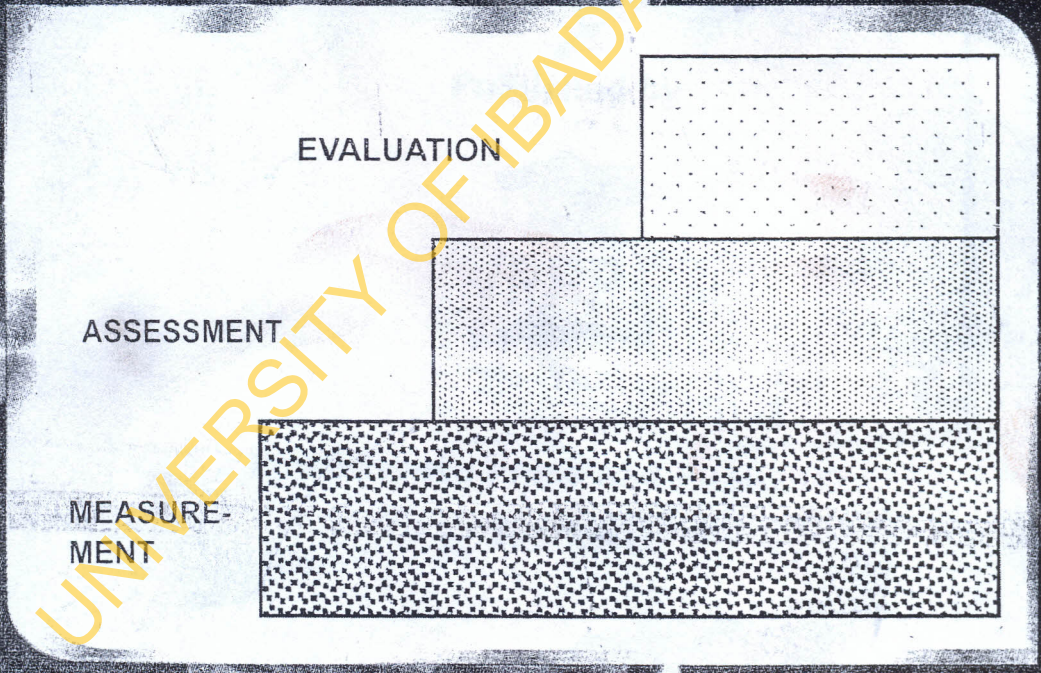
Evaluation Studies



By
Sam O. Ayodele,
Joseph A. Adegbile, J. Gbenga Adewale

EVALUATION STUDIES

by

Sam O. Ayodele

Joseph A. Adegbile

J. Gbenga Adewale

Published by

Powerhouse Press and Publishers

Ibadan

ISBN 978 – 35794-6-0

Second Edition Published by

Powerhouse Press and Publishers

Ibadan

Preface

This is a book that addresses one of the most crucial aspects of education - that of evaluation. The focus is on the *what* and the *how* of evaluation. In an era when it is becoming extremely difficult to import texts from the western world, the authors address themselves to the ever-increasing demand by tertiary students for learning tools. To ensure that all needy students can afford this most highly needed tool, we have decided to make the presentation as concise as possible. At the same time, we have tried as much as is practicable to include all that a student of educational evaluation needs within the limited space available.

The approach is to address what evaluation is, the kinds of evaluation processes there are, the place of such concepts as continuous assessment and curriculum evaluation, etc. Aware of the ogre that statistics has become among many undergraduate and postgraduate students, we have tried to present basic concepts in statistics in as elementary a manner as possible. This is the stance too in our treatment of evaluation instruments and later of the role of the computer in the service of the evaluator.

Here then is a complete volume in an introductory manner, made affordable because of its conciseness. It is our strong belief that students and teachers alike will find the volume highly useful and handy in colleges of education, polytechnics and university faculties and institutes of education.

In presenting this small volume, we dedicate it to those forerunners who have guided our own steps in the Institute of Education of the University of Ibadan: Emeritus Professor E. A. Yoloye and Professor Wole Falayajo.

Prof. S. O. Ayodele,
Dr. J. A. Adegbile,
Dr. J. G. Adewale.

CONTENTS

CHAPTER ONE

MEANING OF EVALUATION

Educators and evaluators have viewed evaluation in different ways. Some of the definitions have been considered to be shallow, abstract or ambiguous. The foregoing implies that evaluation is not easy to define apparently because of its complex nature. From all the available definitions, it is patent that the common area of agreement is the idea of "judging merit" which is one of the hallmarks of evaluation. For our discussion to be clear and objective, let us consider a few definitions as offered by scholars.

The first is that given by Tyler [1969]. According to him, evaluation is seen as simply as a process of measuring the success of teaching in terms of pupils' learning [products]. This is a definition given from the point of view of instruction or teaching, or more learnedly, evaluation of teaching. And evaluation is more than teaching.

