

Effects of Scoring by Section and Independent Scorers' Patterns on Scorer Reliability in Biology Essay Tests

Dr. Casmir N. Ebuoh

Department of Science and Computer Education
Enugu State University of Science And Technology, Ebeano City

Prof. S. A. Ezeudu

Department of Science Education, University of Nigeria, Nsukka

Abstract

The study investigated the effects of scoring by section, use of independent scorers and conventional patterns on scorer reliability in Biology essay tests. It was revealed from literature review that conventional pattern of scoring all items at a time in essay tests had been criticized for not being reliable. The study was true experimental study of post-test only control design. All the 42 Biology teachers from the 23 secondary schools in Enugu Education zone were used. The 42 Biology teachers were assigned to three groups namely: experimental group I: scoring by section, experimental group II: scoring by independent scorers and group III: conventional pattern of scoring all items. The treatment lasted for 13 days (three days coordination and 10 days scoring exercise). The research question was answered by calculating the mean scores and Kendal's coefficient of concordance (w). The hypothesis was tested using t-test. The results revealed that the use of independent scorers was found to be the most effective followed by the scoring by section while conventional pattern was not significant. Recommendations were made based on the findings.

Impact name: Effects of scoring, section and independent patterns, scorer reliability, Biology essay tests.

Introduction

The scoring of essay test had been criticized for not being reliable. There was evidence too to show that the level of unreliability of essay tests appears more likely to be more in the internal than in the external examinations. The low level of the scorer reliability had been attributed to a number of factors. It was claimed that low level of scorer reliability is caused by the use of inappropriate scoring patterns in scoring essay tests (Cox, 1989).

Scoring patterns in this study mean the various methods that are employed by scorers to obtain the quantitative performances of learners. Various scoring patterns have been reported in literature.

They include:

- i. Scoring all items in a candidates script before picking up another script (Ezeoke, 1986 and Maduabum (1984).
- ii. Use of independent scorers (Ezeoke, 1986 and Harbor – Peters, 1999).
- iii. Scoring by section pattern (Ukeje, 1984 and Ezeudu, 1997).
- iv. Another reported pattern of scoring essay test items is where the scoring involves the task of scoring into sessions (Lovegroove, 1984).
- v. Re-arrangement of the script according to the order of quality of item responses (either from the poorest quality or from the highest) before scoring (Horrocks and Shoonover, 1968).
- vi. Ranking all scripts before scoring all items (Ezeoke, 1986).
- vii. Scoring an item across board (Harbor – Peters, 1997 and Quereshi, 1974).

However, among these scoring patterns reported above, patterned scoring in form of scoring by section, use of independent scorers and scoring all items in a script before picking up another script appears to be most popular in scoring essay tests.

In view of the above, the essay test scoring patterns employed in this study are

- (i) Use of independent scorers pattern (UISP)
- (ii) Scoring by section pattern (SBSP)
- (iii) Scoring all items in a script before picking up another script conventional pattern (CP).

Independent Scorers Pattern (UISP), use two or more independent scorers to score a student's script and use the average as a final score. In a situation where this is employed, the scores should not be recorded on the test booklet without finding the average of the scores but should be written on a separate sheet (Ezeoke, 1986).

This method has the advantage of checking possible biases arising from rater-ratee interaction. Again, it may rectify errors caused by oversight on the part of the rater. This advantage holds if the idea of independent scoring is maintained by ensuring that two scorers do not sit close enough to discuss their scores as they may influence each others, thereby defeating the idea of independent scoring (Harbor-Peters, 1999 and Quereshi,

1974).

Scoring by Section Pattern (SBSP) is where one scorer scores a section (part) of the test. The scorer scores his section and passes on the script to the next scorer. Evidence abounds in support of this method. For instance Lovegroove (1984) advised that in an examination in which crucial decision may be taken, two or more scorers should be allowed to score a section of the script independently. It is argued that it increases the speed of scoring and gives opportunity of inter-item comparison rather than complete test comparison (Ezeoke, 1986).

