

# THE RELATIVE EFFECTIVENESS OF THREE TEST FORMATS IN EVALUATING STUDENTS' PERFORMANCE IN BIOLOGY

*Gbore, L. O.*

## **Abstract**

This study examines the relative effectiveness of matching test, completion test and multiple choice items format on students' performance in Biology achievement test. A sample of one hundred SSSIII students were randomly selected from ten senior secondary schools in Akure South Local Government Area of Ondo State. The three postulated hypotheses were tested using ANOVA at 0.05 level of significance. The results showed that: (1) there was no significant difference in the mean scores of students in the matching test, completion test and multiple choice test items formats; (2) there was no significant difference in the performance of boys and girls in the three Biology test formats. Hence, it was recommended that classroom teachers and examining bodies should also regularly use matching and completion test items to complement the commonly used multiple-choice test items or use a combination of the three test formats for effective evaluation in Biology.

## **Introduction**

When questions are prepared and given to students to give response, then a test is said to have been conducted. Bandele (1992) said that test is an instrument given to measure a specific type of behavioural change. It is true that tests have been the oldest and still the dominant instrument used by teachers and examining bodies for measuring and assessing students' academic performance in the various teaching subjects in Nigeria. Decisions of today are made about persons in and out of school and test of individual's achievement guides the school in providing very vital and useful information concerning such decision making. A student preparing to pursue medicine as a course, would among other credits passes be expected to possess a credit pass in Biology, and today, the O'Level Biology syllabus has become very voluminous, and students broad field of competency in understanding the stipulated content of the syllabus can be conveniently evaluated using matching, completion, and multiple choice tests formats bearing in mind the satisfaction of the six levels of Bloom's (1956) taxonomy of educational objectives. Therefore, any steps taken to improve the reliability of the testing instruments employed by teachers will be a worthy exercise. The usual and commonest test among the test instruments used for administration purpose in schools, placement purpose to schools and for classroom teacher-students interaction are the essay type of tests and the multiple-choice test items while other objective type of test items such as completion and matching types have not been utilized despite their merits. Advocates of essay tests have contended that it allows students to employ higher mental processes, organize materials logically and concisely and to be creative. Researchers such as Damrin (1969) have indicated the supremacy of essay type of test over the objective type of tests. However, Itsuokor (1988) putting forward three problems of essay type of test claimed that it lacks: content validity, scorer reliability and scoring economy. Oladunni (1996) also claimed that essay test is too subjective and do not always sample half of the content area demonstrated by an average pupil in the objective tests. Borisade (1997) noted that in the case of objective test, there is only one predetermined correct key and hence subjectivity in marking is eliminated. Also, he found out that there was a significant difference between the performance of boys and girls in both matching and completion test formats. Herrmann (1992) quoting Deaux and Emswiller (1974) noted that success or failure on a task were related to both sex of the subject and the gender-typing of the task. Papalia and Olds (1958) claimed that in most cognitive areas like general intelligence, learning reasoning tasks, problem solving and creativity, there is no difference between males and females. Atkinson, Smith and Ben (1990) pointed out that the sex difference in cognitive abilities which had been observed since almost the beginning of systematic testing appears to be vanishing.

The new National Policy on Education brought with it as one of its features "continuous assessment". The term "continuous assessment", according to Alonge (1989) is the systematic or regular testing of students between terms, sessions, or throughout the duration of the course of study for the purpose of measuring or assessing the students' progress or lack of it on a continual basis.

*Gbore, L.O.*

Considering the content of Alonge's (1989) definition of continuous assessment and its demands on teachers, coupled with the present day large number of students in classes in schools, many teachers may not have much time and endurance to mark voluminous written materials produced as response to questions by students. Nwana (1979) claimed that in this type of situation, the objective test items become a useful and easy assessment technique. Today, there is need to assess the type of evaluation instruments that classroom teachers use for measuring achievements among the students if only to reduce the cumbersomness of demands of continuous assessment in the schools.

Karmel and Karmel (1978) claimed that the most important variable in judging the adequacy of a test is its validity. Nnamdi (1991) explained that validity is superior to reliability, but both are very useful in the determination of the usefulness of a measuring instrument. Therefore, to improve the effectiveness of the teaching-learning processes and evaluation of students' achievements in classroom, the use of moderately difficult, discriminating, valid and reliable instruments is what is required and it is better achieved by means of objective test items.

Ebel (1972), Ahmann and Glock (1981) both agreed that objective test items may be designated as 'supply' types or 'selection' types of items. The multiple choice test item has been identified and considered the most popular and commonest of all the objective types of tests. Bracht and Hopkins (1970) and Coffman (1971) have revealed that if multiple choice type of objective test items are carefully developed, it will bring better valid and reliable evaluation of instructional objectives in varied teaching areas. Denga (1987) and Oladunni (1996) are of the opinion that multiple choice type of objective test is useful for measuring both lower and higher order mental processes.

