
CHALLENGES OF ASSESSING STUDENTS IN MEASUREMENT AND EVALUATION SPHERE

DR. FELICIA KING-AGBOTO
Department Of Physical Education
Imo State University, Owerri-Nigeria

DR.DANIEL ORIFAMAH
Department Of Physical Education
Imo State University, Owerri-Nigeria

And

DR. CHIZOMA CATHERINE OKPARA
Department Of Education Psychology,
Imo State University, Owerri-Nigeria

Abstract

The main purpose of this study is to determine the challenges of assessing students in measurement and evaluation sphere. The study adopted descriptive survey research design. A purposive sampling technique was used to select lecturers teaching courses in measurement and evaluation in Rivers State University. The population of the study constituted of 200 universities lecturers of which 82 were randomly selected and used as sample for the study. A researcher designed 40-itemed, 4 point modified likert type questionnaire titled: Challenges of assessing students in Measurement & Evaluation Sphere Questionnaire (CASMES) was used to obtain data for the study. Respondents indicated their levels of agreement to the items by ticking one of the response options of strongly agree (SA) 4 points, Agree (A) 3 points, Disagree (D) 2 points and strongly Disagree (SD) 1 point for each of the items. Criterion mean of 2.5 was set for decision. That is item and cluster mean of 2.5 and above indicate agreement while below 2.5 indicate disagreement. The instrument was given content and face validation by two experts in measurement and evaluation. The stability of the instrument was established through Cronbach Alpha. This yielded an acceptable value of .83. The resulting data were analyzed using frequency counts and mean rating scores. The resulting data were analyzed using mean and standard deviation. According to the findings, it showed that the lack of time devoted in scoring, marking scheme of work, recording and construction of test blue print affects assessment.

Challenges of Assessing Students in Measurement and Evaluation Sphere - Dr. Felicia King-Agboto; Dr. Daniel Orifamah and Dr. Chizoma Catherine Okpara

Keywords: challenges Assessment, Scoring, Recording, Marking Scheme Construction of Blue print

Introduction

Recently, education has become complex and competitive. Evaluation in recent world has to be completely understood the different way of current challenging. The basic aim of education is the transformation of human disposition into the right arrangement by using a holistic process of body process, mind and sublimation of the emotions and the illumination of the spirit (Nduwayezu, 2020). The term evaluation may be completely wrong if you start understanding and analyzing modern students the traditional way. The paradigm shifts that are happening in today's world of education, clearly suggest that the way the students needs to be evaluated should be completely dynamic and well planned.

Kirkendall, Gruber & Johnson in Inko-Tariah(2004) sees evaluation as the process of determining the value or worth of collected data while Gronlund in Inko-Tariah sees evaluation as the systematic process of collecting, analyzing and interpreting information to determine the extent to which instructional objectives have been achieved. In order words it is a systematic process of passing value judgment base on the result test, measurement and assessment. Thus when a classroom teacher makes remarks like promoted to repeat etc he is passing value judgment and it is very important that such judgment are passed with minimum risks .Baykul & Yildrin in Murat & Nurdan(2015)indicates that evaluation on the other hand means making a decision by comparing results of the measurement to a criteria or a set of standard. Pandey & Smith (2001) state that in essence aims of evaluation are to direct students to investigating, to determine at what levels student understand mathematics, to determine their level of ability of using the processes, to find out development of the students in the process of learning, to improve abilities of the students in considering all possibilities providing meaningful answers in answering questions and to provide them the opportunity of using mathematics at so high levels that they can organize complex events.

Measurement

Moreover, Turgut in Murat & Nurdan(2015) states that measurement in our education means “observing a feature and presenting the results of it by using numbers and symbols while Tekin in Murat & Nurdan defines measurement as to observe whether a quality has any feature or not and if yes, investigating to what extent the feature is observation with symbols. Specifically, with numbers as both of the definitions are the most widely used and adopted definitions in the educational system. Measurement in education is simply means determining the attributes or dimensions of an object, skill or knowledge. Some standard measurements in education are raw

scores, percentile, ranks and standard scores. Moreso, information gotten with measurement and evaluation is kept and recorded as data.

Tam & Erdogan in Murat & Nurdan (2015) states that in this sense, measuring and evaluating determines lack and insufficient aspects of a teaching program, improves education and education and teaching process, provides information in the process of program development and provides guidance. At the end of the measurement and evaluation process, students are said to be successful or unsuccessful.