The conventional pattern presupposes that the teacher/scorer scores all items in a script before attempting to score any other script. The method has been strongly criticized by scholars. One of its major criticisms is that teacher/ scorer biases are made manifest in the way they score the scripts. It would be most difficult to be consistent when handling different items in a script one after another. That is, to give the same score to similar answers to a particular question in all the scripts that are scored (Ezeoke 1986).

Another disadvantage of this pattern is that it causes "halo effect." This means that if the answer to a question is a very good one. It influences the scoring of the next answer/items. Likewise, if the first item to be scored is a poor item, the subsequent items could be scored within the same context.

In view of the above the avoidance of the halo effect can be achieved according to Powell and Lobster (1974) if each response is judged on its merit without regard to other success or failure. The scorer must guard against allowing his scoring to be influenced by any general impression the scorer formed of the subject/candidate's ability. There is a natural tendency to over-estimate the ability of a bright and self-confident student. Scoring must not be tempered with any conviction that the subject could have been answering correctly. The task is to score the response which has actually been given.

It appeared that in scoring of senior secondary essay tests, scorers are more conversant with the use of the conventional pattern of scoring all items on a student script than they are with other patterns. A question that arises is: is the conventional pattern better than the other patterns in achieving higher scorer reliability? It appears that there is no empirical study so far that compared relative effectiveness of the various patterns of scoring biology essay test in terms of which one engenders higher scorer reliability than others using patterns above.

The purpose of the study was to investigate the effects of three different scoring patterns (scoring by section, independent scorers and conventional) patterns on scorer reliability in biology essay tests.

The research question was formulated to guide the study.

What is the mean difference in the scores awarded by scorers in the three different scoring patterns?

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 the three different scoring patterns.

Research Method

The research is a true experimental study of a post-test only control groups 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 differences between the groups are attributed only to chance and therefore followed the laws of probability (Ary, Jacobs and Razavieh, 1979). The scorers were randomly assigned to experimental group I, experimental group II and control group III. The treatment of the subject (scores) was done as indicted below:

Table I:

Assignment of scorers to treatment groups

	Groups		Independent Variables	Pos-test scores
Experimental Group I	R	E ₁	UASP	O ₂
Experimental Group II	R	E ₂	SBSP	O ₂
Experimental Group III	R	C	CPSAL	O ₂

Where,

E₁ = Experimental Group One

E₂ = Experimental Group Two

C = Experimental Group Three

O₂ = Post-test treatment and observations

R = Randomization

SBSP = scoring by section pattern treatment on experimental group

UISP = Use of independent scorers pattern treatment on experimental group two

CPSAI = Conventional pattern of scoring all items on Control group three

The study covered all the schools in Enugu Education zone of Enugu State. The researcher adopted the educational administrative structure in which Enugu State is 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 saw the zones as a thickly populated zone in terms of Biology teachers among all the six zones in Enugu State.

- (a) The population for this study comprises all the 42 secondary school Biology teachers in all the 23 secondary schools in Enugu Education Zone.

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, Experimental Group 11 and Control Group 111 (See Table 1)

All the schools in Enugu Education zone were stratified into Enugu East, L.G.A; Enugu North L.G.A and Isi-Uzo L.G.A. 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 to administer the test items that were scored by the scorers.

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

1. Biology Essay Test (BET) with scoring Guide on Biology Essay Test (SGBET)

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 Blooms (1956) 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 (1990) 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 (Sturoges, 1972)

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 five experts drawn from the sub-department of science Education (two Biology specialists and three Measurement and Evaluation specialists of University of Nigeria, Nsukka and Enugu state University Science and Technology 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, 1990).

Scoring guide was developed for the scoring of the Biology Essay Test (SGBET). The SGBET contained all the answer to the five items of the BET. The responses are restricted response type.