Other selection types of item include True-False item which Oladunni (1996) claimed to be widely used by classroom teachers because they can and are used to writing mediocre items which they can score rapidly with ease. Although, this type of test format provides room for sampling large content area of subject matter in a test, it is limited by its 50% susceptibility to guessing. Of all the types of objective test formats, Ebel (1979) has found out that completion and matching test item formats are adaptable to the measurement of most important educational outcomes such as knowledge, understanding, evaluation or ability to solve problems. The construction of completion items and matching test items is relatively easy for a skilful tester and both of these test formats make evaluation of a large amount of related materials in a relatively short time possible. It is, therefore necessary that a thorough and careful assessment of some of these forms of objective tests employed in the evaluation of achievement of students in the various teaching subjects, especially in Biology, in the classroom be carried out.

### **Statement of the Problem**

Testing is an aspect of teaching-learning process through which the measurement of the development and achievement of students in schools is determined. Most test items given to the students as diagnostic, formative and summative tests are solely and poorly constructed by the classroom teachers. Alonge (1989) claimed that most teachers suddenly realize that it is time to give test, sit down to write items and never plan for it. Most teachers choose and use just any test format without careful and thoughtful consideration over these varied test formats and their merits.

This study is therefore out to construct three biology test formats (completion, matching and multiple choice types of tests) with the purpose of determining whether there is difference in effectiveness in determining the performance of students in subjects like Biology.

### **Research Hypotheses**

To guide this study, the following Null hypotheses were postulated for verification and testing at 0.05 level of significance:

#### **Hypothesis One**

There is no significant difference in the mean scores of students in biology completion test items, matching test items and multiple choice test items formats.

#### **Hypothesis Two**

There is no significant difference in the performance of boys in biology completion test items, matching test items and multiple choice test items formats.

## *The Relative Effectiveness of Three Test Formats in Evaluating Students' Performance in Biology*

### **Hypothesis Three**

There is no significant difference in the performance of girls in biology completion test items, matching test items and multiple choice test items formats.

### **Methodology**

The population for this study consists of all the year two Senior Secondary School students (SSII) in Akure South Local Government Area of Ondo State with a total number of 19 secondary schools. The stratified random sampling technique was employed to select one hundred SSSII students from ten secondary schools randomly selected using simple random sampling.

The instrument used for the collection of the data are three objective test formats made up of 30 completion test items, 30 matching test items and 30 multiple choice test items. The researcher drew the items from the Ondo State joint biology scheme and record of work of senior secondary school two. Initially, 50 items on each of the test formats were constructed using table of specification or test blueprint. These items were then try out on a sample of 20 non-participating SSII students from Owo Local Government Area of Ondo State. The items were then subjected to item analysis and 30 items on each of the test formats were finally selected based on items falling between 0.30-0.80 p- value (difficulty level) and 0.35-0.65 d-value (discriminating index).

The Kuder-Richardson formula-twenty one ( $KR_{21}$ ) was used to determine the reliability coefficient of each of the instrument. The coefficients were 0.67 for the completion test items format, 0.78 for the matching items format and 0.70 for the multiple choice test format.

### **Procedure**

Each of the one hundred subjects were given copies of the biology completion test format, matching test format and the multiple-choice format in a successional manner for them to respond to for twenty five minutes for each of the test formats. The researcher served as invigilator for the tests in each of the participating school to avert examination malpractices. For the purpose of data analysis, the scripts of the participants were collected and scored at the end of the testing session. The range of the scores of the students in the matching test format fell between 12-27, completion test format fell between 10-28 while multiple choice format test fell between 13-27.

### **Data Analysis**

The three hypotheses were analysed and tested using one way Analysis of Variance (ANOVA) at 0.05 level of significance.

### **Testing of Hypotheses Hypothesis One**

This hypothesis states that there is no significant difference in the mean scores of students in

**Table 1: One Way Analysis of Variance (ANOVA) of the Mean Scores of Students in**

Source of Variance	Df	SS	MS	F
Between Group	2	107	53.5	1.55
Within Group	297	10221.197	34.41	
TOTAL	299	10328.197		

biology completion test items, matching test items and multiple choice test items formats.

### **Completion Test, Matching Test and Multiple-Choice Test Formats.**

The result presented in Table 1 shows that the calculated value of F is 1.55 which is less than the critical value of 3.00 at degree of freedom of 2 and 297 with 0.05 level of significance hence this

$P < 0.05$  (Result not Significant).

hypothesis is upheld. In other words, there is no significant difference in the mean scores of the students in the three test formats.

Gbore, L.O.

### Hypothesis Two

This hypothesis states that there is no significant difference in the performance of boys in biology completion test items, matching test items and multiple choice test items format.

This hypothesis was tested using one way Analysis of Variance (ANOVA) at 0.05 level of significance.