Ezewu [1980] also defines evaluation in terms of students' achievements. He sees it as a process of determining the extent to which learners have achieved. The term evaluation means more than measurement. It involves quantitative and qualitative descriptions of the learner behaviour and the value judgment with respect to the desirability of the students' behaviour.

Yoloye [1971] opines that evaluation should be discussed operationally, basing it on the aspect one is evaluating. This explains why he alludes to the story of the six blind men of Hindustani, each giving his own description of the elephant after exploring, with his hands, a specific part of the animal. Thus, it stands to reason that there is a lot more to evaluation than meets the eye.

Another definition of evaluation worth considering is that given by Mervin C. Alkin [1970]. His definition appears to describe the whole process of evaluation. Evaluation, according to him, is "the process of ascertaining the decisions to be made, selecting related informa-

tion and collecting and analysing information in order to report summary data, useful to decision makers in selecting among alternatives". This is a detailed and objective definition. It is also clear from the definition that evaluation does not only involve decision making but helps us to make objective decisions. It, however, involves measurement and non-measurement descriptions of the learner's behaviour.

The major distinction between Alkin's definition and those by other scholars is that he has viewed evaluation beyond a process of ascertaining the degree of learner's performance in a test or the extent to which a teacher has attained the teaching objective. We can establish, from his definition, that evaluation helps the decision makers to take rational decisions through a formalized process consisting of 'data collection', 'data analysis' and 'preparation of summary reports'.

Relationship between Education and Evaluation

Just like evaluation, education has been defined in different ways. Though both are philosophical concepts, education embraces evaluation. One of the major activities in education is evaluation. Take for instance, the evaluation of a teaching strategy or students' performance; this cannot be carried out in a vacuum. It has to be done in the teaching-learning environment where instructional objectives are set before the teacher and the learner with the aim of establishing the effectiveness of the teaching strategy.

It is necessary to point out that education is a society's social institution and because of this it is an element of culture. As cultures across countries are not identical, education, which is an element of culture, cannot be expected to be the same. The way the Athenians viewed education was quite different from the way the Spartans did. This implies that education is not the same in the different societies of the world. A man who defines education as only through schools and colleges is seeing education in terms of schooling; and this is very narrow. The two Greek philosophers, Aristotle and Plato, define education as "moral and character development".

This definition is a total reflection of the needs of their society at that particular time. This is still a shallow way of viewing education because education is wider than moral and character development.

Education is therefore the means by which an individual born into a human society learns the mores and ways of life of that society, and this includes knowledge, skills and values of the given society through the older members of the society. This learning is both consciously and unconsciously imparted, and it is acquired so that he can function successfully as an adult member of that society. This is quite all embracing. It stands to reason that education can be obtained from school, homes and other spheres of life.

Since one of the activities in education is for the individual to learn the ways of life of the society, especially knowledge and skills, it is therefore necessary to evaluate the kind and amount of knowledge and skills acquired in order to establish their adequacy or otherwise. And if it is a more complex evaluation, such as the evaluation of any component of the 6-3-3-4 system of education, this kind of evaluation should be carried out within the school system. This is in order to establish whether the component should be abandoned or be improved upon. Thus, evaluation is seen within the realm of education as a complementary factor.

In the teaching-learning process, evaluation plays a lot of roles. It clarifies the instructional objectives in such a way that they are stated in terms of specific behaviours since the goal of the teacher is behavioural change. Through the aid of evaluation, objectives are stated in a more achievable and measurable way and thus teacher's attention or focus becomes very sharp. Also, through the aid of evaluation, the needs of the learners are pre-assessed in line with the learning outcomes to be achieved. The learners that still do not possess the necessary skills then benefit from the information gathered from what evaluation processes.

Evaluation and education are closely related in terms of providing quality control. This is patently so because the teacher appraises

each step of the teaching-learning process, and as he does this, the effectiveness of the process is established.

For the teacher to ascertain the extent to which the learners have achieved the instructional objectives, the process of evaluation serves as a useful aid to do this. In order to do this effectively, achievement tests and other relevant evaluation instruments are designed to measure the intended learning outcomes. Also, in the teaching-learning process, evaluation contributes greatly towards the improvement of students' achievement in terms of providing feedbacks on students' performance and thus giving useful information on how the students' problems can be solved. It is noteworthy that more reliable information could be obtained from well-developed evaluation instruments. This kind of information could improve teaching and learning tremendously in terms of helping to appraise the extent to which instructions strategies and materials are appropriate.

Thus far, efforts have been made in this chapter to show how much evaluation is related to education. It has been pointed out that evaluation is very useful in appraising the educational effectiveness of the quality of teaching and instructional materials. It can be concluded that without evaluation, education is not complete. By implication, the amount of learning acquired cannot be said to have been sufficiently sound without evaluation of such acquired learning.

Relationship between Measurement, Assessment and Evaluation

Many students of evaluation have often used the terms *assessment, measurement* and *evaluation* interchangeably or inappropriately. This should explain why it is necessary to discuss these terms very clearly in this unit. It should be emphasized that the three terms are often used in the teaching-learning environment. These will be discussed in detail and objectively for the purpose of clarity.

