CRITERION RELATED VALIDITY AS A CORRELATE OF STUDENTS' RELIABILITY IN BIOLOGY ESSAY TESTS.

CASMIR N. EBUOH Ph.D.

DEPARTMENT OF SCIENCE AND COMPUTER EDUCATION
ENUGU STATE UNIVERSITY OF SCIENCE AND TECHNOLOGY, ENUGU-NIGERIA
E-mail: ebuohcas@yahoo.com

Abstract

Literature showed that there are low enrolments of students into Biology related courses such as medicine, pharmacy, radiography, nursing, dentistry and so on as result of students' poor performances in biology and inadequate use of method of evaluation. The purpose of the study was to investigate into the criterion related validity as a correlation of student's reliability in biology essay test scores in secondary schools. The design of the study was correlation and the population was all the 1750 biology students of the senior secondary school two (SS 2) in the 2015/2016 academic session. Simple random sampling (battling without replacement) was used to sample 450 students from 14 secondary schools in Ezeagu Local Government area of Enugu Slate. The instrument for data collection was biology essay test (BET). Pearson 'r' and t-test were used to correlate and test the hypotheses at 5% level of confidence. Based on the analyzed data, it was found out that the correlation coefficient value was both positive and high for the concurrent and predictive validity in essay test scores. The reliability was significant for predictive and current validity in Biology essay tests. Some recommendations were made based on the findings of the study.
Keywords: Criterion related validity, correlate, biology, essay tests

Introduction

Evaluation is a process of assessing the effectiveness of a programme of study in order to bring about desired behavioural changes in the learner. These processes take into account all the skills, attitudes, abilities, behavioural changes and knowledge in subject acquired by the students in a particular programme. The reasons for evaluation could be to judge the level of achievement. Teachers need to know the successes of their students and in doing so, weakness in learning is revealed. To enable students realize their changes and how they improve by their efforts (Ebuoh, 2014).

For evaluation to be effective, the teacher must be responsible for the success of his students' learning and to be ready to receive feedbacks from them in order to enable the teachers to be sensitive to the reliability of instrument (tests) for data collection. Evaluation techniques include theory (objectives and essay) practical, projects, questioning, marks and corrections. These different methods of assessment or evaluation are useful in the sciences in general and Biology in particular. It has been observed that the type of examination that takes place in our schools strongly influence the type of study procedure use by students preparing for them (Maduabum, 2004).

Validity of evaluation instruments/tests are the extent to which the tests measure what they purport to measure. That is, it is the degree to which a test measures what it is intended to measure. Validity of

testing instruments is also seen as the truthfulness of the instrument. Consider for instance, if a form is given to a student requiring him to state his age and the student responds by stating his age as 16 years and later was confirmed by his birth certificate that his age is 16 years then the test item in the form is valid. It is valid because it elicits the expected and true response. But if the student's response is 15 years then the test item is not valid.

Validity however subsumes reliability. This is because a test item adequately valid for predictive purposes is sure to be adequately reliable. Since a single test may be used for many different purposes, there is no single validity index for a test. A test that has some validity for one purpose may be invalid for another. There are three major types of validity, namely, the content, construct and criterion related validity. More so, there is another type of validity, usually, referred to as "face" validity which is actually not considered as a type of validity in the strict or technical sense.

It pertains to aptitude test used for predicting a student's future performance. It is for example, a test used to admit students into various programmes in institutions. The validity of the aptitude test is therefore its ability to discriminate between those students who have basic abilities required for success in a specific type of programme and those who lack such abilities required for success in a specific type of programme.

Criterion-related validity is further sub-classified according to the time interval between the test (concurrent and predictive) measures (Ebuoh, 2004).

(1) Concurrent Validity

Where the criterion is obtained about the same time as the test score, that is concurrently, it is then referred to as concurrent validity. This type of validity is important for tests used in measuring existing status. Normally such tests provide quicker, more objective and easier means of assessing the behaviours than the previous methods. Otherwise, there would not be the need for such tests since the criterion measure is usually available during the time of validating the test. Consider for instance, after several observations, a teacher was able to classify the members of his class according to their study habits. He later developed a test of study habits. This test can now be validated against his former grouping of his pupils

(II) predictive

In a situation where the criterion is obtained at a future date, this is referred to as predictive validity. It is concerned with how well a test can predict subsequent behavior. For example, a teacher may give a class one student a test of intelligence which he would like to see how well it can predict performance at Junior Secondary School Certificate Examination (JSSCE). After three years period, he obtains his

score on Junior Secondary School (J.S.S) examination. This now serves as the criterion. When the scores on intelligence test and criterion measure (JSSCE) are correlated, you obtain a co-efficient of predictive validity for the test.