Assessment

Assessment is broader than measurement; therefore it can describe the worth of pupil's behavior, but fails to pass judgment. Assessment has been described as a process of obtaining information for decision making. These assessments according to UNESCO in Ukozor & Amadi (2015) are school based assessments which are derived by teachers based on content they have actually exposed their students to. They are more of criterion referenced and tend to be tied to what was actually taught by teachers. The focus of assessment in most cases is mainly academic and they are important in determining opportunities. McCormick & James in Inko-Tariah (2004) pointed out a common concern of assessment as ranking of individuals within a classroom but evaluation is concerned with what a pupil or class understands for improvement of teaching. Assessment according to them is restricted to classroom situations while evaluation has to do with what happens both inside and outside the classroom. In assessment, teachers gather information by giving tests, conducting interviews and monitoring behavior. A good assessment should be carefully prepared and administered to ensure its reliability and validity. It must provide consistent results and must measure what it ought to measure.

Kpolovie (2002) sees test as the presentation of a standard set of questions to be answered and which qualify as relevant information gathering instrument for effective evaluation of the examinee's cognitive, affective and or psychomotor traits. Test is a task or series of tasks used to obtain systematic observations presumed to be representative of educational or psychological traits or attributes (Sax in Inko-Tariah 2004). To Onunkwo (2005) a test is an instrument which can be utilized in detecting some qualities, traits, characteristics, attributes etc possessed by a person, an object or a thing. Furthermore, it is any task presented to an individual with the aim of obtaining a sample of his behavior.

Scoring

The first step after examination or testing is scoring. Scoring is the process of measurement Ukozor & Amadi (2015). Scoring demands commitment and thoroughness on the part of the teacher because of it is very sensitive and tasking nature. Test scoring is a process of marking or scoring or awarding scores/marks to the responses of the testees after the administration of the test. The primary function of scores or marks is to communicate to the testees and or other interested members of the

Challenges of Assessing Students in Measurement and Evaluation Sphere - Dr. Felicia King-Agboto; Dr. Daniel Orifamah and Dr. Chizoma Catherine Okpara

public, the testees degree of achievement, performance or competence. Because of their individual peculiarities, essay and objective tests have different scoring methods. Essay test are scored either analytically or globally (wholistically). Annalistic Scoring points on a marking scheme, the scorer then gives marks for organization of work, command of English, neatness of work, expression etc. Global scoring involves the scorer reading the answer, comparing it to what is in the textbook and then awards his mark. Objective tests are scored with marking schemes; big exam bodies like WAEC, use scoring machines, scoring should be within two weeks after the tests are taken (Inko-Tari, 2004).

Scoring model is a procedure of focusing on aspects of learner's response and allocates them to categories according to the level of proficiency. A model of scoring can take the form "Rubric " (Goldstein,2017) yet test scores are tools used to assign categories of learners and to judge what they can do or not, it is not the best way to judge the knowledge and skills of learners because test can only measure a sample of needed skills in a subject area.

Marking Scheme

According to Esomonu (2006) marking scheme should contain all the major points sought for by the test items since it is a model answer or representation of perfect answers to questions. It should specify how many marks are to be awarded to each point made in a question and also specify other factors that would be taken into consideration while scoring. For example, whether spelling and handwriting would be penalized, whether articulate discussions or good general presentation of the facts, accuracy of factual information, skill of organization, clarity of thought would attract extra marks. The total score in a marking scheme should therefore tally with the total obtainable score. In order words, marking schemes should clearly indicates: instructional about the award of marks, specific marks for procedure, accuracy, etc, minimum number of point required to give maximum marks and penalties for wrong spelling, no diagram etc if need be. Also, in order to make a good marking scheme, teacher has to be familiar with how to make assessment rubrics. The use of rubrics in marking scheme guide help in assigning marks to the questions based on the answers required from the questions and the time needed to do each given question. (Ander & Gunille, 2007) Marking scheme is prepared along with the test, which is used in scoring the testees responses' marking scheme is an outline of the main points which should be contained in a good response, and the marks which are allocated to each of the point's .It helps to increase the scorer reliability of the test, thus reducing the subjectivity. The WAEC and other public examining bodies that administer essay questions use this method. It is also used in internal exams. Marking scheme should contain all the major points sought for by the test items since it is a model answer or representation of perfect answers to questions .Esomonu (2005)

Recording

Recording is part of continuous assessment, so there is need to keep and maintain accurate record on every student in school. Since continuous assessment is cumulative and transferable, it is also implied that the records must be cumulative and transferable. Therefore there is need to ensure that the same types of records are kept in all schools in the country. Unfortunately this needed uniformity in types of records kept is lacking, what practically operates in a situation where each school keeps different types of records as the school deems right.

