorphic with this one, consist of having the examinees rank the distracters by order of ease recognition or mark as many as alternatives as they think include the right answer; the less they mark, the bigger their credit. Another similar approach consist of asking the examinees to respond to a test item by assigning a probability between 0 and 1 to the event that certain choice is the correct answer, so that the score on every item ranges from -1 to +1, depending on the amount of confidence exhibited in the selection of the right answer. According to this approach, the examinees can maximize their scores if they honestly report what appears to the themes as true probability for the correctness of each alternative

4.4. Order-of-Preference Schemes

In an order-of-preference scheme, there is only one correct answer per question. Candidates are required to assign an order of preference to all the answers for each question; their score is dictated by the preference they assign to the correct answer. For example, 4 marks could be awarded if the correct answer is the candidate's first choice, 2 if it is their second choice, 1 if it is their third. Order-of-preference schemes have been investigated by several authors (Wood,1999***).***

4.5. Superitems

Another recent approach to the scoring of multiple-choice test items has been attempted through the development of what has been coined as 'superitems'. A superitem consists of a combination of more than one multiple-choice item selected on the basis of their inter-correlations so that each pair within a superitem has difficulty indexes symmetrically distant from the 0.50 value, leading to a score distribution approaching rectangularity. The advantages of superitem from the perspectives of test theory are many-fold. The directions to accompany such test exclude telling the examinees which items are to be grouped into superitems, since this information are rarely available prior to the analysis of the test results. A cutting score for each superitem is to be reached for the examinees to obtain a score of 1 on the superitem in the absence of which he gets a score of 0. If n_j represents the number of people who marked j of the subitems correctly, then

$N_j = n_{j+} n_{j+1+} n_{j+2\ldots+} n$ for j=0, 1,2,…

4.6. Liberal Multiple Choice Tests

The use of negative marking opens the door to allowing candidates to select more than one answer to a question if they are uncertain what the correct answer is (Bush, 2012). To appreciate its implications consider the following, which assumes four-answer questions:

i. If a candidate knows the correct answer to a question, he/she gets 3/3 = 100% for that question.
ii. If the candidate knows that the correct answer is one of two options, he or she gets (3–1)/3 = 67% for that question, compared with an equal chance of getting either 0 or 100% in a standard multiple choice test.
iii. If the candidate knows that the correct answer is one of three options, he or she gets (3–2)/3 = 33% for that question, compared with having a 33% chance of getting 100% in a standard multiple-choice test.

The probabilities assumed here may not always reflect the true situation, because a candidate may not have equal faith in each of the answers he or she has chosen to select. Nevertheless it seems reasonable to suppose that the mark achieved in a liberal multiple-choice test is more likely to be a better indicator of a candidate's true knowledge than the mark that same candidate would have achieved when faced with the same set of questions in a conventional test (even after normalization). This is because candidates are able to express partial knowledge explicitly in a liberal test. In effect, candidates are forced to choose between a much richer set of alternative responses.

4.7. Confidence Assessment

Multiple choice tests incorporating confidence assessment require candidates to assign a confidence level to each of their selections to reflect their degree of certainty. One such scheme as used on medical students, according to Gardner-Medwin (1995), requires candidates to attach a confidence level of 1, 2 or 3 to their selected answer for each question; this is the mark awarded if their selection is correct, while 0,–2 or –6 are awarded (respectively) otherwise. Given the nature of the medical profession, it does seem particularly appropriate to penalize medical students whenever they give a confidence but incorrect diagnosis!

4.8. Signal-Detection Theory

A different approach to the problem of guessing was undertaken by Miles (2011), and is based on the theory of signal detection. The underlying rationale consists of using the concept of signal-to-noise ratio, that is, the ability to discriminate some particular signal in a defined context of noise, to give an estimate of risk-taking tendencies. This can be done in psychometrics if the signal is defined as the relationship detected between a choice and the stem, and noise is the uncertainty attached to the correctness of this relationship. Using

two experiments, one with a verbal test and the other with a nonverbal test, Miles found that the data in the two cases fit the signal-detection model satisfactorily.

