

Effects of Analytical and Holistic Scoring Patterns on Scorer Reliability in Biology Essay Tests

Casmir N. Ebuoh^{1,*}

¹Department of Science and Computer Education, Enugu State University of Science and Technology, Enugu, Enugu State Nigeria

*Correspondence: Department of Science and Computer Education, Enugu State University of Science and Technology, Enugu, Enugu State, Nigeria. Tel: 234-803-744-5718. E-mail: ebuohcas@yahoo.com

Received: November 19, 2017

Accepted: February 20, 2018

Online Published: February 25, 2018

doi:10.5430/wje.v8n1p111

URL: <https://doi.org/10.5430/wje.v8n1p111>

Abstract

Literature revealed that the patterns/methods of scoring essay tests had been criticized for not being reliable and this unreliability is more likely to be more in internal examinations than in the external examinations. The purpose of this study is to find out the effects of analytical and holistic scoring patterns on scorer reliability in biology essay tests in Enugu Education Zone of Enugu State. Post-test only control group design was used. All the 212 Biology teachers from the thirty one secondary schools in Enugu Education Zone of Enugu State were used. The researcher constructed Biology essay test (BET) with scoring guide on biology essay tests (SGBET) for data collection. The research question was answered using Pearson's Product moment correlation coefficient (r). The hypothesis was tested using t-test. The results showed that the use of the holistic scoring pattern was found to be the most reliable followed by analytical scoring pattern. Recommendations were made based on the findings of the study.

Keywords: effects of analytical, effects of holistic scoring, scorer reliability and biology essay tests

1. Introduction

One of the claimed factors that are responsible for the unreliability in scoring essay tests is the scoring pattern of essay tests. Scoring patterns are those methods that are employed in the scoring of essay tests in order to achieve higher scorer reliability. Some of the scoring patterns are conventional pattern of scoring all items, ranking all scripts before scoring all items, re-arrangement of the order of the papers before scoring, dividing the task of scoring into session, scoring an item across board use of independent scorers, analytical and holistic scoring patterns. The scoring of essay tests had been criticized for not being reliable because of the use of inadequate scoring patterns. There was evidence too to show that the level of unreliability of essay tests appears more likely to be more in the internal than external examinations (Ebuoh, 2007).

This is more apparent if we consider the fact that despite the standardized nature of the senior secondary school certificate Examination, significant variations are found in the scores. This is exemplified by the result of SSCE Biology essay scores in the May/June, 2011. This is shown in Table 1.

Table 1. Scores Given by Team Leaders and Assistant Examiners in Scoring Biology Essay Test

S/N	Script Number	Examiners Number	Examiners	Scores	Deviation (Highest Lowest)
1.	525260/313	5024/870	Assistant	Original	15
		5024/130	Team Leader	Vetted	17
2.	525260/017	5024/873	Assistant	Original	33
		5024/030	Team Leader	Vetted	35
3.	525260/056	5024/892	Assistant	Original	19
		5024/291	Team Leader	Vetted	28
4.	525260/001	5024/820	Assistant	Original	37
		5024/120	Team Leader	Vetted	39

Source: West African examination council (2014).

Table 1 shows that out of the four randomly selected vetted scripts of assistant examiners by Team leaders, all were under scored. Further, analysis of the relationship between the assistant examiners' scores and team leaders' scores using spearman's rank order statistic indicated a coefficient value of -0.10. This very low negative relationship is a pointer to the problem of different scores in scoring student's scripts even with the same scoring/marking guide. It was claimed that low level of scorer reliability has been blamed as a result of inappropriate use of scoring patterns in scoring essay tests such as analytical and holistic scoring patterns.

In analytical scoring, the teacher develops a list of the major points/elements he expects the students to include in the response.

The analytical scoring type is more suitable for scoring the restricted response format. E.g. of analytical scoring scheme

Enunciate four features of a computer (4marks)

Answer key: One mark each for any of the following:

- i. Reliability
- ii. Versatility
- iii. Durability
- iv. Fast or speed (score full mark (1) for either fast or speed
- v. Storage capacity
- vi. Programmable Strengths of the Analytical Scoring

Strengths of the Analytical Scoring

1. It helps teachers to keep the full range of writing features in mind as they score.
2. The diagnostic nature of analytical scoring helps students to know areas they need to improve on in the essay writing skill.
3. It helps to minimize Halo effects and Leniency error. Halo effect is the tendency for an impression in one area to influence opinion in another area. While leniency error occurs when a teacher rates or scores students too positively.
4. It is good for scoring large number of short, specific items.