Assessment and grading

Grading in education is the process of applying standardized measurements for varying levels of achievements in a course. Grades can be assigned as letters (usually A through F), as a range (for example, 1 to 6), as a percentage, or as a number out of a possible total (often out of 100).Salvo (2009). Grades are averaged to create a grade point average (GPA). GPA is calculated by using the number of grade points a student earns in a given period of time. GPAs are often calculated for high school, undergraduate, and graduate students, and can be used by potential employers or educational institutions to assess and compare applicants. A cumulative grade point average (CGPA), sometimes referred to as just GPA, is a measure of performance for all of a student's courses.

Construction of test blue print

Inko-Tariah (2004) sees the blue print as a two-dimensional grid showing the content area covered on the row and the cognitive levels to be tested on the column. The grid shows how many questions should come from each topic considering the duration of teaching for each topic. Construction of test blue print otherwise known as table of specifications is a two –fold test content which clearly specifies the different areas of subject matter and types of behavior (ability) to be elicited by each area. It indicates the relative emphasis of the various topics and types of behavior to be treated. A table of specification defines in the most precise manner the scope and emphasis of a test in terms of the various topics of subject matter and the different levels of taxonomy of educational objective. According to Chikwe (2017) steps involved in preparing table of specification are (1) determine the total number of weeks used in teaching all the content areas (2) compute the number of items each content area would contribute to the 196 items needed.

(a)Basic concepts	$2/12 \times 100 = 17$ items
(b)Qualities of test	$3/12 \times 100 = 25$ items
(c) Test construction	$1/12 \times 100 = 8$ items
(d) Test Administration	$2/12 \times 100 = 17$ items
(e) Continuous Assessment	$1/12 \times 100 = 8$ items
(f) Non cognitive assessment	$3/12 \times 100 = 25$ items
Total	100 items

Challenges of Assessing Students in Measurement and Evaluation Sphere - Dr. Felicia King-Agboto; Dr. Daniel Orifamah and Dr. Chizoma Catherine Okpara

Table showing 1 test Blue print for 100 items of measurement and evaluation								
No of weeks	CONTENT AREAS	KNOWLEDGE	COMPREHENSION	APPLICATION	ANALYSIS	SYNTHESIS	EVALUATION	TOTAL
		25%	25%	20%	15%	10%	5%	100%
2	Basic concepts	4	4	3	3	2	1	17
3	Qualities of test	6	6	5	4	3	1	25
1	Test construction	2	2	2	1	1	0	8
2	Test Administration	4	4	3	3	2	1	17
1	Continuous Assessment	2	2	2	1	1	0	8
3	Non cognitive assessment	6	6	5	4	3	1	25
12	Total	24	24	20	16	12	4	100

No. of Weeks

Total No. of weeks x Total No. of questions

(iii) Compute the number of item each cognitive level would contribute to each content area bearing in mind respective percentage emphasis.

For Basic Concepts

Knowledge level = $25/100 \times 17 = 4$

Comprehension = $25/100 \times 17 = 4$

Application = $20/100 \times 17 = 3$

Analysis = $15/100 \times 17 = 3$

Synthesis = $10/100 \times 17 = 2$

Evaluation = $5/100 \times 17 = 1$

Qualities of test

Knowledge level = $25/100 \times 25 = 6$

Comprehension = $25/100 \times 25 = 6$

Application = $20/100 \times 25 = 5$

Analysis = $15/100 \times 25 = 4$

Synthesis = $10/100 \times 25 = 3$

Evaluation = $5/100 \times 25 = 1$

Test construction

Knowledge level = $25/100 \times 35 = 9$

Comprehension = $25/100 \times 35 = 9$

Application	$=20/100 \times 35 =7$
Analysis	$=15/100 \times 35 =5$
Synthesis	$=10/100 \times 35 =4$
Evaluation	$=5/100 \times 35 =2$