Let us consider the following hypothetical performance of a group of senior secondary students I in English Language.

Ojo = 46	Andrew = 71	Erinmu = 61
Tunji = 63	Abdul = 63	Okoye = 53
Omoh = 55	Sarah = 70	Dayo = 33
Stella = 47	Bola = 63	Aliyu = 38
Ade = 51	Tobi = 56	Bade = 40
Bayo = 69	Boye = 48	Toye = 45
Thomas = 54		

The scores shown above are students' performance in English Language. We can see that the highest score is 71 while the lowest is 33. The teacher has administered the test on the students in order to measure and determine their performance based on the lessons he has taught during a specific length of time. So, each testee's score is a measure of his performance, a measure of the extent to which he has mastered what he was taught. What the teacher has done is translating qualitative information to quantitative information, and when we do this, we are measuring. In other words, measurement is the process of translating observations into numbers.

Measurement can thus be defined as the process of assigning numbers to human characteristics or attributes or that of objects or events based on certain rules or regulations. So, from the testee's scores indicated above, the characteristics of the learners were measured and not the learners themselves. For instance, the scores of Ojo, Stella, Thomas and Andrew, which are 46, 47, 54, and 71 respectively, are the measures of their performance in the English Language test. And without this measure, it will be difficult to determine learners' performance. Apart from performance, other measurable attributes include achievement, weight, attitude, interest, etc. In the teaching-learning process, measurement is a very important activity expressing in quantitative terms the degree to which learners possess a characteristic.

In the same vein, if a teacher decides to take the length and breadth of a classroom in order to determine the number of chairs and lockers for such a class, he tries to establish the characteristics

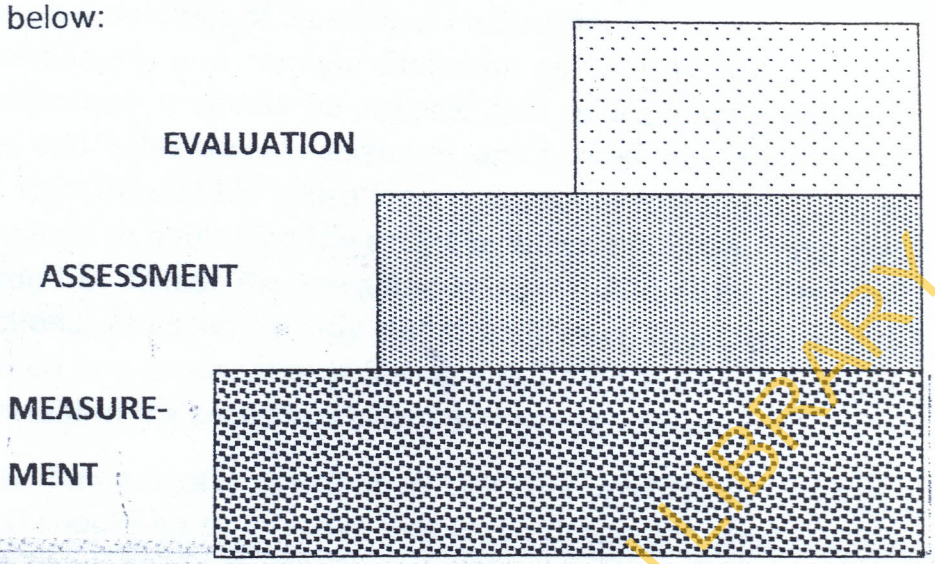
possessed by the object [class]. In this case, the teacher has measured the characteristics of the class and not the class itself. The foregoing implies that measurement is a descriptive process, that is expressed in quantitative terms. However, an evaluator is supposed to use a good evaluation instrument in order to have a reliable measurement. For the purpose of clarity, an evaluation instrument is the tool used to carry out an evaluation process. Examples of these may include tests, rating scales, checklists, questionnaires, etc.

On the other hand, an evaluator or a teacher after measuring the characteristics of the learner in a particular subject may want to know the status of such a student in the subject. In schools nowadays, a combination of letter, number and descriptive adjective conversions is often used. This has often been used to establish the grade equivalent norm of the performance of the testees. The West African Examinations Council in its GCE examinations, adopts this method. Raw scores are converted into a nine-point scale as follows:

Distinction = A1,
Excellent = B2, B3
Credit = C4, C5, C6
Pass = D7, D8
Failure = F9

The raw scores, which correspond to these numbers, vary from year to year. So, on the basis of the foregoing, assessment could be referred to as the process of establishing the status of the performance of an individual or group in a given task usually with reference to the expected outcomes. It should be very clear at this juncture that assessment is concerned with both quantitative and non-quantitative descriptions of objects or events. Measurement, assessment and evaluation do not mean the same thing. It has been pointed out earlier that measurement is limited only to quantitative description of things. Hence, assessment unlike evaluation, does not mean judgment concerning the worth or value of the behaviour being assessed.

