CASMIR N. EBUOH (PH D) AND GABRIEL A OKAFOR (PH.D)

EFFECTS OF SCORING BY SESSION, RANKING AND CONVENTIONAL PATTERNS ON SCORER RELIABILITY IN BIOLOGY ESSAY TESTS IN SECONDARY SCHOOLS IN EZEAGU LOCAL GOVERNMENT AREA OF ENUGU STATE

BY
CASMIR N. EBUOH (PH. D)
Department of Science and Computer Education, Enugu State University of Science and Technology, Enugu.

AND

GABRIEL A OKAFOR
Department of Science and Computer Education, Ebonyi State University, Abakaliki

Abstract:

Literature revealed that conventional pattern of scoring all items (CPSAI) at a time in 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. Post-test only control group design was used. All 48 Biology teaches from 23 secondary schools in Enugu education zone were used. The researcher constructed Biology essay test (BET) with scoring guide on Biology essay test (SGBET) for data collection. The research question was answered using Kendal's (w). The hypothesis was tested using t-test. The results revealed that the use of the Ranking all scripts before scoring all items (RASBSAI) was found to be the most effective followed by dividing the scoring into session pattern (DSSP). Conventional pattern of scoring all items (CPSAI) was not significance.

Background of the Study:

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 test in order to achieve higher scorer reliability. Some of the scoring patterns are conventional pattern of scoring all items at a time, ranking all scripts before scoring all items, re-arrangement of the order of the papers before scoring, dividing the task of scoring by sessions pattern, scoring by section pattern, scoring an item across board and use of independent scorers.

The scoring of essay test had been criticized for not being reliable because of the use of conventional pattern of scoring all items at a time. 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.

This is more apparent if we consider the fact that despite the standardized nature of the senior secondary school certificate (SSC) Examination, significant variations are found in the scores. This is exemplified by the result of SSCE Biology essay scores in the May / June, 2008. 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
2.	525260/017	5024/873	Assistant	Original	33	

243

INTERNATIONAL JOURNAL OF EDUCATIONAL RESEARCH VOL 11. NO1, 2011. ISSN-978-2568-51-1
CASMIR N. EBUOH (PH D) AND GABRIEL A OKAFOR (PH.D)

		5024/030	Team Leader	Vetted	35	-2
3.	525260/056	5024/892	Assistant	Original	19	
		5024/291	Team Leader	Vetted	28	-9
4.	525260/001	5024/820	Assistant	Original	37	
		5024/120	Team Leader	Vetted	39	-2
5.	525260/182	5024/874	Assistant	Original	35	
		5024/380	Team Leader	Vetted	32	+3
6.	525260/073	5024/890	Assistant	Original	16	
		5024/160	Team Leader	Vetted	26	-10
7.	4290708/57	45042/570	Assistant	Original	28	
		5042/131	Team Leader	Vetted	19	-9
8.	4290708/31	5042/723	Assistant	Original	33	
		5042/13	Team Leader	Vetted	20	+13
9.	4290708/88	5042/70	Assistant	Original	33	
		5042/314	Team Leader	Vetted	27	+6
10.	4290708/236	5042/900	Assistant	Original	17	
		5042/60	Team Leader	Vetted	17	0
11.	429072/001	5042/1238	Assistant	Original	36	
		5042/48	Team Leader	Vetted	31	+5
12.	429072/010	5042/1236	Assistant	Original	14	
		5042/148	Team Leader	Vetted	19	-5
13.	429072/111	5042/394	Assistant	Original	26	
		5042/94	Team Leader	Vetted	15	+11
14.	429072/91	5042/1260	Assistant	Original	39	
		5042/230	Team Leader	Vetted	36	+3
15.	429072/187	5042/1336	Assistant	Original	27	
		5042/066	Team Leader	Vetted	17	+10
16.	429072/143	5042/1214	Assistant	Original	16	
		5042/151	Team Leader	Vetted	20	-4
17.	4290703/137	5042/187	Assistant	Original	18	
		5042/65	Team Leader	Vetted	18	0
18.	4290703/119	5042/195	Assistant	Original	12	
		5042/75	Team Leader	Vetted	20	-8
19.	4290703/218	5042/1411	Assistant	Original	33	
		5042/80	Team Leader	Vetted	25	+8
20.	4290703/317	5042/1314	Assistant	Original	30	
		5042/92	Team Leader	Vetted	27	+3

Source: West African examination council (2008).

