

# AN OVERVIEW OF TEST SCORING AND INTERPRETATION OF TEST SCORES IN TEACHING AND LEARNING IN NIGERIAN SCHOOLS

*Esther Amina Akuh*

## **Abstract**

Tests are used to determine what students have learnt or not learnt in the class. Tests show the aspects of the course or lesson that the students have learnt, They also show areas where learning has not taken place. Thus, the teacher can re-teach for more effective learning. Similarly, this paper has discussed what test is, why test is important, how tests are constructed and what precautions are taken to ensure validity of tests. The paper also concludes by explaining how tests are scored and interpreted to improve teaching and learning in our schools,

## **Introduction**

Tests are detailed or small scale task carried out to- identify the candidate's level of performance and to find out how far the person has learnt what was taught or be able to do what he/she is expected to do after teaching (Ogunniyi, 1994). Tests are carried out in order to measure the efforts of the candidate and characterise the performance. Test is therefore an instrument for assessment. Assessment is broader than tests, although the term is sometimes used to mean tests as an "i want to assess your performance in the course"<sup>1</sup>. Some even say they want to assess students' scripts when they really mean they want to mark the scripts.

Tests according to Carroll (1993) are conducted to find out whether the objectives set for a particular course, lesson or topic have been achieved or not. Tests measure the performance of a candidate in a course, lesson, or topic and thus, tell the teacher or course developer that the objectives of the course or lesson have been achieved or not. If the person taught performed badly, we may have to take a second look at the objectives of the course or lesson.

Tests are also used to determine the progress made by the students. Tests are used to place students/candidates into a particular class, school, level or employment. Such tests are called placement tests. The assumption here according to Mitchell (1992) is that an individual who performs creditably well at a level can be moved to another level after testing. Thus, we use test to place a pupil into primary two, after he/she has passed the test set for primary one. and so on.

Suggestively, we can use test to know whether or not the students are improving in the course, lesson, or topic. If progress is made, we reinforce the progress so that the students can learn more. If no progress is made, we intensify teaching to achieve progress. If progress is slow, we slow down the speed of our teaching. Tests can reveal the problems or difficult areas of a learner. Thus, we say we use test to diagnose or find out the problems or difficult areas of a student or pupil. A test may reveal whether or not a learner, for example, has a problem with pronouncing a sound, solving a problem involving decimal, or constructing a basic shape, e.g. a triangle, rectangle, etc.

Tests are used to predict outcomes. We use tests to predict whether or not a learner will be able to do a certain job, task, use language to study in a university or perform well in a particular school, college or university. We assume that if Usman can pass this test or examination, he will be able to go to 100 level of a university and study soil science or crop science. This may not always be the case, though there are other factors that can make a student do well other than high performance in a test (Usman, 2008).

## **Scoring of Tests**

Multiple choice tests are difficult to design, difficult to administer, especially in a large class, but easy to score. In some cases, they are scored by machines. The reasons for easy scorability of multiple-choice tests are because they usually have one correct answer, which must be accepted across the board. According to Lado (1991), essay or subject types of tests are relatively easy to set and administer, especially in a large class. They are, however, difficult to mark or assess. The reason is because essay questions require a lot of writing of sentences and paragraphs. The examiner must read all these.

Whether an objective or subjective test, all tests must have marking schemes. Marking schemes are the guide for making any tests. They consist of the points, demands and issues that must be raised before the candidate can be said to have responded satisfactorily to the test. Marking schemes according to Lado (1991) should carry mark allocation. They should also indicate scoring points and how the scores are totaled up to represent the total score for the question of the test.

Scoring or marking on impression is dangerous. Some students are very good at impressing examiners with flowering language without real academic substance. If you mark on, impression, you may be

carried away by the language and not the relevant facts. Again, Hyghes and Porter (1993) explained that mood may change impression, your impression can be changed by joy, sadness, tiredness, time of the day and so on. Probably, that is why you must always 'insist' on a comprehensive marking scheme.

Furthermore., scoring can be done question-by-question or all questions at a time. The best way is to score or mark one question across the board for all students (Okpalla, 1999). Sometimes, this may be feasible and tedious, especially in a large class. Scores can be interpreted into grades, A, B, C, D, E and F. They may be interpreted in terms of percentages: 10%, 20%, 50%, 60%, etc. Scores may be presented in a comparative way in terms of 1<sup>st</sup> position, 2<sup>nd</sup> position and 3<sup>rd</sup> position to the last. Scores can be coded in what is called BAND (Okpalla, 1999). In band system, certain criteria are used to determine those who will be in Excellent, Very Good categories, etc. An example of a band system is the one given by the International English Testing Services (IETS) and the one by Teaching English as a Foreign Language (TEFL) test.