A definite hierarchy is formed on the basis of the three terms with measurement forming the first step, assessment forming the second, while evaluation forms the last and highest. This is illustrated below:



With the raw scores given earlier on, if the teacher decides to improve on the performance of the testees by adopting a better teaching strategy in re-teaching the same group of students, and with these scores increasing, an assessment is being carried out. The ultimate goal of the teaching-learning process is to change the behaviour of the learner. And since a change has occurred as a result of the teacher's re-teaching the topics or application of the teaching strategy, it is now up to the teacher to establish how much change has taken place. An attempt to ascertain this involves the process of 'assessment'. We should stress at this juncture that in the process of carrying out assessment some measurement is done while the process of evaluation involves carrying out some assessment procedures.

General Principles of Education

The process of evaluation involves some fundamental principles, which could be regarded as criteria for appraising the degree to which certain procedures or practices are effective. It is expected that, if the process of evaluation is based on sound principles, effectiveness of such a process will be enhanced.

In the first place, for the process of evaluation to be effective, it is necessary to identify and delineate the entity to be evaluated. This has to do with the identification of instructional objectives, which is an intended outcome of a period of instruction. It is stated to show the behavioural and content dimension as well as performance level expected. It should be stressed that educational evaluative studies and other related exercises are carried out against the stated objectives of the educational process or programme. This is done mainly to determine the extent to which the programme has succeeded or failed. The foregoing implies that if there were no instructional objectives clearly stated earlier on, there would not have been any opportunity of having a standard against which we could establish the success or failure of the programme.

Another fundamental principle is that the evaluation strategy adopted should be so effective that it can determine what we intend to know about the learner or programme. This implies that the purpose for which the evaluation technique is selected should be seriously considered. And it should be borne in mind that one of the major purposes of evaluation is to furnish information which will facilitate making decisions as to whether to continue, to adjust or to discard an on-going programme in the educational system. Therefore, if an effective evaluation strategy is adopted, an attempt at making an objective decision in respect of an evaluation process or programme will be enhanced.

For an effective evaluation to be carried out, it is very important to determine the type of data we should collect. This is to enable the evaluator to decide how to go about the analysis of such data and the statistical tool to be adopted. It is quite absurd if the data collected are not relevant to what the evaluator is doing. And as long as the type of data is determined, the whole work becomes straightforward; the findings are reliable and useful to the decision makers. One of the duties of evaluators is to give a description of the research findings and make necessary inferences from the data collected. This explains why an evaluator needs to have a sound

knowledge of statistical techniques, and thus he will be able to report effectively.

An evaluator has to consider the principles of determining the individuals or groups of people from whom data will be collected or obtained. Data that are supposed to be collected from secondary school students should not be collected from university undergraduate students. Data should be collected from the appropriate group of people in order to get appropriate findings. Data collected from market women are definitely going to be different from those collected from nurses in clinics. The way and manner people respond to questions or items in a questionnaire may somehow be different. The information supplied varies from people to people and this is why a careful attempt must be made at deciding on who should furnish information on a particular research work. And since evaluation is a process of gathering information in order to aid decision-making, it is necessary to decide objectively on the type of people from whom data will be collected.

Evaluators should also bear in mind that evaluation techniques have their weaknesses and strengths. Errors may arise in the area of sampling or in the process of using it or in the wrong interpretation of the results from such techniques. So, if an evaluator is aware of this inadequacy he will be in a better position to use it differently by constructing them in such a way that the findings will not be faulty.

Programme or comprehensive evaluation involves the use of different evaluation techniques such as testing, observational techniques, socio-metric techniques, etc. Only one evaluation technique cannot effectively do what the evaluator really wants in a comprehensive evaluation. As a matter of fact, it is not easy for a teacher to achieve all the instructional outcomes with a single evaluation technique. Thus, it is very important that an evaluator has at his fingertips the use of different evaluation techniques. This is also a very important evaluation principle.

Concerns of Evaluation

Evaluation has a lot of concerns, and the following are of paramount importance in this respect:

a. Evaluation is concerned with clarifying the instructional objectives. It serves as an effective method of clarifying instructional objectives when they appear to be ambiguous. This is done in terms of the desired learning outcomes. In order for the behaviour of learners to be appropriately appraised by the evaluation technique, it is necessary that the teacher should state his instructional objectives for each lesson in an achievable or measurably manner. When the instructional objectives are well clarified and stated in behavioural ways, the teacher should then be in a better position to follow the teaching process appropriately and also provide enough opportunity for an effective evaluation of the learning outcomes. It stands to reason that if teaching objectives are not well clarified, there may be some problems in achieving the desired learning outcomes.

b. The clarification of instructional objectives, which we have mentioned above, is not the only concern of evaluation but also how to determine the extent to which the instructional objectives have been achieved is another primary concern. One of the ways of establishing this is by using tests or any other evaluation instrument. This evaluation instrument is expected to provide relevant data in respect of the learners' behaviour since this is specified in the instructional objectives.

