

Comparative Analysis Of Students' Performance In Two Formats Of A Biology Practical Test In Taraba State

Chris Igomu Amuche¹ Naomi Thomas² Bulus Onesimus Shiaki³

1. Faculty of Education, Taraba State University, Jalingo, P.M.B 1167, Jalingo Taraba State, Nigeria
2. Faculty of Education, Taraba State University, Jalingo, P.M.B 1167, Jalingo Taraba State, Nigeria
3. Faculty of Education, Taraba State University, Jalingo, P.M.B 1167, Jalingo Taraba State, Nigeria

*E-mail of correspondence author: ehcumch2000@yahoo.com

Abstract

The study sought to compare the performances of senior secondary school students in two formats of a Biology practical test. A quasi-experimental study, involving two random groups, treatments post-test design was adopted. The two tests are multiple-choice test and fill in the answer tests formats (MCT and FIT). A random sample of 220 SS3 Biology students drawn from 1540 students that constitute the population was used. The sample was randomly grouped into two groups of 110 each and was randomly assigned to treatments. Then the tests were administered and the students' responses to the test were scored. Three hypotheses were tested. Mean scores, in the researcher-made Biology test, of students who took the MCT is significantly different from those who took the FIT format. Also, variances of student performances, in the researchers' made Biology test, of those who took the MCT is significantly different from those who took the FIT format, and, the mean score after correction for guessing of students in the MCT is not significantly different from the mean score of students in the FIT format of the same researchers' made Biology test. The study recommends that the sole or dominant use of MCT should be minimized. FIT should be incorporated in our secondary schools' Biology practical testing programmes.

Keywords: Performance, Biology, Practical, Test

1. Introduction

There is no gain saying about the fact that biology occupies a very sensitive position in medical science and related discipline. This informs several efforts geared toward studying biology at a secondary level of education. Hence, it is one of the science subjects one must pass so as to qualify to offer some science courses at tertiary level of education. It is however, very disheartening and heartbreaking that despite the key role and much emphasis, being laid on biology, students at secondary school level of education are still performing woefully in this subject has being an issue of great concern to stake holders in education, most especially those in the field of science. This has been attributed to myriad of factors such as poor parenting, poor attitude of students towards their studies (FRN, 2004). Bassey (2005) opined that several problems are associated with conventional method of teaching. This, indirectly result to poor performance of students. Bassey (2005) attributed this syndrome in educational sector to lack of perceived competence among the learners. However Adegbite [(2006) and Olaleye (2010) views on the causes of poor performance differs. They posited very strongly that teacher skills in assessment are related to students' achievement in a subject. This could be responsible for the mass failure particularly in science subjects in external examinations over the years.

Okooboh, Afolabi and Asilika (2004) stressed that the unimpressive response to science and technical education is particularly evident in students' poor performance in science subjects at secondary school level. In the words of Ajileye (2006) insufficient resources for the teaching and learning of science constitute a major cause of student underachievement. The insufficient resources include laboratories, science equipment, and specimens to be used as teaching aids. Onuoha (1997) identified shortage of qualified and dedicated teachers as the factor affecting student performance in science and that poor practical orientation will lead to poor understanding of the theory. In his opinion teachers are no more dedicated to their assignments. They give more time to trading, petty contracts, farming etc. They sneak in and out of the classrooms and laboratories at will. Ukwuma (1990) in his investigation of factor impair science education confirmed that over 80% of failure in science and technology are due to the inability of students to perform well in practical.

Test can be defined in different ways but for the purpose of this study, test is nothing more than a group of questions given to pupils to respond in order to know their achievement levels. The questions might require the pupils to give the correct meaning of a word, solve a mathematical problem or identify the missing parts. Anikweze (2010) conceptualized test (in a formal school system) as any kind of device for measuring, ability, achievement, interest and other traits. It could be a set of questions; task or problems intended to measure an individual's knowledge, skills, aptitude, intelligence etc.