### **Objective Scoring: Correction Formula**

As stated earlier in the paper, objective (est is very easy to score. All other advantages of objective test are well known to all teachers. However, the chances of guessing the correct answer are high (Ogunniyi, 1994). To discourage guessing, some objective tests give instructions (o candidates that they may be penalised for guessing. In such situation, the correction formula is applied after scoring. This is given as:

$$\begin{aligned} & \text{No. of questions marked right (R)} \\ & - \text{No. of questions marked wrong (W)} \\ & \text{No. of options per item (N) - 1} \end{aligned}$$

If in an objective test of 50 questions where guessing is prohibited, a candidate attempted all questions and gets 40 of them correctly, then the actual score after correction is  $S = 40$  (assuming the options per item is 5)  $S = 40 - 2.5 = 37.5 = 38$  out of 50.

### **Using Test Results**

Test is not an end in itself. However, before tests could be used for those purposes, the teacher needs to know how well designed the test is in terms of difficulty level and discrimination power, then he should be able to compare a child's performance with those of his peers in the class. Occasionally, he may like to compare the child's performance in one subject area with another. To do this, he carries out the following activities at various times (Rechard and Cook, 1999).

- i. Item analysis
- ii. Drawing of frequency distribution tables.
- iii. Finding measures of central tendency (mean, mode and median),
- iv. Finding measures of variability and derived scores.
- v. Assigning grades.

#### **Difficulty Level**

This is the number of candidates that got a particular item right in any given test (Davies, 1990). For example, if in a class of 45 students, 30 of the students got a question correctly, then the difficulty level is 67% or 0.67. The proportion usually ranges from 0 to 1 or 0 to 100%. Usually, the formula for difficulty is

$$F = \frac{n}{N} \times 100$$

Where  $p$  = item difficult

$$\begin{aligned} n &= \text{the no. of students who got the item correct} \\ N &= \text{the no. of students involved in the test} \end{aligned}$$

However, in the classroom setting, it is better to use the upper 1/3 of the students that got the item right (U) and the lower 1/3 of the students that got it right (L). Hence difficulty level is given by

$$\frac{U + L}{N}$$

Where N is the number of students actually involved in the item analysis (upper 1/3 + lower 1/3 of the tests).

#### **Item Discrimination**

The discrimination index shows how a test item discriminates between the bright and the dull students (Oiler, 1997). A test with many poor questions will give a false impression of the learning situation. Usually, a discrimination index of 0.4 and above are acceptable. Items which discriminate negatively are bad. This according to Oiler (1997) may be as a result of wrong keys, vagueness or extreme difficulty. The formula for discrimination index is:

$$\frac{U - L}{\frac{1}{2}N} \quad \text{or} \quad \frac{U - L}{0.5 N}$$

where  $U =$  the number of students that got it right in upper group  
 $L =$  the number of students that got it right in lower group  $N =$   
the number of students usually involved in the item analysis.

### Derived Scores

Usually, we report on our students after examinations by adding together their scores in the various subjects and thereafter calculate the average or percentage as the case may be. This according to Davies (1990) does not give a fair and reliable assessment. Instead of using raw scores, it is better to use derived scores. A derived score usually expresses every raw score in terms of other raw score on the test. The commonly used ones in the classroom are the Z-scores, T-scores and percentiles.

### Z-Score or Standard Score

This is the deviation of the raw score from the mean divided by the standard deviation.

$$Z = \frac{X - \bar{X}}{SD}$$

Where  $Z =$  Z-score

$X =$  any raw score  $\bar{X} =$  the mean  
 $SD =$  Standard Deviation

Raw scores above the mean usually have positive Z-scores while those below the mean have negative Z-scores. Z-scores can be used to compare a child's performance with his peers in a test or his performance in one subject with another (Oiler, 1997).

### T-Score

This is another derived score often used in conjunction with the Z-score. It is defined by the equation

$T = 50 + 10Z$  ' It is also used in the same way as the Z-score except that the negative signs are eliminated in T-scores.

