

## Towards Ensuring Qualitative Assessment Of Learners: Determination Of Scorer Reliability Of Integrated Science (Its) Achievement Scores.

M. C. Anaekwe, Ph.D<sup>1</sup>, F. S. Apata, Ph.D<sup>2</sup>, Prof. U.S.A.Osuji<sup>3</sup>

<sup>1</sup>Department of Science Education Faculty of Education National Open University of Nigeria

<sup>2</sup>Department of Science Education Faculty of Education National Open University of Nigeria

<sup>3</sup>Department of Educational Foundations Faculty of Education National Open University of Nigeria

---

**Abstract:** Global issues in instructional delivery strive towards ensuring quality of assessment techniques. This study investigated the scorer reliability of ITS achievement scores as measured by assessors from National Teachers Institute (NTI) and National Open University of Nigeria (NOUN) respectively. The study adopted an ex post facto cum correlation research design for the study. A total population of 745 scripts already marked by NTI examiners using the marking guides prepared for the purpose was involved in the study. One hundred and twenty-two (122) scripts which were randomly sampled by NOUN moderators, re-assessed using the same marking guides, constituted the sample for the study. Data were analyzed using SPSS computer program to determine the means, standard deviation and Pearson r correlation coefficient statistics. Results of the study indicated that the ITS achievement scores have high inter-item scorer reliability. Again, there was a significant correlation between the achievement scores as assessed by the two groups of assessors.

**Keywords:** Qualitative assessment, Scorer reliability, Integrated Science, Achievement scores.

---

Date of Submission: 20-12-2019

Date of Acceptance: 03-01-2020

---

### I. Introduction:

Global issues in instructional delivery emphasize a collaborative approach both on the part of the teacher and the learner in ensuring effective learning outcomes. Hence, the techniques of team teaching, peer tutoring, cooperative learning among others, are topical issues. Educational assessment is equally not left out on this collaborative trend hence, peer assessment which emphasizes a situation whereby more than one person or groups of persons are involved in generating the final score of a candidate in a learning task. The guiding philosophy is to ensure that the measured score approximates an objective reflection of the measured trait. This approach strives to minimize subjectivity in the instructional process to the barest minimum. It has great implication for ensuring the quality of teaching, learning and assessment in the classroom.

The justification for this paper stands from the backdrop of determining the reliability of the achievement scores awarded in the Bachelors Degree Programmes (BDPs) of the National Teachers Institute (NTI), which is an affiliate of the National Open University of Nigeria (NOUN). The primary mandate of NTI is the award of Nigeria Certificate in Education (NCE). However, NTI runs a degree programme in affiliation of Faculty of Education, NOUN, in response to societal needs. Towards achieving the goals of this affiliation, senior academics of NOUN over-see the affairs of all BDPs in the areas of review of course materials, monitoring teaching and learning of programmes, moderation of all semester examination questions and moderation of marked answer scripts and examination scores, thereby determining the candidates' achievement on the course. It may therefore, be desirable to determine the quality of instructional delivery to the learners vis a vis the quality of assessment and interpretation of learning outcomes.

Qualitative teaching and learning cannot be said to have taken place without appropriate and adequate evaluation process. The evaluation process could be undertaken through administration of tests, (a set of questions), to measuring the extent to which the learners have mastered the content(s) to which they have been exposed. The result of such tests/examinations serves a number of useful purposes including grading, selection, certification, employment, placement, etc. Therefore, great care should be taken in preparing questions or testing items to ensure that the questions meet the expected standard and possesses appropriate psychometric qualities. We can therefore look at assessment or examination question as "a set of question administered to the learner for the purpose of determining the extent they have attained previously identified or exposed objectives or contents (Iwuji, 1990).

From the foregoing, it is important to note that every examination question must: be tied to a previously taught content area, be consistent with the identified objectives of the course, consists of a standard set of questions that will yield valid and reliable learning outcomes, yields a quantitative value (score) on the

basis of which we can infer the presence or absence of the measured attribute or construct. For a test score to be able to effectively serve as a basis for inferring the presence or absence of a measured attribute, the instrument *ab initio*, must be reliable. While validity of a test refers to a test that measures what it is supposed to measure, reliability refers to the degree of consistency with which the test measures what it is supposed to measure. It may therefore, need to be pointed out that an instrument may be valid without being reliable. Therefore, reliability of an instrument is of greater importance in assessing students' learning outcomes.

Standard objective tests usually have perfect scorer reliability if there are no mistakes unlike the essay tests which are fraught with shortcomings bordering on standard of grading. When tests are not scored objectively, as is usually the case with essay tests, scorer reliability becomes an issue of concern. According to Mehrens and Lehmann cited in Mkpa (1992), in grading essay responses, one must use appropriate methods to minimize biases, pay attention only to significant and relevant aspects of the answer and apply uniform standards in all the papers. The uniformity of grading standards which ensures the reliability of the scores is the most crucial aspect of essay grading. When uniformity is lacking, there will be no valid way of comparing students' achievement. This underscores the need to investigating the scorer reliability of essay test. By scoring samples of essay papers independently by two or more different scorers and correlating the scores so obtained, the scorer reliability coefficient of the essay test may be estimated and interpreted. This is the approach adopted in this study. The ITS achievement scores assessed by NTI assessors were correlated with the moderated scores assessed by NOUN moderators.