Test is very important in education. Some tests are “verbal”, where the test is given verbally, no pen or paper is needed to answer the questions while some other tests are written, meaning a test where the pupils have to answer question using a pencil or pen, on paper. Tests can be classified according to modes of response, purpose of testing, mode of interpretation and method of administration. Teachers are familiar with paper-and pencil tests of the objective and essay types; but tests could also be oral or performance tests. No one type of test can be said to be the best. What really matters is the ability of the test type to measure aspects of learner achievement by recall or application of knowledge and by any other reliable demonstrations of change in behaviour after instruction. The focus of this study is on written (paper-and-pencil) tests which have well respected testing instruments that are widely used in schools. Written tests are of various formats which include;

- a) Supply types such as essay, short answered and, fill in the gap/completion.
- b) Selection types such as true/false or alternative response, matching and, multiple choices.

The interest of this study is on the Multiple Choice Test format (MCT) and Fill in the Gap test Format (FIT). MCTs usually have many options for each question, the tested is supposed to select the “best” choice among a set of four or five options.

Test markers often promote MCT as “objective” because there is no human judgment in the scoring, which could be done by machine. However, human beings decide what questions to ask, how the questions could be phrased and what distracters to use. All these are subjective decisions that can be biased in ways- that unfairly reward or harm some testees. Therefore multiple choice tests are not really objective. Anikweze (2010) highlighted some advantages and disadvantages of MCT. The advantages are as follows;

- (i) There is opportunity for extensive sample of the topics that have been treated during instruction thereby ensuring validity of the test in terms of comprehensiveness.
- (ii) Grading is objective and fast. It could be machine-scored but in the absence of machine, the teacher can still adopt ‘window marking’ to quicken the process of scoring.
- (iii) It eliminates bluffing associated with prolific writers - there is no chance for persuasive candidates to blow English that could carry away the teacher as many bright learners often do in essay type tests.
- (iv) It is easily adaptable for several teaching objectives as the different variants of objective type tests suit different instructional objectives.
- (v) It can be made highly valid for some teaching objectives particularly in the cognitive domain, i.e. for measuring knowledge of specific facts.
- (vi) An objective test can be made highly reliable through item analysis and further refinement.

Some disadvantages of the multiple-choice test were also highlighted:

- (i) It frequently neglects the measurement of higher thought processes such that creative and analytical students suffer.
- (ii) It may encourage, and indeed, over-emphasize rote learning to the detriment of meaningfulness of tasks.
- (iii) It promotes poor study habits if used as teaching instrument. The implication is that students may not bother to learn since they can always guess the answer.
- (iv) It encourages guessing in spite of the formula for discouraging guessing.
- (v) The setting of a meaningful objective test is difficult and time consuming.
- (vi) In general, it costs more than essay-type test to prepare and reproduce.

It is doubted that MCT process is natural. In fill-in the gap (FIT), the testees respond to the test items by supplying a word, phrase, number, symbol, or formula, etc. it normally consist of an incomplete statement. Some advantages of FIT format are that;

- i. It is easy to construct as options are eliminated.
- ii. It provides little opportunity for guessing as students are required to supply information.
- iii. It requires more than recognition and recall.

- iv. A relatively large content can be covered. Also, some disadvantages of FIT format are that;
 - i. Scoring is less objective than MCT.
 - ii. It may penalize the better students unduly and unduly help the weaker ones than MCT.

Some related studies have been done by other researchers. Stephen (2007) observes that MCT format has become the engine of most examinations. It has become the main testing format in most school both for internal and external examinations, but also disagreed with the belief that MCT items are very cheap to pass. It was opined that many students struggle with MCT questions and as a result do more poorly than they should. Anxiety and distractions can make a student to miss out some marks when writing a test of this kind. Cramming is possible in this test format but not in MCT. Sadier (1998) totally disagreed with the belief that MCT format is cheaper to pass than others. Accordingly, he posited that although some learners claimed that MCT items deals with the subjective views of any one teacher who may be biased or have low expectation, but there are many ways of addressing these problems such as introduction of distracters and by having independent group of teachers. Linda and Alicia (1987) suggested that MCT format focused upon helping examinees improve their scores by employing strategies for guessing when they do not know the correct answer, identifying unintended clues in the item or taking advantage of the idiosyncratic habits of the test constructor. Toube (2009) asserts that MCT format does not reveal test takers underlying reasoning for choosing a particular answer nor does it reflect test takers' ability to think critically under unprompted situations but measurement that allowed for responses in both MCT and FIT format makes it possible to assess individuals' spontaneous application of thinking skills on top of their ability recognize a correct response.