Test Administration

Knowledge level	$=25/100 \times 17 =4$
Comprehension	$=25/100 \times 17 =4$
Application	$=20/100 \times 17 =3$
Analysis	$=15/100 \times 17 =3$
Synthesis	$=10/100 \times 17 =2$
Evaluation	$=5/100 \times 17 = 1$

Continuous Assessment

Knowledge level	$=25/100 \times 8 = 2$
Comprehension	$=25/100 \times 8 = 2$
Application	$=20/100 \times 8 =2$
Analysis	$=15/100 \times 8 =1$
Synthesis	$=10/100 \times 8 =1$
Evaluation	$=5/100 \times 8 = 0$

Non cognitive assessment

Knowledge level	$=25/100 \times 25 =6$
Comprehension	$=25/100 \times 25 =6$
Application	$=20/100 \times 25 =5$
Analysis	$=15/100 \times 25 =4$
Synthesis	$=10/100 \times 25 =3$
Evaluation	$=5/100 \times 25 = 1$

Statement of the problem

It been observed that large classroom have made teachers to faced with a lot of scripts to mark and score after each examination, particularly in secondary and tertiary levels, as a result of this problem, teachers scores a question by only focusing on the answer that is related to theme and also they asked difficult questions. Moreover they lack time to carry out their responsibility.

Research Questions

1. What are the problems of scoring in assessment?
2. What are the problems of recording in assessment?
3. What are the problems of marking scheme in assessment?
4. What are the problems of construction of test blue print in assessment?

Challenges of Assessing Students in Measurement and Evaluation Sphere - Dr. Felicia King Agboto; Dr. Daniel Orifamah and Dr. Chizoma Catherine Okpara

The study adopted descriptive survey research design. A purposive sampling technique was used to select lecturers teaching courses in measurement and evaluation in Rivers State University. The population of the study constituted of 200 universities lecturers of which 82 were randomly selected and used as sample for the study. A researcher designed 40-itemed, 4 point modified likert type questionnaire titled: Challenges of assessing students in Measurement & Evaluation Sphere in Assessment Questionnaire (PMESA) was used to obtain data for the study. Respondents indicated their levels of agreement to the items by ticking one of the response options of strongly agree (SA) 4 points, Agree (A) 3 points, Disagree (D) 2 points and strongly Disagree (SD) 1 point for each of the items. Criterion mean of 2.5 was set for decision. That is item and cluster mean of 2.5 and above indicate agreement while below 2.5 indicate disagreement. The instrument was given content and face validation by two experts in measurement and evaluation. The stability of the instrument was established through Cronbach Alpha. This yielded an acceptable value of .83. The resulting data were analyzed using frequency counts and mean rating scores.

Results

Research Question 1: What are the problems of scoring in assessment?

Table 2: Mean Responses in Respect of the Problems of Scoring in Assessment

S/ N	Items	SA	A	D	SD	TO TA L	ME AN	DECISION
1	No record of the affective	30(120)	30(90)	10(20)	12(12)	82(242)	2.95	Agreed
2	No record of the psychomotor	30(120)	20(60)	12(24)	20(20)	82(224)	2.73	Agreed
3	Having errors in questions and providing no bonus question.	40(160)	30(90)	6(12)	6(6)	82(268)	3.27	Agreed
4	Cognitive domain is not kept meticulously as it should be.	240(16)	10(30)	10(20)	2(2)	82(292)	3.56	Agreed
5	Scoring students scripts haphazardly.	70(280)	5(15)	5(10)	2(2)	82(307)	3.74	Agreed

6	Lack of knowledge and skill for good marking	50(20)	10(30)	10(20)	12(12)	82(262)	3.20	Agreed
7	Teachers lack time to in terms of scoring	52(208)	10(30)	10(20)	2(2)	82(292)	3.27	Agreed
8	Many lectures give their scripts to their students to score	60(240)	10(30)	10(20)	2(2)	82(292)	3.56	Agreed
9	Lack of details in score allotment	40(160)	30(90)	5(10)	7(7)	82(267)	3.26	Agreed
10	Scoring essay test script by script instead of one number across the script	30(120)	30(90)	20(40)	2(2)	82(252)	3.07	Agreed
Grand mean							3.28	Agreed

SA =Strongly Agree

A= Agree

D= Disagree

SD= Strongly Disagree

Criterion Mean=2.50

The data presented in table 2 showed that the mean ratings of the responses of scoring in assessment on the 10 items in the table ranging from 2.73 to 3.74 which are all greater than the cut-off point value of 2.50 on 4-point rating scale. This indicated that the 10 identified items in the table are in agreement with the statement about the problems of scoring in assessment. The conclusion is that problems encountered during scoring affect assessment.