Weaknesses of Analytical Scoring

1. It is time-consuming. For each piece of writing the teacher is expected to make not less than 11 separate judgments.
2. Students most times do not make use of the analytic comments (feedback) from the teachers.
3. Negative/sarcastic feedback can be pedagogically destructive.

Guidelines to be used to Maximize Effective Analytical Scoring

- i. An analytic scale should be designed /written as it helps to define grading criteria clearly. It should also be shared with the students to help them understand what is expected of them and how their responses would be assessed.
- ii. Criteria are weighed according to their importance. For example if the goal of a test is to ascertain the level of assimilation of course material, then logic, ideas, organizational skill and ingenuity are scored higher than grammar and mechanics:
- iii. Feedback becomes formative and effective when the comments are balanced and both support and challenge students.
- iv. Teachers should as much as possible avoid sarcasm in their comments, cancelling students work with lines and all other forms of destructive criticism (Maduabum 1994 and Ali 1998).

2. Holistic Scoring

The holistic scoring involves the teachers selecting some students' answers that are graded as high, average or low achievement. This selected answers become the model by which the teacher assesses and scores the other answers. This style of scoring is most appropriate for scoring long essay.

2.1 Strengths of Holistic Scoring

- i. It is more reliable than the analytical scoring because it requires that the teacher and one or two other readers/scorers read the essays to determine which is stronger or weaker among the model essays.
- ii. It is efficient and takes less time than analytical scoring.

2.2 Weaknesses of Holistic Scoring

- i. Unlike the analytical scoring where the teacher includes formative comments, the holistic scoring does not make this available for student. The model essays cannot be given to the students for comparison.
- ii. Holistic scoring is impractical for individual use. It is better used as a team (Harbor-Peters 1999).

The purpose of the study is to investigate into the effects of analytical and holistic Scoring patterns on scorer reliability in biology essay tests.

The research question was formulated to guide the study.

What is the difference in the scores awarded by scorers in the two different scoring patterns (analytical and holistic Scoring patterns).

2.3 The Null Hypothesis Was Tested at 0.05 Level of Significance

There is no significant relationship in the correlation coefficient of the scoring patterns of scorers who scored biology essay test using analytical and holistic Scoring (the two different) scoring patterns.

3. Methodology

3.1 Design of the Study

The research is a quasi- experimental study of a post-test only control group design. The subjects are assigned to the groups by randomization. No pre-test is used. The randomization controls for the possible extraneous variables and assures that any initial difference between the groups is attributed only to chance and followed the laws of probability (Nworgu, 2006, Ary, Jacobs and Razavieh, 2001). The scorers were randomly assigned to experimental group I, and experimental group II. The treatment of the subject (scores) was done as indicted below:

Table 2. Assignment of Scorers to Treatment Groups

		Groups	Independent Variables	Pos-test scores
Experimental Group I	R	E ₁	ASP	O ₂
Experimental Group II	R	E ₂	HSP	O ₂

Where,

E1 = Experimental Group One

E2 = Experimental Group Two

O₂ = Post-test treatment and observations

R = Randomization

ASP = Analytical scoring pattern treatment on experimental group one

HSP = Holistic scoring pattern treatment on experimental group two.

3.2 Area of the Study

The study covered all the schools in Enugu Education zone of Enugu State. The researcher adopted the educational administrative structure in which Enugu State was divided into six education zones. These are Enugu, Udi, Nkanu, Awgu, Nsukka and Oboloafor Zones. The choice of this area is because of logistical convenience and the researcher sees the zones as a thickly populated zone in terms of Biology teachers among all the six zones in Enugu State.

3.3 Population of the Study

The population for this study comprises all the 212 secondary school Biology teachers in all the 31 secondary schools in Enugu Education Zone.

3.4 Sample and Sampling Techniques

In consideration of the fact that only secondary school Biology teachers were used for the study and because the number of Biology teachers is not too large, the researcher used all the Biology teachers for the study. The use of all the Biology teachers further helped the researcher in avoiding sampling errors. The Biology teachers were randomly assigned to experimental group 1 and experimental group 11.

All the schools in Enugu Education zone were stratified into Enugu East, Enugu North and Isi-Uzo Local Government Areas. The random sampling technique was used to select three secondary schools from each of the three local government areas making a total of nine secondary schools. The schools selected had up to one or more streams (classes) of SS III students. A simple random sampling technique was used to select 20% of students in the schools picked. A total of 220 SS III students were finally selected from 2215 students.

