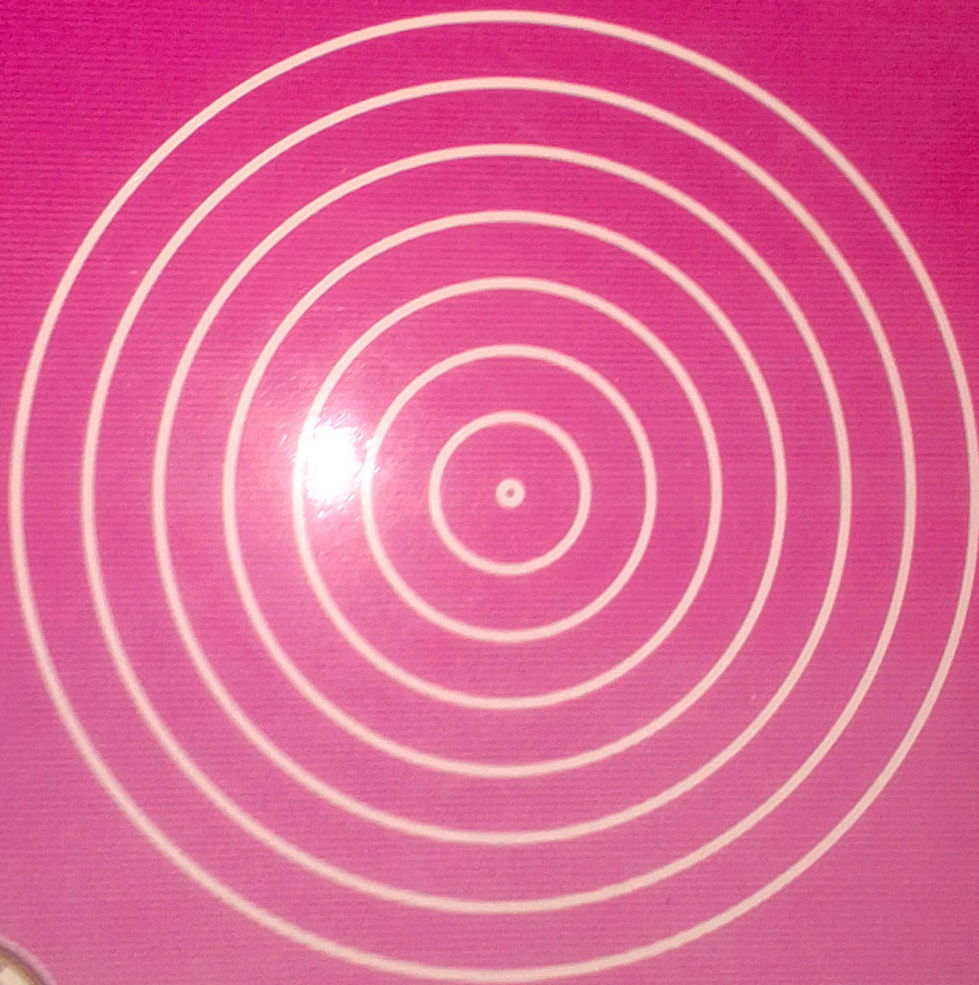# A Book of Reading

INSTRUCTIONAL PEDAGOGY

School of Science & Technology Education,
Federal University of Technology,
Minna, Niger State.

EDITORIAL TEAM

1. Prof. Ohize, E. J.   -   Chairman
2. Dr Falode, C. O.   -   Member
3. Dr Tukura, C. S.   -   Member
4. Dr Hassan, A. M.   -   Member
5. Mr. Abutu, F.   -   Member
6. Dr Chike-Okoli, F. C   -   Member
7. Dr Rabiu, M. B.   -   Member

TABLE OF CONTENTS

Chapter 14

Measurement and Evaluation

Idris U. S. B.
&
Gimba R. W.

Introduction

Unit three is all about some aspects of measurement and evaluation. It discusses different scoring systems and techniques, scales of measurement and types of evaluation with examples.

Behavioural Objectives

At the end of this unit, the participants should be able to:

ee. Differentiate between measurement and evaluation;

ff. List and explain four scales of measurement with examples; and

gg. Differentiate between formative, summative and ultimate types of evaluation.