The importance of providing the examinees with scoring procedures and recommended strategies cannot be over-emphasized in a testing situation. It is worth mentioning that the scoring system which comprises the scoring formula and the accompanying instructions serve as an indivisible entity. For the number-right scoring formula, the maximizing strategy consists of answering every item even if there is no idea about the right choice. The accompanying instruction may read thus:

If you are sure of an answer to an item then mark it as explained
If you are not sure of an answer to an item, you can either guess or omit
a. if you guess correctly you will receive the point
b. if you guess incorrectly, you will be penalized by one-third of the point
c. if you omit the item, you will receive zero point

Since there is a penalty for wrong answers, in order for examinee to maximize his or her scores, it is expedient that examinees mark only the answers they know and avoid making random guesses It is thus absurd to give instruction pertaining to a certain scoring method, then compute the test statistics on the basis of a number of different scoring formulae for comparison purposes.

5. Views of Authors about Multiple-choice Item Scoring Schemes

From the viewpoint of learning, knowledge is accumulated continuously rather than on an all-or -nothing basis. The conventional multiple-choice item examination scheme requires examinees to evaluate each option and select one answer. Examinees are often absolutely certain that some of the options are incorrect, but still unable to identify the correct response (Bradbard, Parker, & Stone, 2004). The conventional scoring format of the multiple choice item examination cannot distinguish between partial knowledge and the absence of knowledge (Coombs, Milholland, & Womer, 2006). In conventional multiple-choice item tests, students choose only one response while the number of correctly answered questions is counted, and the scoring method is called number scoring. Akeroyd (2007) stated that number scoring makes the simplifying assumption that all of the wrong answers of students are as results of random guesses, thus neglecting the existence of partial knowledge.

Coombs, et al, (1956) first proposed an alternative method for administering multiple choice tests. In this procedure, students are instructed to mark as many incorrect options as they can identify. This procedure is referred to as elimination testing. Elimination testing using multiple choice test items makes guessing a futile effort. Bradbard, et al, (2004) suggested that the greatest obstacle in implementing elimination testing is the complexity of grading and the analysis of test items following traditional paper assessment. Accordingly, examiners are not very willing to adopt elimination testing. Bush (2014) presented a multiple-choice test format that permits an examinee that is uncertain of the correct answer to a question to select more than one answer. Incorrect selections are penalized by negative marking. The aim of these schemes is to reward examinees with partial knowledge over those who are simply guessing.

There are many situations in educational measurement in which students' attempts at a task can be categorized into several ordered levels of outcome. The use of multiple outcome categories is a common practice when scoring performances of complex task. But even in situations in which it is usual to score students' performances dichotomously (wrong/right), it is often possible to identify among students' 'incorrect' answer varying degrees of partial understanding, and so to define more than two levels of outcome on an item. For instance, using a test of basic skill in mathematics, a calculator shows the figure 25.634817. If the students are asked to express this value by correcting to two decimal places, the students may give a variety of answers to this item, but by far the most common answers given may be 25.63, 25.64, 2563.4517, and .25634817.

The usual dichotomous scoring of this item would give credit for the first of these answers and no credit for any other. However, the second answer, 25.64, shows a partial understanding: students who give this answer understand that correcting a number to two decimal places involves reducing to two the number of digits after the decimal point. These students appear to believe that because the original number is greater than 25.63 it must be rounded up to 25.64. The last two answers indicate no understanding of rounding and moving the decimal point two places (as in multiplication and division by 100). The most that can be said of these two answers is that they show some understanding of "two decimal". This is more than can be said for the other answers that students may give to this (e.g., 25.634.817). Most of these incorrect answers, which may be given by students, could display some understanding of rounding but reveal confusion about when to round up or down.

Consequent upon this, in an instructional setting, it would be inappropriate to treat every student giving an "incorrect" answer to this mathematics item in the same way. The type of instruction in rounding decimal numbers required by low-scoring students in this kind of mathematics test is likely to be very different from the instruction required by most high-scoring students and partial credit model could be applied

One major objection to non-traditional marking schemes has been that the marking process is more error-prone. Another is that unfamiliar marking schemes can be distracting to examinees at first, but this is only a transitional problem. A number of authors have reported that candidates do come to appreciate a sensible marking scheme once they become familiar with it.

Educational researchers have been continuously concerned not only about how to evaluate students' partial knowledge accurately but also about how to reduce the number of unexpected responses. It can be noted that the number of correctly answered questions is composed of two numbers: the number of questions to which the students actually know the answer, and the number of questions to which the students correctly guess the answer (Bradbard et al., 2004). A higher frequency of the second case indicates a less reliable learning performance evaluation.