c. Evaluation is concerned with the appraisal of the effectiveness of the instructional materials used by the teacher during the teaching-learning process. Since it is necessary for the teacher to decide on what materials to use as well in the strategy to adopt, it is the evaluation process that helps in taking such decisions so as to establish the effectiveness of the learning outcomes.

d. Evaluation is also concerned with the quality control in the teaching-learning process. This implies that the teacher is expected to find out at each stage, whether his method of teaching is effective or not, whether the teaching materials are appropriate or not. He is also expected to determine whether to make some changes in the methods adopted or in the instructional materials used. This is very essential because quality is one of the major yardsticks for measuring the achievement of the expected learning outcomes by the learners.

e. It is the concern of evaluation to furnish the teachers, parents, guardians, curriculum developers and policy makers with students' results or progress reports. These reports could be useful for administrative purposes. They guide the learner in taking rational or objective decisions in the areas of admission, the choice of a vocation, and in tackling their personal life problems. This should explain why testing, one of the evaluation techniques, plays a very important role in this regard.

f. It is clear from the points made in (e) above that it is also the concern of evaluation to contribute towards the improvement of student performance by providing useful information on how students can solve their learning problems and giving relevant and adequate feedback on students' performance. The foregoing is highly essential in order to achieve the expected learning outcomes.

CHAPTER TWO

KINDS OF EVALUATION AND EVALUATION PROCESS

It should be affirmed that the evaluation procedures adopted are determined by the evaluation techniques used. So, each evaluation procedure is based on the purpose for which evaluation is being carried out. This should explain why there are various kinds of evaluation. Let us consider the following, which are very significant in the teaching-learning process.

Test and Testing

We would first direct our attention on the terms 'tests' and 'testing' to enable us understand the subtle differences between them, and to make clear the functions of tests in the classroom instruction and types of tests.

When the word *test* is used as a verb, it is synonymous with the word *examine* for when you test, you are in fact examining a testee. The obvious purpose of this is to get responses from the testee. So, the tests which the teacher gives his learners could be regarded as 'stimuli' while the answers offered by the testees are called the 'responses'. The responses given may or may not be correct. If they are correct, it clearly shows that the testee has learnt what he is expected to learn. If the responses are incorrect, then the testee has not sufficiently learnt what is expected of him.

The foregoing implies that 'test' is an instrument. So, the items we print, type or handwrite is administered to a group of students at the end of a unit of a course or end of the

course as a whole. If, for example, a group of JSS 1 students are taught the pronunciation of certain English words for a period of time, the teacher might be interested in the following:

i. How many students were able to pronounce the words correctly?
ii. Who, specifically by name, can pronounce each of the words correctly?
iii. Who, specifically by name, cannot pronounce the words correctly?

In order to obtain detailed information about these, the teacher needs a test, as an instrument for measuring how correctly the students have grasped the pronunciation of the words. So, a classroom achievement test could be defined as follows:

> A means or instrument for measuring how much learners have learnt what they were expected to learn, who are those that have learnt them and who are those that have not.

Testing, on the other hand, is a process or procedure. In other words, testing is a procedure for carrying out a test. It is a process of administering the test constructed by the teacher. So, the difference between 'test' and 'testing' should by now be clear: that a 'test' is an instrument while 'testing' is a procedure. It is a procedure for administering the test on the testees to elicit the desired information from them.

Context Evaluation

Some evaluators believe that this kind of evaluation is diagnostic in nature. It tells us how to carry out the diagnosis of a particular programme. Context evaluation provides information on the state of the art. In this kind of evaluation, it is possible to carry out a state of the art on input, process and/or output parts of the programme. For instance, an evaluator could embark on the context evaluation of a state unity school. He may want to consider what is being done in the school or what makes it a unity school. In this way, the evaluator is providing information on the state of the art.

In the 6-3-3-4 system of education, a context evaluation of the JSS aspect of the system could be carried out to determine how it is being run or to establish its worth. If it is done like this, then a diagnostic evaluation is being carried out. JAMB is another good example here. An evaluator may want to look at the context and provide information on the state of the art. This kind of evaluation deals broadly with the educational system.

Input Evaluation

This is another kind of evaluation where input variables are seriously considered and their effects or impacts on the programme are measured. There are different input variables out of which we may have books, students, teachers, administrators and other facilities. If for instance there are large classes, there may be a problem. This may seriously affect the teaching-learning process. When there are large classes, teaching effectiveness might be seriously affected – marking of scripts is not thoroughly done, students may not have a

detailed feedback and the teacher eventually finds it difficult to cope in the classroom.

If an evaluator decides to do an input evaluation of the JSS programme in the 6-3-3-4 educational system, he is going to consider the quality of students admitted and the effects of other facilities on the programme.

Process Evaluation

Process evaluation concentrates very much on the question, "How?", that is, how things are being done. To put it in our present context, how a programme is being run. An evaluator may want to consider such variables as school and schooling. He may want to know how much of schooling there is. He may be interested in continuous assessment as a process, that is, how it is organised in schools.