3.5 Instruments for Data Collection

The researcher constructed Biology Essay Test (BET) with scoring Guide on Biology Essay Test (SGBET) for the study.

3.6 Biology Essay Test (BET)

The Biology Essay Test (BET) was developed based on the following Biological topics: cell organization, sense organs, nutrition and transportation in living things. The BET contained five essay items with 3 sub items in each item, which ranged from A to E. The BET was both restricted and unrestricted Biology essay tests. The items measured objectives in the cognitive and psychomotor domains of taxonomy of educational objectives. The weight of the objective levels were based on the proportion of low and high order levels of cognitive and psychomotor domains as suggested by Margret (2000) in the same units of study in the senior secondary school Biology Curriculum. This is because, it was observed that students do not normally exceed the comprehension level (higher cognitive level) by the time they had completed senior secondary school programme (Ezeudu, 1997 and Gagne, 2000)

SGBET

Scoring guide was developed for the scoring of the Biology Essay Test (SGBET). The SGBET contained all the answers to the five items of the BET with 3 sub answers in each item which ranged from A to E. The responses are restricted response type. SGBET was developed to guide the scorers to score the BET using the two scoring patterns (analytical and holistic Scoring patterns).

3.7 Validation of the Instruments

BET

The test blueprint (table of specification) helped to measure the content validity of the instrument. The test blue-print on the BET as well as the BET were face validated by three experts drawn from the sub-department of science Education (two Biology specialists and two Measurement and Evaluation specialists of University of Nigeria, Nsukka. Their criticism and vetting helped in modifying and/or replacing some test items. The weight of objective level on cognitive domain was based on the proportion of low order level (memorization of facts) and higher order level (application) (Margret, 2000).

SGBET

The SGBET was face validated by three specialists one in Measurement and Evaluation and two specialists in Biology Education of Enugu State University of Science and Technology. Their criticism and vetting helped in modifying and/or replacing some answers and items.

3.8 Reliability of the Instruments

In order to establish the coefficient of internal consistency of the instruments used for the study, the following steps were taken. In establishing the coefficient of internal consistency of BET, scores generated from the 20 SS III Biology students used for the trial test were subjected to the Cronbach Alpha formula and was found to be 0.88. The Cronbach Alpha was considered appropriate since BET consisted of essay items. This internal consistency gives homogeneity of the test items in the instrument.

3.9 Experimental Procedure

The 212 Biology teachers drawn from the 31 secondary schools in Enugu Education zone were randomly assigned to the treatment conditions as Experimental Group I and Experimental Group II. 106 Biology teachers were assigned to each experimental group through simple random sampling. balloting without replacement was done using the names of the Biology teachers.

1. Experimental group I: Analytical Scoring patterns (ASP)
2. Experimental group II: Holistic Scoring patterns (HSP).

The experiment lasted for three days of coordination and 10 days of scoring the Biology essay test. This was also the only period the school authorities in the schools could allow the researchers to use the scorers.

3.10 Method of Data Analysis

The research question was answered using the mean and Pearson's Product Moment Correlation Coefficient (r). The T-test was used in testing the null hypothesis.

4. Results

What is the mean difference of the scores awarded by scorers using the two different scoring patterns in scoring Biology Essay Tests?

Table 3. Mean Scores of the Scorers Who Used the Two Different (Analytical and Holistic) Scoring Patterns in Scoring Biology Essay Test

Scoring patterns	Mean scores
ASP (Group I)	14.84
HSP (Group II)	19.96

Table 3 showed that the mean scores for experimental group II and experimental group I were 19.96 and 14.8 respectively. This means that scorers who scored using holistic scoring pattern had the highest mean scores of 19.96 followed by those who scored using analytical scoring pattern with a mean score of 14.84.

There is no significant relationship in the correlation coefficient of the scoring patterns of scorers who scored Biology essay test using the two different scoring patterns.

Table 4. Pearson's Product Moment Correlation Coefficient (r). and T-Test of the Relationship in the Scorer Reliability of the Two Group of Scorers

Scoring patterns	No of Scorers	Pearson's (r)	Calculated- t	Critical -t
ASP	16	0.42*	3.57	2.15
HSP	16	0.64*	4.68	2.15

* = **Significant relationship**

It was observed that significant relationship existed between group I and II. This implied that group II scored significantly higher than group I. The implication of the results is that the use of holistic scoring patterns (HSP) indicated superiority over analytical scoring patterns (ASP) of scoring biology essay test. This is because; mean level is a measure of superiority.