Table 1 shows that out of the 20 randomly selected vetted scripts of assistant examiners by team leaders, 8 was under scored, and 10 was over scored while only 2 scripts have no deviation. Further, analysis of the relationship between the assistant examiners' scores and team leaders scores using spearman's rank order statistic indicated a coefficient 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 guide. Cox, 1989 also indicated the unreliability in scoring mathematics essay tests.

The table 2 indicates the various scores awarded by five different scorers on scripts of 10 students.

Table 2. Scores given to students by five scorers in mathematics essay test.

Students	Scores					
	A	B	C	D	E	Deviation (highest-Lowest scores)
1	0	1	17	13	8	17
2	1	1	18	11	7	17
3	1	6	13	5	2	12
4	11	2	10	8	5	9
5	2	10	10	3	0	10
6	5	14	24	3	1	23
7	0	9	15	3	3	15
8	2	9	19	13	3	17
9	1	1	9	15	21	20
10	1	3	20	14	6	19

Source: COX (1989)

It appeared that in the scoring of senior secondary essay tests, scorers are more conversant with the use of the conventional pattern of scoring essay test 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 scoring by session, ranking and conventional patterns in scoring biology essay tests in terms of which one engenders higher scorer reliability than others using the patterns above.

Statement of the problem
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 likely to be more in the internal than in external examinations as shown in tables 1 and 2 above. 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 that low level of scorer reliability is caused by the use of inappropriate scoring patterns in scoring essay tests.

It appeared that in Nigeria, scorers are more conversant with the conventional pattern of scoring essay tests than they are with other methods. A question that arises is: is the conventional scoring 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 high scorer reliability than others.

Purpose of the study:
The purpose of the study was to investigate the effects of three different scoring patterns (dividing the scoring into sessions, ranking all scripts before scoring all items and conventional patterns) on scorer reliability in biology essay tests.

Research Question
The research question was formulated to guide the study.
What is the difference in the scores awarded by scorers in the three different scoring patterns (scoring by session, ranking and conventional patterns).

Research Hypothesis:
The null hypothesis was tested at 0.05 level of significance.

245

INTERNATIONAL JOURNAL OF EDUCATIONAL RESEARCH VOL 11. NO1, 2011. ISSN-978-2568-51-1
CASMIR N. EBUOH (PH D) AND GABRIEL A OKAFOR (PH.D)

There is no significant relationship in the correlation coefficient of the scoring patterns of scorers who scored biology essay tests using the three different scoring patterns.

Review of scoring patterns;

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

They are;

i.   Scoring by section pattern (SBSP)(Ukeje, 1984 and Ezeudu, 1997).
ii.   Ranking all scripts before scoring all items (RASBSAI) (Ezeoke, 1986).
iii.   Scoring an item across board (SAIAB)(Harbor peters, 1997).
iv.   Use of independent scores pattern (UISP)(Ezeoke, 1986, Harbor peters 1999).
v.   Another reported pattern of scoring essay test is where the scoring involves dividing the task of scoring into sessions (DSISP) (Lovegroove, 1984).
vi.   Rearrangement of the order of papers scoring pattern (ROPSP) (Horrocks and Shoonover, 1998).
vii.   Conventional pattern of scoring all items (CPSAI) in a script before picking up another script (Maduabum, 1984 and Ezeoke, 1986).

However, among these scoring patterns reported above, patterned scoring in form of scoring by session, Ranking of the order of the paper and conventional pattern of scoring all items in a script before picking up another script appear to be most popular in scoring essay tests.

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

i.   Dividing the scoring into sessions pattern (DSISP).
ii.   Ranking all scripts before scoring all items (RASBSAI)
iii.   Conventional pattern of scoring all items (CPSAI).

Dividing the Scoring into Sessions Pattern (DSISP).