Meaning of Measurement

Measurement differs from testing. Kolawole (2005) defined measurement as the process of using numbers to describe quantity and frequency according to a set of rules. In the same vein, Anikweze (2005) defined measurement as the assigning of numbers to observed events or objects based on acceptable principles. In other words, it is a logical quantitative process. Examples of measurement can include: getting length of objects or places in millimetres, centimetres; weights in grams and kilograms; sizes in diameters, square meters; temperature in degrees; among others. All these are done based on certain rules before reaching the assigned quantity. Educational measurement therefore, can be referred to as a process of assessing skills, knowledge and affective status giving certain values to indicate quantity or magnitude of the variable. For example, a test score of 30/40 simply indicates that the student scores 30 points out of 40. Therefore, the score 30 is a product of measurement.

Scales of Measurement

Generally, scales of measurement are scales used in quantifying variables in scientific investigations. A variable can change in value and can also be assigned a numerical value (Coolidge, 2006). A variable can be discrete or continuous.

Discrete Variable: This is any variable that is unique and different qualitatively from one another. For example, gender can be said to be discrete variable as male can be valued as 1, while female takes 2 but it is important to know that their values do not range between 0 and 1, i.e. nothing like being more male or female.

Continuous Variable: This is any measurement that can be made along a continuum (line scale) varying from smaller to larger number. Examples include time, test scores, temperature, heights, *et cetera*.

There are four basic types of measurement scales namely: nominal, ordinal, interval and ratio scales.

Nominal Scale: This is also called categorical scale. It assigns people or objects to qualitative groups. Gender for example, is a nominal scale assigning people to groups of males and females. Males may be assigned 1 while females are assigned 2. There is no intermittent value in between, that is to say no male can be more or less male and no female can be more or less female. Also, in a yes or no response, there can be no more yes or no than the other.

Ordinal Scale: This is another scale that involves ranking individuals, events or objects on some variables. A variable is ranked number one if it assumes the highest value in distribution. An example is class position, where students are ranked based on the magnitude of their total scores. The drawback of this scale is that the difference between ranks are not statistically equal. Difference between first and second may not be the same with third and fourth.

Interval Scale: This is a scale of measurement that has attributes of the two mentioned earlier, and in addition, having equal interval between units. Interpretation is made considering how far apart events, individuals or objects are on the variable. For example, in an attitude scale, usually the units are rated as 5, 4, 3, 2, 1 which depict that difference between 5 and 4, is equal to differences between 3, 2 and 1 respectively.

Ratio Scale: This scale contains the properties of all the preceding scales mentioned with addition of an absolute zero. Example can be seen in measuring

this is the most rarely used scale in educational research as zero is not appropriate in describing some variables like knowledge, motivation and so on. But in business studies and economics, concept like zero income is used.

Some examples of variables are matched with the appropriate scales in Table 2.1.

Table 2.1: Variables and Their Scales.

S/N	VARIABLE	SCALE
1	IQ scores	Interval
2	Performance/Exam scores	Interval
3	True/False and Yes/No	Nominal/Categorical
4	Gender	Nominal/Categorical
5	Colours	Nominal/Categorical
6	Temperature	Ratio
7	Rank	Ordinal
8	Grades	Ordinal
9	Height	Ratio

Scoring Systems

Scores are numerical values assigned to an attribute in a measurement to indicate quantity, magnitude or amount of certain variables. There are several scoring systems used in educational measurement. They include:

1.   *Raw Scores*: These are marks assigned to an individual in a test to indicate level of achievement in the area assessed. They are crude or unrefined scores that show what an individual obtained out of a given task. For example, student scoring 10/10 signifies that he/she achieved all that was given as a task. Raw score gives room for further interpretations of performances and traits.

2.   *Percentage*: This is more advanced than the raw scores. It can be seen as the proportion of students' achievement in the total task. It is the

conversion of raw score to percentage. If a student scores 40 out of 60 tasks, then the percentage will be 40/60 x 100 = 67%.