6. Conclusion and Recommendation

There are many objections that may be made against multiple-choice tests such as the difficulty in designing good questions and the fact that a lot of questions are required. Nevertheless, multiple-choice tests have many important benefits, which of course is why they are in such widespread use. Multiple choice tests are especially appropriate when there is a pressure for a speedy reporting of scores as often witnessed in many universities in the recent time and when the group to be tested is large as occasioned by the 21st century population explosion in tertiary institution. It has been affirmed that simple normalization or negative marking can be used to counter the effects of lucky guesses without changing the nature of the test itself. Other marking schemes require candidates to do more than just tick one answer per question, and can be used to explicitly assess confidence or partial knowledge of examinees. This approach has the potential of providing various kinds of examinee feedbacks that may invariably enhance learning.

One objection to non-traditional marking schemes has been that the marking process is more error-prone. Another is that unfamiliar marking schemes can be distracting to examinees at first, but this is only a transitional problem. A number of authors have reported that candidates do come to appreciate a sensible marking scheme once they become familiar with it.

To make our higher-level multiple-choice questions more effective, it is imperative that we can use justification to assess reasons behind an answer. We can ask students to specify why they choose their answers. This allows students to demonstrate knowledge at higher order level.

Above all, multiple-choice test designers should consider taking the opportunity that automation provides to move away from the traditional marking scheme to a richer, more sophisticated one.

7. References

i. Abu-Sayf, F. K (2005). Relative effectiveness of the conventional formula score. Journal of Educational Research, 69, 269-270
ii. Akeoyd, F.M. (2007). Progress in multiple-choice scoring methods. Journal of Further and Higher Education, 6,87-90
iii. Bradhard, D.A., Parker, D. F. & Stone, G. L.(2004). An alternate multiple-choice scoring procedure in a macroeconomics course. Decision Sciences Journal of Innovative Education, 2 (1), 11-26.
iv. Bush, M. (1999). Allowing for uncertainty in multiple-choice tests. South Bank University Technical Report SBU-CISM-99-01.
v. Bush, M. (2014). Alternative marking schemes for on-line multiple-choice tests. London: South Bank University, UK.
vi. Coombs, C.H., Milholland, J.E. & Womer, F.K. (2001). Assessment of partial knowledge. Educational and Pyschological Measurement, 16,13-37.
vii. Culwin, F. Web Hosted. Assessment possibilities and policy proceedings of 3rd annual conference on integrating technology into computer science education (iticse' 98)
viii. Ebel, R. L., & Frisbie, D. A. (2007). Essentials of educational measurement (5th ed.). Englewood Cliffs, NJ: Prentice Hall.
ix. Emaikwu, S O. (2012). Fundamentals of test, measurement and evaluation with psychometric theories. Makurdi: Selfer Academic Press.
x. Emaikwu, S.O. (2014). Recent developments in the conduct of examination in Nigerian
xi. Universities: The clarion call for urgent utilization of objective test items. Indian Journal of Applied Research, .4 (7), 121-127
xii. Farthing, D.W. (1998) Permutational multiple-choice questions: An objective and efficient alternative to essay-type examination questions. Proceedings of 3rd Annual Conference on Integrating Technology into Computer Science Education (ITiCSE' 98)
xiii. Gardner-Medwin, A.R. (1995). Confidence assessment in the teaching of basic science. Journal of Association for Learning Technology, 3, 80-85.
xiv. Gronlund, N. (1988). How to construct achievement tests. New Jersey: Prentice- Hall.
xv. Marjorie, D & Nevart, Y (2013). Test construction manual: Construction of objective tests.Center for Teaching Excellence, University of Oklahoma. USA.
xvi. Miles, J. (2011). Eliminating the guessing factor in the multiple tests. Educational and Psychological Measurement, 75 (3), 637–651.
xvii. Millman, J. (2011). An analysis of test-wiseness. Educational and Psychological Measurement,25 (3), 707–726.
xviii. Murayama, K (2009). Improving your test questions. Center for Innovation in Teaching and Learning, University of Illinois
xix. Traub, R. E. (1998). The effect of scoring instructions and degree of speediness on the validity and reliability of multiple-choice tests. Educational and Psychological Measurement, 32 (3), 537-558.
xx. Wood, R. (1991). Assessment and testing: a survey of research . Cambridge University Press