Research Question 2: What are the problems of recording in assessment?

Table 3: Mean Responses in Respect of the Problems of Recording in Assessment

S/N	Items	SA	A	D	SD	TOTAL	MEAN	DECISION
11	No main point	30(120)	30(90)	12(24)	10(10)	82(244)	2.98	Agreed
12	Marks are not properly allocated	50(200)	20(60)	10(20)	2(2)	82(282)	3.44	Agreed
13	No marking scheme	70(280)	5(15)	5(10)	2(2)	82(307)	3.74	Agreed
14	Lack of concentration	60(240)	10(30)	20(40)	2(2)	82(312)	3.80	Agreed

Challenges of Assessing Students in Measurement and Evaluation Sphere - Dr. Felicia King-Agboto; Dr. Daniel Orifamah and Dr. Chizoma Catherine Okpara

15	It is slow and laborious	50(200)	12(36)	10(20)	10(10)	82(266)	3.24	Agreed
16	Improper record keeping	60(240)	5(15)	5(10)	12(12)	82(277)	3.38	Agreed
17	Lack of time	40(160)	20(60)	20(40)	2(2)	82(262)	3.20	Agreed
18	Some lecturers are not meticulous in recording	30(120)	30(90)	12(24)	10(10)	82(244)	2.98	Agreed
19	Carefree attitude of the lecturers	50(200)	10(30)	20(40)	2(2)	82(272)	3.32	Agreed
20		62(248)	10(30)	7(14)	3(3)	82(295)	3.60	Agreed
Grand mean				3.37		Agreed		

SA =Strongly Agree

A= Agree

D= Disagree

SD= Strongly Disagree

Criterion Mean=2.50

The data presented in table 23 showed that the mean ratings of the responses of recording in assessment on the 10 items in the table ranging from 2.98 to 3.80 which are all greater than the cut-off point value of 2.50 on 4-point rating scale. This indicated that the 10 identified items in the table are in agreement with the statement about the problems of recording in assessment. The conclusion is that problems encountered during recording affect assessment.

Research Question 3: What are the problems of marking scheme in assessment?

Table 4: Mean Responses in Respect of the Problems of Marking Scheme in Assessment

	Items	SA	A	D	SD	TOTAL	MEAN	DECISION
21	Late preparation of marking scheme	40(160)	20(60)	12(24)	10(10)	82(234)	2.85	Agreed
22	Non familiarization with the marking	25(100)	35(105)	20(40)	2(2)	82(247)	3.01	Agreed
23	Carefree attitude of the teacher	30(120)	30(90)	10(20)	12(12)	82(242)	2.95	Agreed
24	Distractions of the teacher by his/her private business	52(208)	20(60)	5(10)	5(5)	82(283)	3.45	Agreed
25	Lack of knowledge and skill for good marking	40(160)	22(66)	10(20)	10(10)	82(256)	3.12	Agreed
26	There is no standardized marking scheme.	60(240)	5(15)	5(10)	12(12)	82(277)	3.38	Agreed

27	Non familiarization with the marking scheme before marking	62(248)	10(30)	7(14)	3(3)	82(295)	3.60	Agreed
28	Improper allocation of marks	50(200)	10(30)	20(40)	2(2)	82(272)	3.32	Agreed
29	Large classes is a major problem	30(120)	30(90)	12(24)	10(10)	82(244)	2.98	Agreed
30	Laziness on the part of the teacher.	50(200)	20(60)	10(20)	2(2)	82(282)	3.44	Agreed
Grand mean					3.21 Agreed			

SA =Strongly Agree

A= Agree

D= Disagree

SD= Strongly Disagree

Criterion Mean=2.50

The data presented in table 4 showed that the mean ratings of the responses of marking scheme in assessment on the 10 items in the table ranging from 2.85 to 3.60 which are all greater than the cut-off point value of 2.50 on 4-point rating scale. This indicated that the 10 identified items in the table are in agreement with the statement about the problems of marking scheme in assessment. The conclusion is that problems encountered during marking scheme affect assessment.