However, this form of validity is very important to all tests used in selecting candidates for a particular purpose e.g. Common Entrance Examination, Joint Admission and Matriculation Board, Civil Service selection examination. This type of validity is pertinent in the case of intelligence tests or any other tests designed to predict subsequent behaviours.

Experience has shown that the use of inadequate evaluation techniques in the determination of the students performance in Biology, particularly has ultimately denied the students the feeling of participation and reality. What makes a Biologist is how much information he/she has stored in his memory and the actual practice he receives in biological rigorous process, how he wonders, sets up a controlled experiment, his willingness to withhold judgments and how he realized the limitations of Biology. These are affected by the use of essay tests in evaluating Biology.

The importance of essay tests in the evaluation of Biology cannot be over-emphasized, because the essay tests provide opportunity to promote the scientific method of thought. The scientific method of thought entails inculcating into the learners, the habit of drawing conclusions based on observation, experimentation and practical. The use of essay tests can also enable students to learn much about the interrelationship between biology and other science subjects.

The evaluation techniques (objective tests) employed in the evaluation of Biology are designed ultimately to produce educated individuals. Some of whom may or may not take to biological studies in their professional pursuits. However, in whatever profession they finally find themselves, it is hoped that the Biology education they have acquired in school will be of value to the totality of their education.

Correlation study is highly useful in studying problems in education or in other behavioural sciences. This permits one to measure a great number of variables and their interrelationship simultaneously. In behavioral science, we are frequently confronted with situation in which a large number of variables are contributory causes of a particular pattern of behaviour. The classical experimental method which manipulates one variable and attempts to hold others constant often introduces a high level of artificiality into research situation encountered in the behavioural science.

This section discusses the essay types as acclaimed method of evaluating students' performance in Biology. Omelewa (2007) noted that it is needless to re-emphasis the fact that Biology by its nature and characteristics is an intellectual subject. The less it is evaluated otherwise the more it ceases to be Biology that is examined. The essay tests are useful in evaluating and developing students' intellectual abilities in biology. Essay test types are more successful in evaluating higher order level of cognitive domain in Biology. Essay test types appear to offer means of development of students' performance in the learning of

biological concepts, principles and methods. Okeke (2005) noted that the essay test types are indispensable tools for the evaluation of the affective, psychomotor and cognitive domains of students' performances. It was observed that essay test types are more suitable to brighter and average students when compared with objective test items. Ekwuzie (2005) stated that essay test type is the best because it creates fertile ground for the examination of the objective of the topic and because it sometimes discovering new terms. Therefore, the opportunity of making students the researchers or discovers full of curiosity, interest that were desired at formative stage of appreciation and seeing Biology as their best alternative to other subjects. Essay test types need to be dovetailed into practical type of assessment so that the entire content is a unique whole and evaluated in a manner conducive to thinking. Other methods of evaluating Biology (such as observation, questioning, projects, corrections and marks) lack the necessary intellectual development and creative thinking which are found in essay test types. Ibekwe (1995) while comparing two evaluative methods in the evaluation of integrated science found out that essay method is more appropriate in evaluating integrated science (Biology chemistry and physics). Evaluation should be made more reliable so that examinees can be more actively involved and makes for recall of facts. Biology cannot be evaluated from but through essay and written expressions. Ebuoh (2014) and Udeigwe (2011) found out that practical and essay types are the best methods of evaluating Biology and indeed other sciences. The test format allow students not to guess, put down, experience, observe, experiment and draw conclusions so that they can perceive Biology as it is. Maduabum (2004) claimed the superiority of essay from the advantages they have over other evaluative techniques.

Essay test is classified into two namely,

i. Restricted response/short answer type
ii. Extended response/long essay type.

In this response type the respondent is required to provide written answers of a few times in length (specified numbers) to brief questions. Example,

1. List five major social functions of education
2. Write eight sentences on any two of the following: Reproduction, growth, movement and irritability.

Here, the respondent is required to provide a long comprehensive written answer of three or more pages to a question. Consider for instance, discuss the common dimensions for classifying test.

In either case, the respondent is faced with the problem of providing the answer unaided and in whatever manner he may deem appropriate. Thus he is free to choose his own words and organize his ideas in his own way bearing in mind the relevance of these to the questions asked. The restricted and the extended types are not applied always and have advantages over each other as well as disadvantages.