Students' achievement in a task is a measure of the quality and quantity of knowledge, skills, techniques, positive attitudes and behavior that students achieve or acquire in a course of study. Achievement according to Gronlund (1979), is evaluated by mark or grade that students attain in a term or education cycle. It is quantified as a measure of the student's academic standing in relation to those of other students of his/her age. This implies that academic achievement is a result oriented construct that includes the extent of performance of a desired task. The students' achievement in ITS was determined and analyzed for this study. The researchers, being part of the moderating team for the BDPs, deemed it necessary to determine the scorer-reliability of ITS achievement scores by comparing the set of scores awarded by NTI examiners and NOUN moderators.

#### **Statement of the Problem:**

Students' achievement is a measure of the quality and quantity of knowledge, skills and attitudinal disposition of those students towards the measured course of study. They therefore, ought to be objectively determined. When such quantitative values are compromised by way of shortcomings arising from preparation of the test instrument, the decisions made, based on such tests becomes misleading. This underscores the need to ensure that test instruments are of appropriate standard and possesses desired psychometric qualities. Reliability of a test refers to the degree of consistency with which the test measures what it is supposed to measure. Reliability index of a test is one of such qualities which if found wanting, would obviate the essence of the test. It is for this purpose that great care should be taken in ensuring that tests are of appropriate reliability standard. The problem of this study therefore, was to determine the scorer reliability of ITS achievement scores as assessed by NTI and NOUN assessors respectively.

#### **Purpose of the Study:**

The purpose of the study was to determine the scorer reliability of ITS achievement scores. Specifically, this study sought to determine:

- i. The mean achievement score of ITS students as assessed by NTI examiners
- ii. The mean achievement score of ITS students as assessed by NOUN moderators
- iii. The reliability coefficient of ITS achievement scores as assessed by both team of assessors.

#### **Research Questions:**

The following research questions guided the study:

- i. What is the mean achievement score of ITS students as assessed by NTI examiners?
- ii. What is the mean achievement score of ITS students as assessed by NOUN moderators?
- iii. What is the reliability coefficient of ITS achievement scores as assessed by both team of assessors?
- iv. How reliable are the students' achievement scores on integrated science?

#### **Hypothesis:**

There is no significant correlation between ITS achievement scores as assessed by NTI and NOUN assessors.

## II. Methodology:

This study adopted a combination of ex post facto and correlation designs in this research. By determining the outcome of students' achievement on ITS without manipulating any independent variable, ex post facto design was implied. On the other hand, by comparing two sets of achievement scores on ITS as determined by NTI examiners and NOUN moderators respectively, to determine their extent of relationship, a correlation design was implied. The population of the study consisted of seven hundred and forty five –five (745), ITS candidates of the NTI in the 2018\_1 semester examination. The rule of thumb for moderation of examination scripts required the moderators to sample 10% of the scripts for the exercise. However, a total of 122 scripts representing 122 candidates were sampled and used for the study, as illustrated in the table 1.

**Table 1:** Population and Sample for the study.

S/N	Course Codes	Course Titles	Population	Sample
1	CHM201	Physical chemistry	110	15
2	ESM211	Global environmental laws	60	8
3	BIO205	Introductory development of Cell Biology	71	11
4	ECE231	Science in early years	130	22
5	CHM205	Inorganic chemistry	130	22
6	CHM314	Environmental Chemistry	107	19
7	ESM343	Climate Change and the Environment	96	14
8	ESM303	Environmental laws and policies	71	11
	<b>Total</b>		<b>745</b>	<b>122</b>

The test instruments prior to administration to the candidates underwent all the statutory processes of ensuring academic standard of tests including internal and external validation and were therefore, adjudged to be valid and reliable for the purpose. Data analysis was done SPSS computer program to determine the means, standard deviation and Pearson r correlation coefficient.

## III. Results:

Results of the study were presented in line with the research questions and hypotheses.

### Research Questions:

- What is the mean achievement score of ITS students as assessed by NTI examiners?
- What is the mean achievement score of ITS students as assessed by NOUN moderators?
- What is the reliability coefficient of ITS achievement scores as assessed by both team of assessors?
- How reliable are the students' achievement scores on integrated science.