Anastasi (1981) and Ezeudu (1997) showed evidence to suggest that it is moere reliable to divide the task of scoring into sessions. Moreso, in a situation where a large number of papers (scripts) are to be scored, the scorer could find it necessary to divide the task of scoring into several sessions. This is because, every person has changes in mood which affects his values and tolerance.

The scorers are no exceptions. A times, the scorer is more lenient and willing to accept work indicating less degree of competence. This tendency can be reduced if the scorer at the second session of scoring will read through some of the papers which were scored during the first scoring period/session. This helps to ensure reliability in scoring standards and maintain that all papers were checked nearly the same. It is pertinent to point out her that greater effectiveness of this theory as opposed to comparing the whole examinations to other examinations.

Ranking all scripts before scoring all items (RASBSAI)

This is a modified form of scoring all numbers. A recommended pattern of scoring is to rank the essay test by reading through the papers and placing them in piles. The piles could be in five. Then the piles could be ranked in quality from very superior to very inferior. Then, a scorer/teacher scores all answers in each candidate's script before picking up another script. The scorer does this for all the piles. The scorers determine how the scripts are regrouped. Then very superior pile of scripts would be given "A" The superior is "B", Average superior is "C", also inferior is D and very inferior is "E" ((Horrocks & Shonoover, 1968). However the National Examinations Council and West

African Examination Council use stanine. The piles are ranked in quality ranging from excellent (AI), very good (B2), Good (B3), credit (C4), credit (C5), credit (C6), pass (D7), and pass (E8) to fail (F9). Ezeoke (1986) opined that the pattern is reliable not time-consuming and very effective when a large number of scripts are to be scored. The ranking method of scoring essay test would seem to be the normal one, it is very doubtful if in fact the matter is so simple.

Certainly, the examiner may try to obtain an over all impression of the candidate's performance, but the examiner is bound to be influenced by this and that detail. The details sometimes may not be relevant to the assessment. It may consider things such as the work being well typed or not. Even when the details are relevant one or two of them may be so pervading such that what purports to be an overall assessments is badly biased (Cox, 1999). The advantages notwithstanding, it has peculiar outstanding deficiencies. For instance, it is not possible to determine the accuracy, comprehensiveness and the quality of the individual items in the test.

Conventional Pattern of scoring all items (CPSAI)

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.

1. One of its major criticism 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).

2. 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, self-confident, talkative child. An examiner has to be very careful to avoid scoring the responses of such a child too leniently. 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.

Research Methodology
Design of the study
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 difference 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 (scorers) was done as indicted below:

Table 3:
Assignment of scorers to treatment groups

	Groups	Independent Variables	Pos-test scores

Experimental Group I	R	E_1	DSISP	0_2
Experimental Group II	R	E_2	ROPSP	0_2
Experimental Group III	R	C	CPSAI	0_2

Where,

E1 = Experimental Group One
E2 = Experimental Group Two
C = Experimental Group Three
O_2 = Post-test treatment and observations
R = Randomization
DSISP = Dividing the scoring into session pattern treatment on experimental group.
RASBAI = Ranking all scripts before scoring all items treatment on experimental group.
CPSAI = Conventional pattern of scoring all items on Control group three

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 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 see the zones as a thickly populated zone in terms of Biology teachers among all the six zones in Enugu State.

Population of the study

(a) The population for this study comprises all the 48 secondary school Biology teachers in all the 23 secondary schools in Enugu Education Zone.

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, Experimental Group 11 and Control Group 111 (See Table 3)

All the schools in Enugu Education zone were stratified into Enugu East Local Government Area; Enugu North Local Government Area and Isi-Uzo Local Government Area. 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.

Instruments for Data Collection

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

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

CASMIR N. EBUOH (PH D) AND GABRIEL A OKAFOR (PH.D)

objectives in the cognitive and psychomotor domains. 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)

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 three scoring patterns (scoring by session, re-arrangement of scripts and conventional pattern of scoring all items).

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 five experts drawn from the sub-department of science Education (two Biology specialists and three 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, 1990).

SGBET
The SGBET was face validated by three specialists in Measurement and Evaluation as well as two specialists in Biology Education of sub-department of science education of University of Nigeria, Nsukka. Their criticism and vetting helped in modifying and /or replacing some answers and items.