**Table 2: One Way Analysis of Variance (ANOVA) of the Performance of Boys in Biology**

Source of Variance	Df	SS	MS	F
Between Group	2	76.44	38.22	0.92
Within Group	147	6119.72	41.63	
TOTAL	149	6196.16		

$P < 0.05$  (Result not Significant).

Table 2 revealed that the F calculated was 0.92 while the critical value was 3.00 at degree of freedom 2 and 147 with the level of significance at 0.05. Since the calculated F value of 0.92 is less than the critical value of 3.00, the hypothesis which states that there is no significant difference in the performance of boys in biology completion test items, matching test items and multiple choice test items format is upheld.

### Hypothesis Three

Hypothesis three states that there is no significant difference in the performance of girls in biology completion test items, matching test items and multiple choice test items formats. In testing this

**Table 3: One Way Analysis of Variance (ANOVA) of the Performance of Girls in Biology**

Source of Variance	Df	SS	MS	F
Between Group	2	36.05	18.03	0.59
Within Group	147	4498.78	30.60	
TOTAL	149	5434.83		

hypothesis, the one way Analysis of Variance was employed at 0.05 level of significance.

### Completion Test, Matching Test and Multiple-Choice Test Formats

The result of the Analysis of Variance in Table 3 revealed that F-calculated value was 0.59 while the critical value was 3.00. The F-calculated value of 0.59 fell within the critical region. This

$P < 0.05$  (Result not Significant).

means that there is no significant difference in the performance of girls in the three test formats.

### Conclusion

The findings of this study show that the three Null hypotheses raised were accepted, indicating that there is no difference in the performance of students in the three test formats. This result also followed the same outcome or finding of the study of Coffman (1971) who claimed that after the correction for reliability of multiple-choice test items format and completion test items format, performance on the two test formats over the same content measures essentially the same factors. It is therefore concluded that there is no marked difference in the relative effectiveness of matching test format, completion test format and multiple choice test format in serving as evaluation instrument for Biology. In other words, completion test format and matching test format are as effective as the popular multiple choice test format in evaluating student's achievement.

### Recommendations

Finally, it is recommended that classroom teachers and examining bodies should ensure that

### ***The Relative Effectiveness of Three Test Formats in Evaluating Students' Performance in Biology***

they use the three varieties of the test formats regularly in their tests to make objective test items function efficiently as measures of both higher and lower mental abilities, and more so to improve the level of performance of the students. The three test formats can be combined and employed in a single test for evaluating performance in Biology.

#### **References**

- Ahman, J.A. and Glock, M.D. (1981) *Evaluating Students Progress: Principles of Test and Measurement*. Sixth Edition Boston: Allyn and Bacon Inc.
- Alonge, M.F. (1989): *Measurement and Evaluation in Education and Psychology*. Ado-Ekiti: Adebayo Printing (Nig.) Ltd.
- Bandeke, S.O. (1992) *Principles of Test Construction* (Mimeograph) Ondo State University, Ado- Ekiti.
- Bloom, B.S. (1956): *Taxonomy of Educational Objectives Handbook 1. The Cognitive Domain*. New York: David McKay Co., Inc.
- Borisade, T.F. (1997) Comparative Analysis of the Difficulty and Discriminating Indices of Two Mathematics Test Formats. *Unpublished M.Ed. Thesis*. Ondo State University, Ado-Ekiti.
- Bracht, G.H. and Hopkins, K.D. (1970) The Comparison of Essay and Objective Tests of Academic Achievement. *Educational and Psychological Measurement*. 30: 259-364.
- Coffman, W.E. (1971) Essay Examination in Educational Measurement. *American Council on Education*.
- Damrin, L.B. (1960) Can Teachers Write Good Completion Test Items? *School Review*, 11: 320-330.
- Denga, D.I. (1987) *Educational Measurement, Continuous Assessment and Psychological Testing*. Calabar: Rapid Educational Publishers (Nig.) Ltd.
- Ebel, R.L. (1972) *Essentials of Education Measurement*, Englewood Cliff, N.J: Prentice Hall Inc.
- Herrmann, D.J. (1992): Gender-Linked Differences in Everyday Memory Performance. *British Journal of Psychology* 83: (3), 221-231.
- Itsuokor, D.E. (1988) *Essentials of Test and Measurement*. Ilorin: Woye and Sons (Nig.) Limited.
- Karnel, L.J. and Karnel, M.O. (1978) *Measurement and Evaluation in the Schools*, 2nd ed. New York: Macmillan Publishing Co. Inc.
- Nnamdi, A. (1991) *Research Methodology in the Behavioural Sciences*. Lagos: Macmillan and Co. (Nig.) Limited.
- Nwana, O.C. (1979) *Education Measurement for Teachers*. Ikeja: Thomas Nelson (Nig.) Ltd.
- Oladunni, M.O. (1996) *Test, Measurement and Evaluation in Education*. Ibadan Nigeria: Demilade Omotayo Publishers.