The aspect of training of teachers or lecturers could be considered. The evaluator may want to ascertain how they are trained. He may want to know the number of teachers, books, items of furniture. Are the facilities adequate or inadequate for the programme? When all these are done, then, process evaluation is being carried out.

Product Evaluation

As the name implies, this refers to the product of an evaluation programme. Let us consider the International Centre for Educational Evaluation [ICEE] of the University of Ibadan. The primary concern is what an evaluator does after training and not what he does on training. An evaluator may decide to evaluate the products.

Primary or secondary school leavers may even be evaluated. Such questions as – 'How do they perform after training?' 'What are their contributions after training?' 'Are they worthwhile contributions?' etc. – could be attempted in this regard. Issues such as their impact on the society may be considered. If this is done, then we are in the realm of product-impact evaluation, which is otherwise known as product evaluation.

Formative Evaluation

In the formative evaluation, all mistakes are corrected as the programme is going on. It is the type of evaluation, which guides the implementation of a programme. For this kind of programme to succeed, there must be input corrective measures. The primary aim of the formative evaluation is to obtain evidence about the worth or adequacy of a programme while it is in progress. In the 6-3-3-4 educational system, terminal and promotion exams are a very good example of the formative evaluation. The continuous assessment also plays a very important role in this regard.

Formative evaluation also identifies strengths and weaknesses of a programme. It also gives a feedback about individuals and the extent to which each unit or chapter is mastered. It is now apparent that the kind of feedback in this kind of evaluation is geared towards teachers and principal or headmaster since learning and teaching weaknesses identified are corrected within the school system. Yet another important aspect of this kind of evaluation is flexibility. In other words, one should be ready to change for the better so that the programme can succeed.

Summative Evaluation

Summative evaluation is carried out when the programme has been completed. It aims at collecting evidence in order to judge the worth or adequacy of the whole programme or learning process. In the school, summative evaluation is usually done at the end of a lesson, a term, a semester or a year. Feedback from this kind of evaluation often shows the extent to which the general instructional outcomes are achieved. Examples of summative evaluation include the final senior secondary school certificate examination, or general certificate of education, semester and final degree examinations in higher institutions of learning. It should be pointed out that the mastery errors, which a summative evaluation identifies, do not increase the grade of a student but may provide a sort of self-satisfaction. Generally speaking, students, teachers, guidance counsellors, university or college admissions officers make use of the summative evaluation results.

Goal-free Evaluation

This kind of evaluation does not suggest that there are no goals. It refers to the fact that there are side effects or, to put it learnedly, that there are unintended goals. At times, the unintended goals become important enough to drastically affect the programme. Take for instance, we may be training a group of students and it is discovered that one of them is a robber. Attention will now be focused on the student who has become a social misfit in terms of looking for ways of reforming him or nipping such a case in the bud. This implies that there is a goal but that there are some side effects, and the side effects form the goal of the programme.

Evaluation is a continuous process, which underlies all good teaching and learning. The role of evaluation in many facets of the school programme cannot be over-stressed. It does not only contribute to the teaching and learning process used in the classroom, but also to marking and reporting, programmed instruction, school research programmes and curriculum development. The principal features of the evaluation process are discussed below.

a. *Identification of the entity to be evaluated*

The entity to be evaluated should be well identified and delineated. In the school programme, an evaluator may want to consider teachers. And on this, the major factors worth considering about the teacher include their training, knowledge, classroom management acumen, interpersonal relations and professional ethics. Other groups of people that may be evaluated are learners, administrators, members of the community, and others.

b. *Determining the type of data to be collected*

On the basis of the meaning or definition, which has been given in this book, it is clear that evaluation requires some kind of data collection. The design of the evaluation study depends largely on the conditions under which the data are collected. The quantity and adequacy of the data collected can be judged by the extent to which the results of the data are interpretable and generalisable.

When people are subjected to experimental treatment, the kind of data we collect here can be described as straightforward experimental data. Some variables are manipulated here while their effects on other variables

are observed. This is usually the effect of independent variables on dependent variables. When, however, the information is already available and which does not involve manipulation of variables, the kind of data we collect can be described as quasi-experimental.

c. *Collating the Data*

Another important step in an evaluation exercise is the collation of data that have been collected. It should be pointed out that the data collected in an evaluation exercise have to be organised in order to make them readily usable and accessible. In a small scale evaluation exercise, the problem of data collation is not usually enormous. This implies that as the size of the evaluation increases so do the problems of data collation.

In whichever way the data have been collected, they have to be collated or prepared to ensure good quality data sets and efficient data analysis. After the data have been collected in the field, it is the duty of the field staff or the persons administering the test to collate them. This will prevent the possibility of returning incomplete data and the consequent waste of time.

d. *Analysing the data*

The guiding principle of analysing the data is that the type of analysis is entirely dictated by the kind of questions an evaluator is asking. The questions asked can be categorised broadly into three, viz:

i. There is 'What' type of questions. This is also called descriptive type of questions. The aim of an evaluator here is to describe the situation as it is.

ii. An evaluator can ask questions about relationships between variables. For example, such question as 'What is the relationship between age of school and

performance of the students?' Also, in this category, we use one kind of variable to predict another kind of variable.

iii. The third category has to do with questions of finding out the effect of one variable on another variable. For example, an evaluator may want to consider the effect of teaching style on students' achievement.