### Percentiles

This express a given score in terms of the percentage scores below it. For example in a class of 30, Usman scored 60 and there are 24 students scoring below him. The percentage of score below 60 is therefore:

$$\frac{24}{30} \times 100 = 80\%$$

Usman therefore has a percentile of 80 written as P80. This means Usman surpassed 80% of his colleagues while 20% were better than him. The formula for the percentile rank is:

$$PR = 100 \times \frac{b + F}{N}$$

Where PR = Percentile rank of a given score

$b =$  Number of scores below the score

$F =$  Frequency of the score

$N =$  Number of all scores in the test

## Course Credit System and Grade Points Interpretation

Perhaps the most precious and valuable records after evaluation are the marked scripts and the transcripts of a student. At the end of every examination e.g. semester examination, the marked scripts are submitted through the head of department or faculty to the examination officer. Occasionally, the examination officer can round off the marks carrying decimal, either up or down depending on whether or not the decimal number is greater or less than 0.5.

The marks so received are thereafter translated/interpreted using the Grade Point (GP), Weighted Grade Point (WGP), Grade Point Average (GPA) or Cumulative Grade Point Average (CGPA). Courses are often weighed according to their credit units in the course credit system.

Grade Point (GP):- This is a point system, which has replaced the A to F Grading System.

Weighted Grade Point (WGP):- This is the product of the grade point and the number of credit units carried by the course i.e.  $WGP = GP \times \text{No. of credit unit}$ .

Grade Point Average (GPA):- This is obtained by multiplying the grade point attained in each course by the number of credit units assigned to that course, and then summing these up and dividing by the total number of credit units taken for that semester (total registered for).

$$GPA = \frac{\text{Total Point Scored}}{\text{Total Credit Units registered}}$$

$$\frac{\text{Total WGP}}{\text{Total credit units registered}}$$

Cumulative Grade Point Average:- This the up to date mean of the grade points earned by the student. It shows the student's overall performance at any point in the programme (Davies, 1990).

$$CGPA = \frac{\text{Total Points so far Scored}}{\text{Total Credit Units so far taken or registered}}$$

The scores and their letter grading and interpretation may vary from programme to programme or institution to institution.

## Recommendations and Conclusion

This paper discusses the pattern of scoring of tests and interpretation of scores be the continuous assessment tests or other forms of tests. The following guidelines are recommended for scoring and interpreting test scores for effective teaching and learning.

1. Scoring or making and interpretation should not be done on impression.
2. The best way is to score or mark one question across the board for all students.
3. Whether an objective or subjective tests, all tests must have marking schemes for easy scoring and interpretation.
4. The test scores and interpretation should be able to reveal whether the objectives set for a particular course, lesson or topic have been achieved or not.
5. Scores can be interpreted into grades, A, B, C, D, and F. They may be interpreted in terms of percentages: 10%, 20%, 50%, etc.

Test scoring and interpretation of test score is very important in teaching and learning as it reveals the problems or difficulty areas of a learner. Thus, we use tests to diagnose or find out the problems or difficult areas of a student or pupil. Test scores can be used to predict outcomes and is valid if it measures what it is supposed to measure and reliable if it measures what it is supposed to measure consistently.

## References

Carroll, J.B. (1993). *Psychometric theory and language testing*. Newbury: Rewley Mass Press.

Davies, A. (1990). Validating three tests of English language proficiency: *Language Testing* (1) 50-60.

Hughes, A. and Porter, D. (1993). *Current developments in language Testing*. London: Academic Press.

- Lado, R. (1991). *Language testing: The construction and use of foreign language tests*. London: Longman.
- Mitchel, R.J. (1992). *Measurement in the classroom: A Worktest*. Hunt Publishing Co.
- Ogunniyi, M.B. (1994). *Educational measurement and evaluation*, Longman.
- Okpalla, M.P. (1991). *Measurement and evaluation in education*. Ibadan: Stiching-Horden Publishers.
- Oiler, J.W. (f 997). *Language tests at school*. London: Longman.
- Richard, C. and Cook, T. (1999). *Qualitative and quantitative methods in evaluation research*. California: Sage.
- Usman, M. (2008). A guideline for setting standards of performance on educational occupational Test in agriculture. A seminar paper presented at the First ASSORTVEN Seminar, Potiskum/Gombe North East Chapter. Dec. 4<sup>th</sup>.