1. They help to appraise higher level of mental ability (higher order domains) as reasoning abstract thinking, imagination, induction and deduction.
2. Guessing is minimized
3. They encouraged organization, integration and expression of ideas effectively

There are several disadvantages of essay test such as:-

1. Scoring is slow, subjective and inconsistent which makes marking low reliability.
2. It is also subjective and inconsistent which makes marking low reliability.
3. Sampling of course contents is usually limited due to the small number of questions that can be included in a test.

They also make the phenomena being evaluated more real through thinking and actual experience. Having examined the rationale for the use of co- relational study and essay tests, it is pertinent to carry out a study on the criterion-related validity as a correlate of student's reliability in biology essay tests in secondary school Biology.

Okeke (2003) found out that the method of evaluating biology in particular and sciences in general affect reliability of evaluation instrument. In recent years, there were increases in the poor performance of students in senior secondary certificate biology. Similarly Ebuoh (2004) found out that this low reliability was related to the use of the inadequate method of evaluating Biology in secondary schools. The researcher therefore, wishes to investigate into the criterion-related validity as a correlate of student's reliability in biology essay tests in secondary schools.

The purpose of the study was to find out:

1. the relationship between concurrent validity on students' reliability in biology essay tests.
2. the relationship between predictive validity on students' reliability in biology essay tests.

This study is restricted to senior secondary two (SS 11) biology students in Ezeagu local Government Area of Enugu State. It concerns itself in finding the relationship the criterion related validity as a correlate of students' reliability in biology essay tests.

Research questions

1. What is the relationship between concurrent validity as a correlate of student's reliability in biology essay tests?
2. What is the relationship between predictive validity as a correlate of student's reliability in biology essay tests?

Hypotheses:

The hypothesis was tested at five percent (5%) level of significance that is, ninety five percent (95%) chances of being correct if rejected or failed to be accepted.

1. There is no significant relationship between concurrent validity as a correlate of student's reliability in biology essay tests.

2. There is no significant relationship between the predictive validity as a correlate of student's reliability in biology essay tests.

Research Methods

This study is a co relational study. It was aimed at finding out the relationship between criterion-related validity on student's reliability in biology essay tests.

The population for the study comprised all the 1750 senior secondary two (SS 2) Biology students in the 10 secondary schools in the Ezeagu Local Government Area in 2015/2016 session. The study did not use all the 1750 Biology students in the 17 secondary schools in the Local Government Area, 10 schools were randomly sampled using simple balloting without replacement. Simple balloting without replacement and the sample size was 450 senior secondary school (SS3) Biology students that wrote the biology essay tests (BET).

Validity of the instrument was determined using face validity. The instrument was validated by three experts. One is an expert in measurement and evaluation and other two are experts in Biology Education. Reliability was found using Cronbach alpha after trial testing the instrument and the value was .87

In method of data collection, the scores from the 450 senior secondary two (SS 2) Biology students of the 10 sampled schools were extracted from their scripts after the administration and scoring of BET.

In order to analyse the data, various tools were employed. Firstly, the Pearson's product moment correlation co-efficient was used to analyse the interval data and to find the relationship between the two administrations of the tests. The calculated Pearson's 'r' was tested for significance with t-test at five percent confidence level of significance. That is, ninety five percent (95%) chances of being correct if rejected or failed to be accepted. The correlation coefficient were given qualitative interpretation based on the table one below

Table 1: Quantitative interpretation of co -relation co-efficient.

Correlation coefficient value	Interpretation
0.0-0.20	Very low/virtually no relationship
0.20-0.40	Low/ definite positive relationship
0.40 - 0.60	Medium relationship
0.60-0.80	High relationship
0.80-1.00	Very high/ near perfect relationship

Source: Nworgu(1992).

Results

The result of the research question is presented in table 2 below.

It was shown from table 2 below that the relationship between concurrent validity on student's reliability in biology essay tests was positive and high relationship. There was high relationship between predictive validity on student's reliability in essay tests. In schools where the relationship was positive, the magnitude was between 0.76 and 0.82 which showed a range of high relationship the two group of schools' correlation coefficient were 0.76 and 0.82 for concurrent and predictive validity respectively.

Table 2: Pearson's 'r' and t-test of the relationship between criterion-related (concurrent and predictive validity) on students' reliability in biology essay tests.

Criterion related Validity	Pearson's	Calculated –t	Table –t
Concurrent validity	0.76*****	4.25	2.06
Predictive validity	0.82*****	11.29	2.10

***** Significant correlation co-efficient

The results of the hypothesis are presented in table 2 above.