The kind of question an evaluator is dealing with will dictate the kind of statistical analysis he should adopt.

e. *Interpreting the data*

This is the process of giving meanings to the data collected. There are different ways in which results can be interpreted. They can be interpreted in terms of the types of tasks that can be performed. This is otherwise called *criterion reference.* They can also be interpreted in terms of the relative position held in some reference groups. This is known as *norm reference.* It should be pointed out that both types of interpretations are useful. The criterion reference describes what a person can do, and the second describes how his performance compares to that of others. It should be noted that interpreting data with the aid of norms requires an understanding of the various methods of expressing test scores and a clear understanding of the nature of the norm group.

f. *Communicating*

It is generally believed that if instructional objectives have been well defined in behavioural terms, and evaluation procedures have been effectively applied, the task of reporting the data collected is greatly simplified.

In this unit, it is mandatory to discuss school marks and other reports of pupils' progress. This is obviously because they serve a variety of specific functions in the school. However, it is better to describe these in relation to the users of the reports. They include pupils and parents, teachers, counsellors and administrators. It is necessary to report to pupils and parents so as to facilitate the learning and development of pupils. The specific functions to be served are the same as those of the general evaluation programme. Marks and progress reports also contribute to the instructional and guidance programme of the school by providing much information about pupils. These reports supplement and complement test scores and other evaluative data.

Marks and progress reports also serve a number of administrative functions. In administration, they are used for determining promotion. Finally, it should be pointed out that an effective system of reporting would provide the type of information needed by the users of the reports. It also makes it possible for the report to be presented in a clearly understandable form.

CHAPTER THREE

STATISTICS

Definition of Statistics

Statistics is the study of gathering or collection of data or information, analysing the data and drawing inferences or conclusions from the data.

The needs for statistics are as follows:

i. To acquire the technique of data collection.
ii. To understand statistical procedures or methods for analysing our data, e.g. X^2, Z-score and T- statistics.
iii. To acquire the competency for carrying out or planning research.

The man who introduced statistics was William S. Gosset, who introduced standard error of the mean in sampling. The theory of probability was also introduced to deduce the chance or opportunity for success or failure.

The two general fields of statistics based on the use are *descriptive* statistics and *inferential* statistics.

Descriptive Statistics: This is the statistics used to describe the characteristics of a group. It deals with determination of the mean or average characteristics of a group, e.g. age, weight, height, etc. There are other descriptive measures like the *mode, mean* and *median, standard deviation, minimum, maximum,* etc., which could be used based on the need.

Inferential Statistics: This deals with making deductions or inferences based on the values from a sample, rather than the population provided the sample is representative of the population.

Common Terms, Symbols and Concepts in Statistics

SYMBOLS

a.	Σ	=	Sigma = sum of
b.	$\bar{x}$	=	Mean of sample
c.	μ	=	Mean of population
d.	Sx	=	Standard deviation of sample
e.	δx	=	Standard deviation of population
f.	Md	=	Median
g.	Mo	=	Mode
h.	f	=	Frequency
i.	Σf	=	N = number of group
j.	Σfx	=	Sum of product of f and x
k.	$(\Sigma fx)^2$	=	Sum of sq of the product of f and x
l.	Σfx^2	=	Sum of the product of f and x

where x = variable
and f = frequency

Terms

Population: A population is the large group that possesses or contains all the elements that have the common characteristics. This is the group on which the inferences are generalised.

Sample: Elements or members in the sample possess the same characteristics with the population but the differences are in the size (i.e. sample < population).

Parameter and Statistics: An index or measure described in the population is known as parameter, e.g. μ = mean of population. And indices or measures for the sample are called statistics, e.g. $\bar{x}$ and Sx

Variables and Constants

A variable is the characteristic of a sample that can take different values at any particular time, e.g. age, performance, teaching methods. A constant is the characteristic of a sample that is fixed in a particular situation, e.g. if all the samples come from a school or a city, then school or city is the constant.

Types of Variables

There are (a) Discrete and (b) Continuous variables.

a. Discrete Variables are those that can take a definite or fixed value, e.g. the number of children in a family, the number of books on a shelf, etc. They are also known as discontinuous variables.
b. Continuous Variables are variables like height, size or weight, which can take different values at any particular time. This is because between any consecutive points on the scale, there can be an infinite number of possible values. Although height for example is a continuous variable, we can change this into a discrete form by limiting the number of values between 2 consecutive points, e.g. height correct to one decimal place; it can take infinitesimal values.

Scales of Measurement

The assignment of number of values to the variables being measured, e.g. weight = 45kg, implies that a number 45kg has been assigned as the weight.

There are various ways in which we assign number to the variable. In some cases, we simply classify while in others we arrange in order of size. There are different ways or scales of measurement. These are nominal, ordinal, interval and ratio.