Research Question 4: What are the problems of construction of test blue print in assessment?

Table 5: Mean Responses in Respect of the Problems of Construction of test blue print in Assessment

S/ N	Items	SA	A	D	SD	TOTAL	MEAN	DECISION
31	I do not have the knowledge about the blue print	40(1160)	30(90)	5(10)	7(7)	82(267)	3.26	Agreed
32	Construction of test blue print is a waste of time.	25(100)	35(105)	20(40)	2(2)	82(247)	3.01	Agreed
33	Some teachers do not cover all levels of cognitive and content areas.	40(160)	22(66)	10(20)	10(10)	82(256)	3.12	Agreed
34	I find it difficult to allot the appropriate number of items to the blue print	30(120)	25(1)	25(50)	2(2)	82(247)	3.01	Agreed

Challenges of Assessing Students in Measurement and valuation Sphere - Dr. Felicia King-Agboto; Dr. Daniel Orifamah and Dr. Chizoma Catherine Okpara

35	I need extraordinary teaching in order to compute the behavioral objective.	52(208)	20(60)	5(10)	5(5)	82(283)	3.45	Agreed
36	Teachers are not acquainted with the method of constructing blue print test.	17(62)	35(105)	20(40)	10(10)	82(217)	2.64	Agreed
37	Most lecturers do not trial testing during test construction.	30(120)	30(90)	20(40)	2(20)	82(252)	3.07	
38	Some test constructors does not consider the age and educational level of the students	20(80)	30(90)	20(40)	12(12)	82(222)	2.71	Agreed
39	Some test constructor does not write more number of items in case the constructed items do not stand further scrutiny/validity process	25(100)	25(75)	20(50)	12(12)	8(237))2	2.89	Agreed
40	Some of them do not know that negatively worded items should be discarded.	30(120)	20(60)	12(24)	20(20)	82(224)	2.73	Agreed
Grand mean							2.99	Agreed

SA =Strongly Agree

A= Agree

D= Disagree

SD= Strongly Disagree

Criterion Mean=2.50

The data presented in table 5 showed that the mean ratings of the responses of construction of test blue print in assessment on the 10 items in the table ranging from 2.64 to 3.45 which are all greater than the cut-off point value of 2.50 on 4-point rating scale. This indicated that the 10 identified items in the table are in agreement with the statement about the problems of construction of test blue print in assessment. It also shows the distribution of the problems of construction of test blue print in assessment in terms of the levels of cognitive and content areas. They cover tasks on cognitive levels of understanding, application and analysis but do not assess students that task their ability to synthesize and evaluate. The conclusion is that since the distribution of items to be objectives and content areas did not cover all, no enough evidence that test blue print has a high content validity. However, good constructed table of specification will help to improve the validity of teacher evaluation based on a given assessment.

The quality of evidence gathered from students can be trusted due to quality or coherence of table of specification.

Hypothesis

Hypothesis One: There is no significance difference between male and female lecturers in scoring assessment

Data for testing Hypothesis One are presented in Table 6 below.

Table 6: Test of Significance Difference in the mean ratings of male and female lecturers in scoring assessment.

VARIABLE S	N	MEAN	SD	DF	T-CAL	T-CRIT	DECISION	SIGNIFICANCE
Male Lecturers	44	10.16	4.82	80	0.63	1.98	Agreed	0.05
Female Lecturers	38	10.74	3.24					

The results on the significance in Table 6 showed that the calculated t-value 0.63 is less than the critical t-value (t-tab) value of 1.96 at 80 degrees of freedom. This implies that there is no significance difference between the opinion of male lecturers and female lecturers in scoring assessment at 0.05. Therefore, the null hypothesis of no significant ($p \leq 0.05$) difference in the mean ratings of the responses of male and female lecturers is accepted for hypothesis one.

Hypothesis Two: There is no significance difference between measurement & evaluation lecturers and curriculum lecturers in terms of marking scheme in assessment Data for testing Hypothesis two are presented in Table 7 below.