Out of the 10 schools studied, the relationship between the two groups was significant at 5% confidence level in both the concurrent and predictive validity with their concurrent value of 4.25 and 11.29 respectively. Since the calculated t-value is greater than the table value, the null hypotheses were rejected.

Discussions of results

From the analysis of the results, it was observed that the students' reliability in essay tests in relation to the two types of criterion-related validity (concurrent and predictive validity) were both positive and high for the two forms of validity.

The magnitude of the correlation coefficients among the various schools is comparable to Daniel's (2000) in which essay scores were correlated with objective test scores in mock-WASC with WASC O'Level results. The finding was significant for they have shown some degrees of reliability and validity that existed especially in the objective and essay tests. The variation in relationship between objective and essay test scores which ranged from very high/near perfect to very low/virtually no relationship could be attributed to certain factors.

The high relationship could be as a result of adequate staffing, both qualitatively and quantitatively in the different schools and in Biology. Qualitative staffing involves the handling of the subjects by teachers who were adequately informed in the course of measurement and evaluation and in the subject matter. Quantitative aspect implies appropriate teacher-student ratio. Similarly, Ali (2002) observed that the quality of staff affects students' performance in Biology subjects. The existence of qualitative and quantitative staff attributes lead to high quality instructions and evaluation. These were highlighted in the work of Azikiwe (1999).

This affects among other things the quality of constructions, evaluation instrument and the neglect of other evaluation techniques such as objective tests. This might be why Maduabum (2004) noted that for evaluation to be effective, the professional evaluators/teachers must use various techniques to evaluate the success of his pupils' learning and receive feedback from them to enable him to be sensitive to their needs. Supporting the view, Ebuoh (2014), Agusiobi and Olaitan (2001) corroborated that non standardization is likely to cause difference in students' performance in essay and practical in Senior Secondary mock Biology.

Conclusions

1 The relationship in the students' reliability in biology essay tests was positive and high in the two forms of criterion-related validity (concurrent and predictive validity)

2 There was significant relationship between students' reliability in essay tests in biology in all the two forms of validity (concurrent and predictive validity)

Recommendations

Recommendations were made based on the findings of the study.

1 Topics on validity (criterion-related validity) should be included in the course content of teachers in the tertiary institutions.

2 In a case where the students' performances in the first and second tests are reliable they could be used to determine the students' performance using either the concurrent and predictive validity.

References:

Agusiobo. N.O. & Olaitan, O.S. (2001). *Principles of practical teaching.* New York: John Wiley and Sons.

Ali, A. (2001). *A handbook of science method for secondary school teachers.* Nsukka: Institute of Education.

Azikiwe, U. (1999). Continuous assessment in the post- primary institution: Constraints and strategies for implementation. Implementation of National Policy on education: Theoretical and empirical analysis. *Nigeria Educational Research Association.6 [2]* 13

Daniels, U.C. (2000). 'Testing Geography at O/L of the GCE. *The British Journal of Education Psychology*. 24.

Ebuoh, C.N., (2014). Relationship between students' performance in essay and practical test in senior secondary biology. *Academic Discourse: An International Journal .7* (1)298-307.

Ebuoh, C.N., (2004). An investigation into the extent of use of laboratory method in teaching Biology in Senior Secondary Schools in Enugu urban of Anambra State. *Unpublished B Ed Project Report.* Nsukka: University of Nigeria.

Ewuzie, A.R. (2005). Academic performance at Senior Secondary Biology: Impact of activity and essay question methods. *Unpublished M.Ed dissertation Report:* Nsukka: University of Nigeria.

Ibekwe, C. (1995). A comparative study of two evaluation methods in the evaluation of integrated science to JSS II students. *Unpublished M.Ed dissertation Report.* Nsukka: University of Nigeria.

Maduabum, M. A. (2004). *Teaching biology effectively*. Jos: Jos University press limited.

Nworgu, B.G. (1992). *Educational measurement and evaluation*: *Theory and practice*. Awka: Haliman Publishers.

Okeke, F.N. (2010). Students' Mock-WASC/GCE success in science subjects as a prediction of their achievements in WASC/GCE O/L in Anambra State. *Unpublished Master's dissertation of* University of Nigeria, Nsukka.

Omelewa, M. (2007). Some Earliest Problems of Science education in Nigeria (1 859-1932). *Journal of Science Teachers Association of Nigeria.* 15 [1] 22.

Udeigwe, A. (2011). Relationship between continuous assessment scores and students' performance in SS 2 in Awgu L.G.A of Enugu state. *Unpublished Masters Dissertation Report* of University of Nigeria. *Nsukka:*