3.   *Grading*: In Nigeria, percentages are further converted into letter grades as follows:-

Table 2.2: Percentage, Grades and Interpretation

SCORES IN %	GRADE	INTERPRETATION
70-100	A	Distinction/Excellent
60-69	B	Credit/Very good
50-59	C	Merit/Good
45-49	D	Pass/Fair
44- 0	F	Fail

Ranking: Ranking is the act of converting raw scores to ordinal scale. Usually, the highest score is ranked first, followed by the next lower score and so on until the least score. For example, five students' raw scores of 90, 60, 80, 45 and 50 will be ranked as follows:

Table 2.3: Score Ranking

STUDENT	SCORES	RANK
A	90	1st
B	60	3rd
C	80	2nd
D	45	5th
E	50	4th

In the case of tied scores, the scores are assigned same rank and the next to them is ranked with the cumulated position as follows:

Table 2.4: Ranking of Tied Scores

STUDENT	SCORES	RANK
A	40	1st
B	30	2nd
C	30	2nd
D	25	4th
E	15	5th

Common Scoring Techniques

Scoring techniques are those methods used in assigning scores to given responses in a test or performance task. They vary according to the types of test or task.

Scoring Objective Test (Multiple Choice): There are various techniques used in scoring multiple choice items which include:

Number Right: This is the most widely used technique in scoring objective tests. A minimum score of 1 is assigned to a right option and 0 score to a wrong one. Eventually, the right scores are counted against the total scores to indicate the proportion of what student earned. For example, in a test of 50 marks, if a student got 40 right, his score will be $\frac{40}{50}$, and to make it into percentage, it is multiplied by hundred, e.g. $\frac{40}{50} \times 100 = 80\%$. That means the student got 80% of the task right, failing only 20%.

Logical Choice Weight: In this technique, student is expected to have full, partial or no knowledge about the task. Thus, options are weighted according to how accurate or closer they are to right answer. For example in a five option item, best option has 1 mark, followed by closer option with 0.75, 0.50, and 0.25 until the wrong option with 0. If it is 4 options (A-D), it will be 1, 0.67, 0.33 and 0 respectively. It is more flexible than the number right as it gives room for a testee to earn mark on the little he/she knows about an item.

Corrected Method: This method may not be favourable to students as it contains penalty for guessing answers. It assumes that if student really knows, the correct

option will be chosen, but if not, there will be guessing. It is calculated as $R - (\frac{W}{N-1})$ where

R = right answers, W = wrong answers, N = number of options. For example, if a student scored 40 out of 50 items with 5 options each, the final score will be $40 - \frac{10}{5-1} = 40 - \frac{10}{4} = 40 - 2.5 = 37.5$ (75%). Finally the student drops from 40 to 37.5, and from 80 to 75 per cent.

4. *Essay Scoring (Subjective)*: This is the traditional scoring of written responses of the testee based on the provision of a marking scheme. It is highly subjective but with a good marking scheme, substantial objective scoring can be achieved

Scoring Essay Test: Essay tests, due to the subjective nature of the format, are best scored with a well-developed marking scheme. Nonetheless, examiners can follow the guidelines below so as to achieve a more objective scoring of an essay test:

1. Carefully develop a marking scheme with appropriate allocation of marks. Minimize the use of decimal points in marks as much as possible.

2. Start with scoring question one across the scripts, followed by question 2 and so on. But if it is a team teaching course or subject, each teacher or lecturer should mark definite questions to ensure consistency of scoring.

3. Never allow unnecessary bio data on the scripts

4. Examiner should concentrate more on the expected facts.

5. If examiners are more than one, they should meet and train themselves on how to achieve effective marking of the tests.

Evaluation

Evaluation is seen as making value judgment on certain phenomenon. In other words, it entails making decision on a variable in question based on data obtained (Allan & Francis, 2009). It is also seen as the formal determination of the quality, effectiveness or value of a programme, product, project, process, objective or curriculum (Blaine & James, n.d. in Allan & Francis, 2009). For examples, if a student scores 60% in a class test, decisions like pass or fail, good or poor performance, qualified or disqualified could be made on the score obtained. In this case, one can say that evaluation incorporates measurement. Evaluation in education can be said to be a process of determining the extent to

which stated instructional objectives are achieved by learners (Anikweze, 2005). In summary, testing leads to measurement which also leads to evaluation. Therefore, one cannot evaluate before measurement except on hypothetical ground.