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

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 30 SS III Biology students used for the trial test were subjected to the Cronbach Alpha formula and found 0.91. The Cronbach Alpha was considered appropriate since BET consisted of essay items. This internal consistency gives homogeneity of the test items in the instrument.

The 42 Biology teachers drawn from the 23 secondary schools in Enugu Education zone were randomly assigned to the treatment conditions as Experimental Group 1, Experimental Group II and the third regular scores were the control group. 14 Biology teachers were assigned to each experimental group through simple random sampling. Balloting without replacement was done using the names of the Biology teachers. Experimental group I: Scoring by section pattern all items (SBSP)

1. Experimental group II: Used Independent scorers pattern UISP.
2. Control group III: Used conventional pattern of scoring all items (CPSAI)

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.

The research question was answered using the mean and Kendall's Coefficient of Concordance (W). The t-test was used in testing the null hypothesis one. All analysis was carried out using computer.

RESULTS

RESEARCH QUESTIONS

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

The result in response to research question one is shown in table 2

Table 2: Mean scores of the scorers who used the three different scoring patterns (SBSP, UISP and CPSAI) in scoring Biology essay test.

Scoring pattern	Mean scores
SBSP: scoring by section pattern (Group I)	14.84
UISP: use of independent scorers pattern (Group II)	16.96
CPSAI: conventional pattern of scoring all items (Group III)	10.92

Table 2 showed that the mean scores for experimental group II experimental group I and control group III were 16.96, 14.84 and 10.92 respectively. This means that scorers who scored using independent scorers pattern had the highest mean scores of 16.96 followed by those who scored by section pattern with a mean score of 14.84 and the group who scored with conventional pattern of scoring all items had the least mean of 10.92.

Research Hypothesis

There is no significant relationship in the correlation coefficient of the scoring patterns of scorers who scored Biology Essay Test using the three different scoring patterns.

The hypothesis one above was stated to investigate the significant relationship of the three scoring patterns in scoring Biology Essay Test and to test the statistical significance at five percent level of significance.

The result of the analysis of the significant relationship obtained from the three different scoring pattern were shown in the table 3 below.

Table 3: Kendall's (W) and t-test of the relationship in the scorer reliability of the three group of scorers.

Scoring patterns	No of Scores	Kendal's (W)	Calculated- t	Critical -t
SBSP	14	0.42*	2.57	2.15
UISP	14	0.51*	5.68	2.15
SAIP	14	0.13	1.09	2.15

* = Significant relationship

Table 3 provided data for testing hypothesis of this study. It was observed that significant relationship existed among group I and II, group I and group III, group II and group III. This implied that group II scored significantly higher than group I and group III, and that group I scored significantly higher than group III.

The implication of the results is that the use of independent scorer pattern indicated superiority over first (scoring by section pattern) and third (conventional pattern of scoring all items)S patterns of scoring biology essay Test. Similarly, the scoring by section pattern showed superiority over the third pattern of scoring biology essay Test. Mean level is a measure of superiority.

The results are discussed according to the formulated research question and hypothesis which were presented under the heading below:

Effects of the three scoring patterns on scorer reliability in scoring Biology essay test.

It was found out that the relationship of scores of the scorers who used independent scorers pattern, scored by section and scored all items pattern were positive. The magnitudes of their positive correlations coefficients were 0.51* for use of independent scorers pattern, 0.42* for scoring by section pattern and 0.13 for conventional pattern of scoring all items respectively. The magnitudes of the positive correlation coefficient for UISP, SBSP were medium relationship while CPSAI had very low relationship.

The relationship was significant for those groups of scorers who scored biology essay test using independent scorers pattern (UISP) and scoring by section pattern (SBSP). This is because the calculated t value of 5.68 for UISP was much higher than the critical t value (2.15) and the calculated t value of 2.57 for CPSAI was more than critical t value of 2.15, contrary, the calculated t value of 1.09 for CPSAI lower than critical t value of 2.15. This implied that the use of independent pattern and scoring by section pattern had significant relationship on scorer reliability in scoring biology essay test.