5. Discussion of the Results

It was found out that the relationship of scores of the scorers who used analytical and holistic scoring patterns was positive. The magnitudes of their positive correlation coefficients are 0.64* for holistic scoring patterns (HSP) and 0.42* for analytical scoring pattern (ASP)

The values of the correlation coefficient of analytical and holistic scoring patterns are positive and high medium relationship. The relationship was significant for the group of scorers who scored biology essay test using analytical and holistic scoring patterns. This is because the calculated t value of 4.68 for holistic Scoring patterns was much

higher than the critical t value of (2.15). This implied that the use of analytical and holistic scoring patterns had significant relationship on scorer reliability in scoring Biology essay test. This finding was in line with those of similar experimental studies in science and science related subjects (Ezeudu 1995; Osisioma, 1995 and Okafor 2000) where the experimental treatment groups proved better than the control group.

6. Conclusions

The following conclusions are drawn based on the findings of the study:

- 1 Analytical and holistic Scoring patterns had positive significant relationship in scoring Biology essay test.
- 2 Furthermore, use of holistic Scoring pattern was found to be outstandingly more efficacious than analytical scoring pattern

7. Recommendations

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

- 1) The use of analytical and holistic scoring patterns were found efficacious in engendering scorer reliability in scoring biology essay test and since the techniques are not yet popular in our school system. They should be incorporated in the curriculum for teacher training institutions.
- 2) Obviously, the serving teachers lack the necessary competencies to develop the analytical and holistic scoring patterns to equip these categories of serving teachers, professional association such as Measurement and Evaluation Association of Nigeria (MEAN) and government agencies should organize workshops, seminars and conferences for them on the two patterns.
- 3) On acquiring the necessary skills, the teachers should be encouraged to employ these techniques more in scoring Biology essay test so that scorers (teachers) will no longer be scared because of the tediousness of scoring essay test.
- 4) The opinion of the researcher is that any professional development for pre-service and in-service teachers must include opportunities to learn the scoring patterns. Biology teachers in Nigeria must not be left out. It is necessary to ensure that Biology teachers in Nigeria acquire the required patterns to be used in scoring biology essay test and move beyond the use of the conventional approach in scoring Biology essay tests.

References

- Ali, A. (1998). Inferential Statistic Techniques. In S. O. Olaitan & G. I. Nwoke (eds.), *Practical research methods in Education* Onitsha: Summer Education Publishers.
- Ary, D., Jacobs, L. C., & Razavieh, A. (2001). *Introduction to research in Education* (2nd Ed.). New York: Rinehart and Winston.
- Ebuoh, C. N. (2007). Effects of scoring patterns on scorer reliability in Biology essay tests. *African Journal of Educational Foundations*, 3(2), 288-301
- Ezeudu, F. O. (1995). *Effects of concept maps on students achievement, interest and retention in selected units of organic chemistry*. Unpublished Ph.D. Dissertation, Nsukka: University of Nigeria.
- Ezeudu, S. A. (1997). Grading and reporting of pupils' progress. In S. A. Ezeudu, U.N.V Agwagah & C. N. Agbaegbu (eds.), *Educational Measurement and Evaluation for Colleges and Universities*. Onitsha: cape Publishers' Limited.
- Gagne, R. M. (2000). *Condition of teaching* (3rd ed.). New York Holt Reinhart and Winston.
- Harbour-Peters, V.F.A. (1999). *Noteworthy points on measurement and evaluation*. Enugu: Snaap Press Limited.
- Maduabum, M. A. (1994). *Teaching Biology effectively*. Jos: Jos University Press Limited.
- Margaret, R. (2000). *Introduction to Measurement in Physical Education and Sciences* (2nd ed.). Missouri: Mosby college Publications.
- Nworgu, B. G (2006). *Introduction to Educational Measurement and evaluation: theory and practice* (2nd ed.). Nsukka; Hallman Publisher.
- Okafor, G. A. (2000). *Effects of note taking patterns on students academic achievement, interest and retention in*

Geography. Unpublished Ph.D Dissertation. Nsukka: University of Nigeria.

Osioma, U. I. N. (1995). *Effects of mode of concept mapping and gender on students achievement in and attitude towards integrated science*. Unpublished Ph.D. Thesis. Nsukka: University of Nigeria

The West African Examination (2011). Vetting sheet on Biology.