There are three types of evaluation namely formative and summative and ultimate evaluation

Formative evaluation: This is a pre-and/or intra-teaching evaluation that enables getting feedback which should inform course and content of instruction. It has the following characteristics:

- It concentrates on examining and changing processes as they occur;
- It provides immediate feedback about a programme, students, methods;
- It enables adjustment of programmes, methods of teaching, use of instructional materials in the process, as to help meet the desired or set goals;
- It enables diagnosis of learning difficulties in students as to proffer solution early enough;
- It improves performance of students in the final assessment as it influences students' memory and eliminates anxiety among students in the summative evaluation

Types of Formative Evaluation: Formative evaluation can be any of the following types:

Needs Assessment Evaluation: This determines who needs a programme, how great is the need, and what might work to meet the need;

Structural Conceptualization Evaluation: It helps the stakeholders or the government to define a programme, the target population and as well as the possible outcomes;

Implementation Evaluation: - This is the type of formative evaluation that monitors the fidelity and quality of a programme delivery

Process Evaluation: This helps to investigate the process of delivering the programme including alternative delivery procedures.

The use of continuous assessment in educational system is a form of formative evaluation that can fit the four types mentioned. The students' performance in the process can help to evaluate the process of teaching and learning, how the curriculum is implemented, what is lacking in the curriculum and so on.

Summative Evaluation: This is the second broad type of evaluation that occurs at the end a programme or course of instruction to provide an overall information of the effectiveness of a programme/course of instruction. Thus it examines the outcomes of a programme or course of instruction.

Summative evaluation should provide answers to the following questions:

- Were the set objectives achieved?
- Will there be any need to modify and improve the programme?
- What is the overall impact of the programme?
- What resources will be needed to address the weakness of the programme?

Types of Summative Evaluation: There are several types of summative evaluation as outlined by Pell Institute (2016), but the common ones include:

- *Goal-based Evaluation*: This tries to determine if the intended goals of a programme are attained, for instance, has the programme accomplished the set goals?

- *Outcome Evaluation*: This investigates whether or not the programme caused practical effects on specifically defined target outcomes, for instance, what effects programme participation had on the students or target audience?

- *Impact Evaluation*: This is more complex and tries to assess the overall or net effects, intended or unintended effects of certain programmes. For example, what impact does the programme have on the larger population like the school, college, community or the society as a whole?

- *Cost effectiveness and benefit analysis evaluation*: This tries to address questions on efficiency and standardizing outcomes in terms of their monetary costs and values. For example, it tries to answer question like how efficient is the programme with regards to cost?

Ultimate Evaluation is post-graduation evaluation when programme graduates are already engaged on the job or position for which he/she is trained. This is of particular importance in view of the fact that any educational process is targeted towards acquisition of knowledge, skill and attitude for gainful employment in real life and not just the award of certificates. Ultimate evaluation can be regarded as programme evaluation that may inform review of study programmes.

Summative Evaluation: This is the second broad type of evaluation that occurs at the end a programme or course of instruction to provide an overall information of the effectiveness of a programme/course of instruction. Thus it examines the outcomes of a programme or course of instruction.

Summative evaluation should provide answers to the following questions:

- Were the set objectives achieved?
- Will there be any need to modify and improve the programme?
- What is the overall impact of the programme?
- What resources will be needed to address the weakness of the programme?

Types of Summative Evaluation: There are several types of summative evaluation as outlined by Pell Institute (2016), but the common ones include:

- *Goal-based Evaluation*: This tries to determine if the intended goals of a programme are attained, for instance, has the programme accomplished the set goals?

- *Outcome Evaluation*: This investigates whether or not the programme caused practical effects on specifically defined target outcomes, for instance, what effects programme participation had on the students or target audience?

- *Impact Evaluation*: This is more complex and tries to assess the overall or net effects, intended or unintended effects of certain programmes. For example, what impact does the programme have on the larger population like the school, college, community or the society as a whole?

- *Cost effectiveness and benefit analysis evaluation*: This tries to address questions on efficiency and standardizing outcomes in terms of their monetary costs and values. For example, it tries to answer question like how efficient is the programme with regards to cost?

Ultimate Evaluation is post-graduation evaluation when programme graduates are already engaged on the job or position for which he/she is trained. This is of particular importance in view of the fact that any educational process is targeted towards acquisition of knowledge, skill and attitude for gainful employment in real life and not just the award of certificates. Ultimate evaluation can be regarded as programme evaluation that may inform review of study programmes.

Goal-based Evaluation: This tries to determine if the intended goals of a programme are attained, for instance, has the programme accomplished the set goals?

Outcome Evaluation: This investigates whether or not the programme caused practical effects on specifically defined target outcomes, for instance, what effects programme participation had on the students or target audience?