Table 7: Test of Significance Difference in the mean ratings of measurement and evaluation lecturers and Curriculum lecturers in terms of marking scheme in assessment.

VARIABLES	N	MEAN	SD	DF	T-CAL	T-CRIT	DECISION	SIGNIFICANCE
Measurement & Evaluation Lecturers	41	8.07	3.42	80	0.55	1.98	Agreed	0.05
Curriculum studies Lecturers	41	7.66	3.37					

The results on the significance in Table 7 showed that the calculated t-value 0.55 is less than the critical t-value (t-tab) value of 1.96 at 80 degrees of freedom. The means data shows that measurement and evaluation lecturers are more experienced in marking scheme than their colleagues in curriculum studies department. This implies

Challenges of Assessing Students in Measurement and Evaluation Sphere - Dr. Felicia King-Agboto; Dr. Daniel Orifamah and Dr. Chizoma Catherine Okpara

that there is no significance difference between the opinion of measurement & evaluation lecturers and curriculum studies lecturers in marking scheme assessment at 0.05. Therefore, the null hypothesis of no significant ($p \leq 0.05$) difference in the mean ratings of the responses of measurement & evaluation and curriculum studies lecturers is accepted for hypothesis two.

Hypothesis Three: There is no significance difference between educational management lecturers and Adult education lecturers in recording in assessment Data for testing Hypothesis three are presented in Table 8 below.

Table 8: Test of Significance Difference in the mean ratings of educational management lecturers and adult education lecturers in recording in assessment.

VARIABLES	N	MEAN	SD	DF	T-CAL	T-CRIT	DECISION	SIGNIFICANCE
Educational Management Lecturers	41	9.83	3.56	80	0.71	1.98	Agreed	0.05
Adult Education Lecturers	41	9.24	3.87					

The results on the significance in Table 8 showed that the calculated t-value 0.71 is less than the critical t-value (t-tab) value of 1.96 at 80 degrees of freedom. The means data shows that educational management lecturers are more experienced in recording than their colleagues in adult education department. This implies that there is no significance difference between the opinion of educational management lecturers and adult education lecturers in recording assessment at 0.05. Therefore, the null hypothesis of no significant ($p \leq 0.05$) difference in the mean ratings of the responses of educational management and adult education lecturers is accepted for hypothesis three.

Hypothesis Four: There is no significance difference between male and female lecturers in construction of blue print in assessment Data for testing Hypothesis four are presented in Table 9 below.

Table 9: Test of Significance Difference in the mean ratings of male and female lecturers in construction of blue print in assessment.

VARIABLES	N	MEAN	SD	DF	T-CAL	T-CRIT	DECISION	SIGNIFICANCE
Male Lecturers	42	11.12	2.65	80	0.25	1.98	Agreed	0.05
Female Lecturers	40	11.03	2.79					

The results on the significance in Table 9 showed that the calculated t-value 0.25 is less than the critical t-value (t-tab) value of 1.96 at 80 degrees of freedom. The

mean value shows that male lecturers are good in test construction than female counterparts. Also, it means that there is no significant difference between the opinion of male lecturers and female lecturers on the construction of blue prints at 0.05. Therefore, the null hypothesis of no significant ($p \leq 0.05$) difference in the mean ratings of the responses of male and female lecturers is accepted for hypothesis four

Discussion of Findings

Results in table 1 reveal the problems of scoring in assessment. Murat & Nurdan (2015) mentioned problems teachers faced during scoring, they are: having insufficient time seems to be the other reason, unclear questions asked, students refused to answer questions, even when they asked, teachers don't give clear point, having errors in questions and providing no bonus question, no mutual scoring system. From this finding, the respondents are in agreement with the statement about the problems of scoring in assessment in all items with mean above criterion mean. The conclusion is that problem encountered during scoring affect assessment. The findings from table 2 indicate the problems of recording in assessment such as no main point, marks are not properly allocated, no marking scheme, lack of concentration and it is slow and laborious. However, they agreed that they set more questions from areas considered more important. All the Items are in agreement with the mean above the criterion. The conclusion is that problem encountered during recording affect assessment. The findings of table 3 reveal the problems of marking scheme in assessment for instance late preparation of marking scheme, on familiarization with the marking, carefree attitude of the teacher, distractions of the teacher by his/her private business. This means that respondents are in agreement with the statement about problems of marking scheme. The conclusion is that problems encountered during marking scheme affect assessment. The findings as revealed in table 4 is the problems of construction of test blue print in assessment. Some of the problems are that assessment tasks did not cover all levels of cognitive outcome. This is quite disturbing even as Ogomaka (2005) avers that the inability of an assessment instrument to have content coverage of topics as well as levels of cognitive outcome has far reaching implications for curriculum, teacher and product evaluation. Similarly, Onyeka in Onuekwusi & Agoha (2015) emphasizing the importance of high level cognitive skills (HLCS) in human capital development, posits that its absence, disadvantages the students who are expected to show some level of High thought processes in taking worthwhile decisions. This is quite true because tasks on Lower Level Cognitive Skills (LLCS) encourage mere acquisition of knowledge. Therefore it is discouraging because the result indicated that high order cognitive skills like synthesis and evaluation which are expected to be most useful in assessment are not covered.