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.

Biology Essay Test
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 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.

Experimental Procedure
The 48 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 scorers were the control group. 16 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: Dividing the task of scoring into session pattern of scoring all items (DSISP)
2. Experimental group II: Ranking all scripts before scoring all items (RASBSAI).

3. Control group III: Used conventional pattern of scoring all items (CPSAI)

The scores were tested on one dependent variable namely, reliability after the experimental treatment. 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.

Method of Data Analysis

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. All analysis was carried out using computer.

RESULTS

RESEARCH QUESTIONS ONE

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

The result in response to research question one is shown in table 4

Table 4: Mean scores of the scorers who used the three different scoring patterns (DSISP, RASBSAI and CPSAI) in scoring Biology essay test.

Scoring pattern	Mean scores
DSISP (Group I)	16.84
RASBSAI (Group II)	20.96
CPSAI (Group III)	11.92

Table 4 showed that the mean scores for experimental group II experimental group I and control group III were 20.96, 16.84 and 11.92 respectively. This means that scorers who scored by ranking all scripts before scoring all items pattern had the highest mean scores of 20.96 followed by those who scored by session pattern with a mean scores of 16.84 and the group who scored with conventional pattern of scoring all items had the least mean of 11.92.

Research Hypothesis One

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 patterns were shown in table 5 below.

Table 5: Kendall's (W) and t-test of the relationship in the scorer reliability of the three group of scorers.

Scoring patterns	No of Scorers	Kendal's (W)	Calculated– t	Critical –t
DSISP	16	0.52*	3.57	2.15
RASBSAI	16	0.74*	5.68	2.15
CPSAI	16	0.12	1.08	2.15

* = Significant relationship

Table 5 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 II scored significantly higher than group III.

The implication of the results is that use of ranking pattern indicated superiority over first and third patterns of scoring biology essay test. Similarly, the scoring by session pattern showed superiority over the third pattern of scoring biology essay test. Mean level is a measure of superiority.

Discussion of the Results

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

The 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 ranking pattern, scored by session and scored all items pattern were positive. The magnitudes of their positive correlations coefficients were 0.74* for ranking pattern, 0.52* for scoring into session pattern and 0.12 for conventional pattern of scoring all items respectively. The magnitude of the positive correlation coefficient for RASBSAI, DSISP were high and medium relationship while CPSAI had very low relationship respectively.

The relationship was significant for those group of scorers who scored biology essay test using ranking pattern (RASBSAI) and scoring by session pattern (DSISP). This is because the calculated t value of 5.68 for RASBSAI was much higher than the critical t value (2.15) and the calculated t value of 3.57 for DSISP was more than critical t value of 2.15. Contrary, the calculated t value of 1.08 for CPSAI was lower than critical t value of 2.15. This implied that the use of ranking pattern and scoring by session pattern had significant relationship on scorer reliability in scoring Biology essay test.

Further, analysis in order to determine the mean differences among the various groups of scorers were carried out. From table 4, it was observed that the mean scorers for experimented group I, experimental group II and control group III were 16.84, 20.96 and 11.92 for DSISP, RASBSAI and CPSAI respectively. This means that scorers who scored using ranking pattern (experimental group II, had the highest mean score of 20.96 followed by those who scored by session pattern had 16.84 and the group who scored with the conventional pattern of scoring all items (experimental group III) had the least mean score of 11.92.

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:
Ranking pattern and divide the task of scoring into session has positive significant relationship in scoring biology essay test.

Furthermore, use of ranking pattern was found to be outstandingly more efficacious than

251

INTERNATIONAL JOURNAL OF EDUCATIONAL RESEARCH VOL 11. NO1, 2011. ISSN-978-2568-51-1
CASMIR N. EBUOH (PH D) AND GABRIEL A OKAFOR (PH.D)

scoring by session pattern. The effect of the conventional pattern of scoring all items was not significant on scorer reliability in scoring biology essay test.

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

1. The use of ranking pattern and scoring by session 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 ranking and scoring into session 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.

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*. 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 corporation.

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.

Osisioma, 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 (2008) Vetting sheet on Biology.