A major reason for using constructed-response questions in schools could be that a test taker who can choose the correct answer from a list may not be able to provide the answer without seeing it presented. Is the difference educationally important? Sometimes it is. Students who cannot remember the correct procedure for conducting a science laboratory experiment may recognize the correct next step, or the correct sequence of steps, when they see it. Students who cannot explain the logical flaw in a persuasive message may find it easy to identify the flaw when it is presented as one of four or five possibilities.

Students who cannot state the general scientific principle illustrated by a specific process in nature may have no trouble recognizing that principle when they see it stated along with three or four others. Making the multiple-choice questions more difficult by making the wrong answers more like the correct answer does not overcome this limitation. Instead, it can cause some test takers who know the correct answer (without seeing it presented) to miss the question by choosing one of the nearly-correct answers instead. For example, in biology, the test taker may be asked to describe the way a particular biological process occurs in a type of plant and explain how it enhances the plant's ability to survive or to reproduce. This test format however, does not cater for manipulative skills; rather it emphasizes recall of factual knowledge.

It is obvious that the purposes of instruction do not always correspond with the purposes of testing. Consequently, the types of tests used for assessing achievement must also vary. No one type of test could be described as the best (Anikweze, 2010). However, it is important that in selecting tests at any point in time, the teacher should ask himself/herself the following pertinent questions:

What is to be measured?

What was the objective of instruction?

How will the measurement be used?

What is the best type of test for this purpose?

The type of test commonly adopted by a teacher reveals the type of instructional objectives usually attained by that teacher. For instance, a teacher who prefers objective tests that demand only recall items could be accused of not reaching instructional objectives for reasoning, creativity and logical expression. These higher order cognitive objectives are better assessed with essay type question. There is therefore the great need for teachers to balance their test programme in such a way that various aspects of instructional objectives are assessed through the use of different types of measurement techniques. Some instructional objectives could be achieved through oral tests, some with objective tests, some with essay tests and others with performance tests. As posited by Green (1975.5): "Effective,

well-rounded instruction demands that a balanced testing program be used for adequate assessment of the varied instructional objectives”.

Some people in the field of educational testing have claimed that multiple-choice tests and constructed-response questions provide essentially the same information. Therefore, they argue, multiple-choice questions can be used as a substitute for constructed response questions (Lukhele, Thissen, & Wainer, 1994). These claims are based on research studies showing a high level of agreement between scores on multiple-choice and constructed-response questions (e.g., Godschalk, Swineford, & Coffman, 1966). However, those research studies generally have compared the multiple-choice and constructed-response scores of a single group of test takers who were tested once with both types of questions. The high level of overall agreement can mask important differences between groups of test takers. The acceptability and popularity of the use of MCTs are on the increase despite the various demerits of MCT format. Most external examination bodies for primary and post primary schools in Nigeria use MCTs for their examinations. Also most departments in many universities and tertiary institutions in Nigeria currently use MCTs for all their first and second years' examinations despite the facts that MCTs give room for testees to guess, perpetrate assorted forms of examination malpractices, exercise some undesirable test wiseness acts, work from answer to problems (in mathematical subjects examinations), etc. It is not certain that the formula used to correct for guessing works accurately. It is also doubtful that real life problem situations provide options that do serve as distracters. Could all these errors or problems be applicable to FITs? If not, could mean scores of students' performances in MCT not differ from mean scores in FIT? Could correction for guessing make scores in MCT not differ from scores in FIT? Could variance of scores of students' performances in MCT not differ from that in FIT?

Proffering answers to these posers form the problems of the study. The thrust of the study is to carry out psychoanalysis of the performance of students using MCTs and FITs format in researchers'-made practical Biology test. Specifically, the researchers sought to ascertain whether:

- (i) The mean scores, in the researcher-made practical Biology test, of students who took the MCT and those who took the FIT format do not differ.
- (ii) The variances of student performance, in the researcher-made practical Biology test, of those who took MCT and those who took the FIT formats do not differ.
- (iii) The mean scores of students in MCT format after correction for guessing and the mean score in FIT format in the same researcher-made practical Biology test differ.