Conclusion

One of the primary measurement tools in education is the assessment. Teachers gather information by giving tests, conducting interviews and monitoring behavior.

Challenges of Assessing Students in Measurement and Evaluation Sphere - Dr. Felicia King-Agboto; Dr. Daniel Orifamah and Dr. Chizoma Catherine Okpara

Therefore assessment should be carefully prepared and administered to ensure its reliability and validity. Effective use of marking scheme in scoring among other means reduces subjectivity which is the major disadvantage of essay tests. The necessary procedures for scoring should be imbibed to help teacher achieve high scoring reliability.

Recommendation

1. There should be enhancement on teacher's awareness on the importance of table of specification.
2. Training on how to design table of specification should be organized for teachers.
3. There is need for teachers to acquire the knowledge and skill for proper scoring technique in marking students

References

- Anders & Gunilla(2007)The use of scoring rubrics: Reliability, Validity and educational consequences. School of Teacher Education, Malmo University, SE-205 06 Malmo, Sweden
Research review
- Chikwe, K.C (2017) Fundamentals of Test, Measurement and Evaluation in Education. Published Emmanest Ventures
Port Harcourt
- Goldstein,H (2017) Measurement and Evaluation issue with PISA.
New York and London, Routledge
- Esomonu, N.P.M (2006) Scoring Techniques for Higher Accuracy in Students Assessment: Journal of the Nigerian Society for Educational Psychologists (NISEP) Vol.4, No.1 April
- Inko-Tari,D.C (2004)Educational Measurement and Evaluation
Published by Nissi Books Port Harcourt.
- Kpolovie,P.J (2002)Test,Measurement and Evaluation in Education
Published by Emhai, University of Port Harcourt Nigeria
- Murat,T & Nurdan,O (2015) Students view of the problems faced with the measurement & evaluation system in the primary School mathematics Education system. Published by Elsevier Limited.www.sciencedirect.com
- Nduwayezu,F (2020) Measurement & Evaluation issues in

Education.www.academia.edu

Ogomaka, P.M.C (2005).Some Grey Areas of External Examination Bodies Evaluation of Secondary Students Achievements in Nigeria. Nigerian Academy of Education 149-156

Onuekwusi,C.N &Agoha,T.C (2015) Perception and Validity of Assessment by Entrepreneurship Educators in Selected Tertiary Institutions in Imo State. A publication of the Association of Educational Researchers and Evaluators of Nigeria (ASSEREN) Volume 14, Number 2, 2015

Okonkwo, C.A & Osuji (2006) Measurement and evaluation .National Open University of Nigeria. The REGENT (printing& Publishing) Ltd

Onunkwo,G.I.N (2015) Evaluation instruments and techniques in Education Principles and Applications, Vigo publishers International Onisha, Anambra State

Pandey,T. Smith,T.R (1991) A sampler of mathematics assessment. Retrieved 30th march at www.Eric.com

Salvo,I(2009). "Il liceale con la media del 9,93 "Sono il più bravo d'Italia"". repubblica.it (in Italian).

Retrieved (2021)www.wikipedia.org/wiki/Gradingineducation.Grading in education
Ukozor,F,I & Amadi,V.C (2015) Scoring techniques and Minimizing Subjectivity in Essay test for Accuracy and Reliability in Students ‘ Assessment. A publication of the Association of Educational Researchers and Evaluators of Nigeria (ASSEREN) Volume 14, Number 2, 2015