**Table 2:** Means and Reliability Coefficients of ITS Achievement Scores.

Groups	N	Mean	S.D	Pearson r
NTI Assessors (VAR00002)	122	50.66	20.05	0.984
NOUN Assessors (VAR00003)	122	49.91	19.41	0.984

In response to research questions 1 and 2, as shown in table 2, the mean achievement scores of ITS students as assessed by NTI assessors was 50.66, while that for NOUN assessors was 49.91. The same table also shows a value of 0.984 as the inter-item reliability coefficient of the ITS scores in response to research question 3. In response to research question 4, the reliability coefficient indicates that the achievement scores are highly reliable (98%). In other words, there was a high degree of positive relationship between the achievement scores as measures by the assessors.

### Hypotheses:

- There is no significant correlation between ITS achievement scores as assessed by NTI and NOUN assessors.

**Table 3:** Pearson Correlation Coefficient for ITS Achievement Scores.

		Correlations	
		NTI VAR00002	NOUN VAR00003
NTI VAR00002	Pearson Correlation	1	.984**
	Sig. (2-tailed)		.000
	N	122	122
NOUN VAR00003	Pearson Correlation	.984**	1
	Sig. (2-tailed)	.000	
	N	122	122

\*\* . Correlation is significant at the 0.01 level (2-tailed).

From table 3, the Pearson correlation coefficient value of 0.984 was shown as a measure of the relationship for the two sets of scores of the achievement scores. This implies that the achievement scores are highly correlated (98%). In other words, there was a high degree of positive correlation between the achievement scores as measured by the assessors.

#### **IV. Discussion:**

Table 3 shows a scorer reliability coefficient of 0.984. This value was significant at 0.01 level of significance for a 2-tailed test. This indicates that the NTI assessors are consistent in the scores they award and one can readily rely on the ITS achievement scores. Therefore the null hypothesis of no significant correlation is rejected. This result is encouraging as it is a pointer to the fact that the NTI assessors are dependable in making reliable decisions with respect to award of achievement scores on ITS. Furthermore, the result has buttressed the fact that NTI –NOUN assessment collaboration in the award of BDPs, is in a way, meeting up with the standards expected by the National Goals of Assessment (FRN,2013) which includes :

- i. Ensuring that students' abilities are accurately measured;
- ii. Improving the credibility of examinations conducted in Nigeria;
- iii. Enhancing the global competitiveness of the products of the Nigerian educational system.

#### **Implications:**

Generally, the successful implementation of all educational programmes is the major responsibility of teachers at the classroom level. Educational programmes being direct offshoots of national goals and aspirations are accomplished through the teachers' professional expertise and dedication to duty. He therefore, has to provide the learners with the most effective instruction and the responsibility of assessing them with valid and reliable measuring instruments.

Granted that the result of this study indicated that the ITS achievement tests were of high inter-item scorer reliability, conscious efforts should be made to strengthen the capacity of classroom teachers at all levels of Nigerian educational system towards efficiency in constructing reliable test items. This way, the learners' abilities could be more accurately measured and quality of products from Nigerian educational system assured.

#### **V. Conclusion and Recommendation:**

Effective teaching and learning cannot be said to have taken place without ensuring the quality of the evaluation process. The evaluation process could be undertaken through administration of achievement tests to measuring the extent to which the learners have mastered the content(s) to which they have been exposed. Students' achievement is a measure of the quality and quantity of knowledge, skills and attitudinal disposition of those students towards the measured course of study. They therefore, ought to be objectively determined. When such quantitative values are compromised by way of shortcomings arising from preparation of the test instrument, the decisions made, based on such tests becomes misleading.

This underscores the need to ensure that test instruments are reliable. Reliability of a test refers to the degree of consistency with which the test measures what it is supposed to measure. This study therefore, sought to determine the scorer reliability of ITS achievement scores as measured by NTI and NOUN assessors respectively. The results of the study indicated among others, that the ITS achievement scores have high inter-item scorer reliability. It is therefore, recommended that conscious efforts should be made to strengthen the capacity of classroom teachers at all levels of Nigerian educational system towards efficiency in constructing reliable test items through regular workshops and seminars on test item construction and administration.

#### **References:**

- [1]. Federal Republic of Nigeria (FRN, 2013). National Policy on Education. Lagos:NERDC
- [2]. Gronlund, I.E. (1979). Measurement and Evaluation in Teaching. London:
- [3]. Macmillan publishers.
- [4]. Mkpa, N.D. (1992). Scorer Reliability of Teacher-Made Geography Test Questions Used in Imo State Secondary Schools. In Alaezi, O., Obioma, G. and Akpan, E.U.U.(Eds.). Curriculum Development and Implementation Processes. Owerri: International Universities Press Ltd.
- [6]. Iwuji, V.B.C. (1990). Measurement and Evaluation for Effective Teaching and Learning. Onitsha: Summer Educational Publishers
- [7].

M. C. Anaekwe, Ph.D. " Towards Ensuring Qualitative Assessment Of Learners: Determination Of Scorer Reliability Of Integrated Science (Its) Achievement Scores." *IOSR Journal of Research & Method in Education (IOSR-JRME)*, 10(1), 2020, pp. 01-04.