2. Research Questions

In order to achieve the above objectives, the following research questions were framed;

- (i) What are the mean scores, in the researcher-made practical Biology test, of students who took the MCT and those who took the FIT formats?
- (ii) What are the variances of student performances, in the researcher- made Biology test, of those who took the MCT and those who took the FIT formats?
- (iii) What are the mean scores of students in MCT after correction for guessing and FIT format in the same researcher-made practical Biology test.

3. Research Hypotheses

Some hypotheses were framed and testes at a significant level of 0.05

1. The mean scores, in the researcher-made practical Biology test, of students who took the MCT are not significantly different from those who took the FIT format (H_{01})
2. The variances of student performances, in the researcher-made practical Biology test, of those who took the MCT are not significantly different from those who took the FIT format (H_{02})
3. The mean score after correction for guessing of students in the MCT is not significantly different from the mean score of students in FIT format of the same researcher-made practical Biology test (H_{03}).

4. Methods

The study would adopt the evaluative survey design. This is because the study is a sample survey. Sambo (2005) stated that a sample survey is a study in which a random sample is taken from a well defined population, data is collected from the sample, a statistic is calculated from the data, and the statistic is used to estimate the true parameter in the population. The study focussed on the comparative analysis of the performance of students in two test formats (MCT and FIT) in researcher-made Biology test. It involved the use of random groups, independent

variable; test format that has two qualitative value, MCT and FIT and dependent variable; students' performance scores in the test.

Rg X_1 0

Rg X_2 0

Where Rg = Random group

X_1 =MCT

X_2 = FIT

0=Post-test

The study was carried out on students in selected senior secondary schools in Jalingo, Taraba state. The population comprises about 250 secondary schools and 21,540 Biology students in Jalingo metropolis. From the population, a sample of 10 schools and 220 students were randomly selected using balloting. 20 Biology students were selected in SS3 classes from each of the 10 schools, to form the study sample. The instrument employed for data collection is researchers' - made Biology test. The test is of two formats;

a. Multiple test choice format (MCT). This was given to a group of 10 students in each school visited. The test comprises 35 items.

b. Fill in the gap test format (FIT). This was also given to the other group of 10 students in each school visited and it was also made of 35 items.

The researchers constructed 50 items based on text blue print for each of MCT and FIT format. The researchers presented the draft to other colleagues in measurement and evaluation and also two experts in Biology to check for content coverage, language expression, appropriateness, clarity and arrangement. Based on their vetting, each of the formats was reduced to 40 items, thus ensuring face, content validity and a logical validity index of 0.80. To ensure the reliability of the items, the researchers conducted a pilot study and used the scores of the testees in calculating the internal consistency reliability coefficient employing Kuder-Richardson formula 20. The reliability coefficients obtained are 0.76 for MCT and 0.77 for FIT.

Data analysis was carried out to obtain mean scores and variances of student performances as they concern the research questions. Also, correction for guessing was computed, Z and Fischer-test statistics were used to test the hypotheses.

5. Results

Hypothesis 1. The mean scores, in the researchers' made Biology test, of students who took the MCT is not significantly different from those who took the FIT format.

Table 1: Statistics used in answering Research Question I and testing H_{01} .

N_1	N_2	X_1	X_2	MCT (mean)	FIT (mean)	S^2_1	S^2_2	Z-cal	Z-tab
110	110	3203	3598	35.0	35.0	54.7	71.4	4.87	1.96

Decision: reject null H_{01} .

The students' mean scores, in the researchers' made Biology test, for MCT is 35.0 and FIT is 30.0. When subjected to Z-test, it was found that the Z-calculated (4.87) is greater than the Z-tab value (1.96). Therefore the mean scores, in the researcher-made Biology test, of students who took the MCT is significantly different from those who took the FIT format.