Further, analyses in order to determine the mean differences among the various groups of scorers were carried out. From table 2, it was observed that the mean scorers for experimented group I, experimental group II and control group III were 14.84, 16.96 and 10.93 for SBSP, UIAP and SAIP respectively. This means that scorers who scored using independent scorers pattern (experimental group II, had the highest mean scorers of 16.96 followed by those who scored by section pattern had 14.84 and the group who scored with the

conventional pattern of scoring all items (experimental group III) had the least mean score of 10.92.

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.

Conclusion

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

Use of independent scorers and scoring by section patterns had positive significant relationship in scoring Biology essay test.

Furthermore, use of independent scorer's pattern was found to be outstandingly more efficacious than scoring by section pattern. The effect of the conventional pattern of scoring all items was not significant on scorer reliability in scoring biology essay test.

Recommendations

Based on the findings of the study, the following are the recommendations:

1. The use of independent scorers pattern and scoring by section pattern 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 use of the independent scorers and scoring by section 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

- Ary, D. Jacobs, L. C & Razavieh, A (1979). *Introduction to research in Education*. (2nd ed). New York: Rinehart and Winston.
- Ali, A (1988) *Inferential Statistic Techniques*. In S. O. Olaitan & G. I. Nwoke. *Practical research methods in Education* (eds) Onitsha: Summer Education Publishers.
- Anastai, A (1981) *Psychological testing* (2nd ed.) New York: The Macmillian Company.
- Cox, R. (1969) *Reliability and validity of Education*. London: Evans Brothers
- Ebuoh, C. N. (2007) Effects of scoring patterns on scorer reliability in Biology essay tests, *African Journal of Educational Foundations*. 3 (2) 288-301
- Ezeoke, J. O. (1986) *Theory and practice of continuous assessment*. (2nd ed) Ihiala: Deo Gratis Press.
- Ezeudo, 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*. Ontisha: cape Publishers' Limited.
- Ferguson, A. C. (1989) *Statistical analysis in Psychology and Education*. New York: McGraw Hill.
- Gagne, R. M. (1977) *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.
- Horrocks, J. E. and Shoonever, T. I. (1998) *Measurement for teachers*. Columbus: Ohio Charles E. Meril Publishing Company.
- Lovegroove, M. N. (1984) Evaluating the result of learning. In B. O. Ukeje (ed) *Foundations of Education*. Benin City: Ethiope Publishing coporate
- Maduabum, M. A (1984) *Teaching Biology effectively*: Jos: Jos University Press Limited.
- Margaret, R. (1990) *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
- Sturoges, P. T. (1972) Effects of Instructions and forms of information feedback on retention of meaningful materials. *Journal of Education Psychology*. 6 (2) 77 – 90.
- The West African Examination (2005) Vetting sheet on Biology.

The IISTE is a pioneer in the Open-Access hosting service and academic event management. The aim of the firm is Accelerating Global Knowledge Sharing.

More information about the firm can be found on the homepage:

<http://www.iiste.org>

CALL FOR JOURNAL PAPERS

There are more than 30 peer-reviewed academic journals hosted under the hosting platform.

Prospective authors of journals can find the submission instruction on the following page: <http://www.iiste.org/journals/> All the journals articles are available online to the readers all over the world without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. Paper version of the journals is also available upon request of readers and authors.

MORE RESOURCES

Book publication information: <http://www.iiste.org/book/>

Academic conference: <http://www.iiste.org/conference/upcoming-conferences-call-for-paper/>

IISTE Knowledge Sharing Partners

EBSCO, Index Copernicus, Ulrich's Periodicals Directory, JournalTOCS, PKP Open Archives Harvester, Bielefeld Academic Search Engine, Elektronische Zeitschriftenbibliothek EZB, Open J-Gate, OCLC WorldCat, Universe Digital Library , NewJour, Google Scholar