Impact Evaluation: This is more complex and tries to assess the overall or net effects, intended or unintended effects of certain programmes. For example, what impact does the programme have on the larger population like the school, college, community or the society as a whole?

Cost Effectiveness and Benefit Analysis Evaluation: This tries to address questions on efficiency and standardizing outcomes in terms of their monetary costs and values. For example, it tries to answer question like how efficient is the programme with regards to cost?

CDS (2017) identified impact and outcome evaluation as two types of ultimate evaluation and explained them as follows:

1. **Impact Evaluation**: This is done at appropriate intervals of an existing programme and at its end. It helps to assess the degree to which a certain programme meets its ultimate goal like determining if there has been improvement in primary and secondary education since the introduction of Universal Basic Education (UBE) in Nigeria or whether there has been any decrease in the drop-out or illiterate children in the society? This provides evidence for use in policy and funding decisions by government.

2. **Outcome Evaluation** is another form of ultimate evaluation as it measures programme effects in the target population by assessing the progress in the outcomes that the programme is set to address. In any institution of learning, the following questions could serve as guides for outcome evaluation:

- Were the teachers who received ICT training more likely to effectively teach than those who did not?

- Did the implementation of ICT in schools results in changes in knowledge, attitudes, and skills among teachers?

- Did the programme have any unintended (beneficial or adverse) effects on the target population(s)?

- Do the benefits of the ICT activity justify a continued allocation of resources?

Another example of outcome evaluation is the employers ratings where ex-students' performance is assessed and reported back to the institution he or she graduated from. To some extent, this will portray the outcome of the training received before graduation.

Conclusion

This Unit has taught some aspects of measurement which include measuring scales, scoring systems and techniques which if well digested, will boost the evaluation skills of teachers and lecturers irrespective of institutions of learning. It also gave insight on some aspects of measurement that are paramount for lecturers to be acquainted with and concluded with aspects of evaluation. This should go a long way in refreshing and improving evaluation skills of lecturers thereby ensuring effective evaluation in courses/programmes.

Exercises

1.   Explain the meaning of measurement and evaluation.

2.   Write short notes on measurement scales.

3.   Differentiate between formative, summative and ultimate evaluation.

References

Abiola, O. O. (2007). Procedures in Educational Research. Kaduna: King Nig. Plc.

Anikweze, C. M. (2005). Measurement and Evaluation for Teacher Education. Enugu Sneap Press Ltd.

Allan, C. O. and Francis, P. H. (2009). Curriculum Foundations, Principles and Issues. First Edition. Singapore: Pearson.

Fredrick, L. C. (2006). Statistics: A Gentle Introduction. California: Sage Publications.

Kolawole, E. B. (2005) Test and Measurement Revised Edition. Lagos: Bolabay Publications.

Kolawole E.B. (2010). Principles of test construction. Revised edition. Lagos: Bolabay Publications.

Pell Institute. (2016). Evaluation, approaches and types. Retrieved 10th April, 2016 from www.toolkit.pellinstitute.org/evaluation-101/evaluation-approaches-types

Popham, W. J. (2000). Modern Educational Measurement: Practical Guidelines for Educational Leaders. United States: A Pearson Education Company.

Robert, L. L. And Miller, M.D. (2005). Measurement and Assessment in Teaching. 9th edition. Ohio: Pearson Prentice H.

CDC (2017). Types of evaluation. Retrieved 11th March, 2017 from http://www.cdc.gov/std/program/ProgEvaluation.pdf

Kolawole, E. B. (2005) Test and Measurement Revised Edition. Lagos: Bolabay Publications.

Kolawole E.B. (2010). Principles of test construction. Revised edition. Lagos: Bolabay Publications.

Pell Institute. (2016). Evaluation, approaches and types. Retrieved 10th April, 2016 from www.toolkit.pellinstitute.org/evaluation-101/evaluation-approaches-types

Popham, W. J. (2000). Modern Educational Measurement: Practical Guidelines for Educational Leaders. United States: A Pearson Education Company.

Robert, L. L. And Miller, M.D. (2005). Measurement and Assessment in Teaching. 9th edition. Ohio: Pearson Prentice H.

CDC (2017). Types of evaluation. Retrieved 11th March, 2017 from http://www.cdc.gov/std/program/ProgEvaluation.pdf