Hypothesis 2: The variances of student performances, in the researchers' made Biology test, of those who took the MCT is not significantly different from those who took the FIT format ($P < 0.05$).

Table 2: Statistics used in answering Research Question II and testing H_{02} .

N_1	N_2	S^2_1	S^2_2	F-cal	F-tab
110	110	54.7	71.4	1.31	1.00

Decision: reject null H_{02} .

The variance of student performances, in the researchers' made Biology test, for MCT is 54.7 and FIT is 71.4. Furthermore, F-test result indicates that F-cal value (1.31) is more than the F-tab value (1.00). Therefore, we can

accept the alternate hypothesis that variances of student performances, in the researchers' made Biology test, of those who took the MCT is significantly different from those who took the FIT format.

Hypothesis 3: The mean score after correction for guessing of students in the MCT is not significantly different from the mean score of students in FIT format of the same researchers' made Biology test ($P < 0.05$).

Table 3: Z test Analysis for testing Hypothesis 3.

N_1	N_2	X_1	X_2	1	2	S_1^2	S_2^2	Z-cal	Z-tab
110	110	3628	3598	30.2	30.0	87.9	71.4	0.17	1.96

Decision: fail to reject H_03 .

The mean score of students in MCT after correction for guessing is 30.2 and that of FIT is 30.0. Also, the Z-test result showed that Z-cal value (0.17) is less than the Z-tab value (1.96), therefore the mean score after correction for guessing of students in the MCT is not significantly different from the mean score of students in the FIT format of the same researchers' made Biology test.

6. Results

From the result of hypothesis one, the mean differed probably because of guessing, as those who took the MCT had opportunities of guessing while those who took the FIT could not guess. From hypothesis two, the variances differed because in the MCT, those who know less guess more, so they came closer to those who know more. They became more homogenous in the case of MCT and therefore had a lesser variance. Of course, the finding of hypothesis three is true because when the correction for guessing procedure was employed on the scores of those who took the MCT, the resulting mean score no longer differed significantly from the mean score of those who took the FIT.

7. Recommendations

Based on the findings of the study, the following recommendations were made:

1. For selection examinations, FIT is better since the variation is more. It means it can distinguish those who know more from those who know less.
2. Since the results of the performance of students do not agree, continuous assessment and examination should be made up of the two formats, so that error could be minimized.
3. Since the students' performance in the natural way of testing, FIT, is lower than the students' performance in MCT, then there be errors introduced by other behavioural tendency. Correction for such behavioural tendency should be made use

References

- Adegbite, P. R. (2010). Reflection of a school teacher, Lagos. Tusanmi publications, 2000.
- Anikweze, C. M. (2010). Measurement and evaluation for teacher education (2nd Ed.). Enugu: Snaap Publihers.
- Bassey, M. P. (2005). "Availability of resources for the teaching of science in public secondary schools." Afri Edu J, 2005; 1: 29-36.
- Chakraborty, G. (2000). A comparative analysis of three communication format. Journal of Advertising, vol 3, No 1.
- Federal Republic of Nigeria (2004). National policy on Education, Lagos; NERDC
- Godschalk, F. I., Swineford, F., & Coffman, W. E. (1966). *The measurement of writing ability*. New York: College Entrance Examination Board.
- Linda, C & Alicia, S. (1987), Comparative analysis of two test formats, Journal of experimental education vol 55, No 2.
- Lukhele, R., Thissen, D., & Wainer, H. (1994). On the relative value of multiple-choice, constructed-response, and examinee-selected items on two achievement tests. *Journal of Educational Measurement*, 31, 234–250.

Olaleye A.O. (1985). An education of Nigeria system since independence; Lagos. African University Press publication. 1985

Osuala, R. C & Ogomaka, P. M. C (2005), Effect of some teaching methods on students' performances in senior secondary school Biology, Alvana Journal of Science vol 1, No 5.

Sadier, P. M. (1998), Psychometric models of student conceptions in science. Journal of research in science teaching vol 35, No 3pps 265-296.

Stephen, M. (2007), Multiple choice test fair sage, mastering publisher.

Toube (2009), Assessing students critical thinking performance. Journal of thinking skills & creativity vol4, No1.2