Nominal Scale: In a football match, players wear jerseys with different numbers [usually 1 to 11]. A player wearing jersey number 1 does not indicate that he is the best or the worst in the team. Hence, we use nominal scale for identification. We often categorise and make frequency counts on the group, e.g. categorising male and female as M and F, or 1 and 2.

Ordinal Scale: Here, the order or arrangement is important, i.e. from highest to the lowest, or the lowest to the highest. Consider the following example:

Average Scores:

76.5	1
76.3	2
73.0	3
71.0	4
64.0	5

It is done by ranking. It has no equal interval. In ordinal scale of measurement, we have ranks to indicate positions, which do not give any information about distance/interval between points on the scale.

Interval Scale: The differences between successive points are equal. Hence, they have equal intervals and are called interval scale. For example, the difference between 10 cm and 15 cm in a ruler is 5 cm. This is possible due to the fact that equal intervals are used in the measurement. There is no absolute zero point.

Ratio Scale: This scale has all the quantities such as absolute zero and equal intervals. For example, in thermometer the absolute zero is 273 Kelvin. This makes comparison between points in the scale possible and valid.

ORGANISATION OF DATA

Frequency: A teacher scores a test given to a class and forms a mark list according to alphabetical order of names. The following table illustrates this.

Table 3.1: Students' Scores

NAME	SCORE	NAME	SCORE
Adeola	84	Fassy	82
Ayobami	80	Ganddy	62
Bello	68	Gabby	87
Benga	87	Iyke	85
Benjey	86	Joe	86
Bennedy	70	Lawandi	61
Chistry	79	Lizzy	86
Dolapo	90	Nan	71
Dasola	67	Paul	78
Eli	80	Qudari	86

a. To identify the overall group performance.
b. Organising data enables one to identify an individual's performance in relation to the group performance.
c. To create order in the scores.

Method of Organising Data

a. Arrange the scores from the highest to the lowest.

90	86	80	70
87	86	80	68
87	85	79	67
86	84	78	62
86	82	71	61

Next, we count the number of times each score appears in the array of scores using tallying method.

Table 3.2: Simple Frequency Distribution table (ungrouped data)

SCORE	FREQ	SCORE	FREQ
90	1	78	1
87	2	71	1
86	4	70	1
85	1	68	1
82	1	62	1
80	2	61	1
79	1		

Let us now consider the ages of 30 students in a classroom from a frequency table in the data. The ages are:

13, 13, 14, 12, 15, 13, 12, 13, 13, 16, 13, 12, 14, 14, 15, 14, 16, 12, 13, 13, 12, 14, 13, 14, 15, 14, 15, 16, 15.

Table 3.3: The Frequency Distribution (ungrouped data)

Ages	Tally	Frequency	Cumulative Freq
16	///	3	30
15	/////	5	27
14	///// //	7	22
13	//// //// //	10	15
12	/////	5	5

This is used when the range is narrow.

Grouped Frequency Distribution

When the number of terms in a distribution is large and the range is wide, we need to apply grouping rather than just simple frequency count.

Use a class width of 3 and form a frequency distribution.

76	77	65	70	69	60	68	72	69	72
68	67	73	63	80	68	74	75	71	73
65	64	81	77	76	64	72	73	67	75
71	78	66	78	63	68	76	71	72	74
70	67	67	71	68	72	75	73	74	75

Procedure:

- We subtract the lowest score from the highest score.
- We try to divide the lowest by the class size. If it divides, we use the lowest as the starting point but if not, we use a number closer but lower than the lowest which is divisible by the class size or width.

In this example, the class width is 3. Then, we try to divide the lowest score 60 by 3. It is perfectly divisible, without a reminder, hence, we start to group from 60 and add 2 so that the interval will now be 3, e.g. 60 – 62, the interval is 3 but the difference is 2.

Table 3.4: Frequency Distribution Table

Class	Tally	Frequency	Cum Freq
81 - 83	/	1	50
78 - 80	///	3	49
75 - 77	///// ////	9	46
72 - 74	///// ///// //	12	37
69 - 71	///// ///	8	25
66 - 68	///// /////	10	17
63 - 65	///// /	6	7
60 - 62	/	1	1

Terms

a. Limits of the class – This is the first and the last terms in each class. We have lower and upper limits, e.g. 60 – 62 stated limits but 59.5 – 62.5 real limits.

b. Mid-point of a class – This is mid-way score in a class. E. g. 67-69 midpoint = 67 + 69 ÷ 2

= 136 ÷ 2 = 68

Note: the range in the distribution influences a class width. We use the stated limits in determining the frequency distributions whereas we use the real limits for calculations.

Graphs

Graph is a diagrammatic representation of data. It becomes much easier for some people to interpret information from graphs than from tables. Data could be represented in any of the followings:

a. Pie chart
b. Simple bar or line chart
c. Multiple bar chart
d. Histogram
e. Frequency polygon.

Pie Chart

This is a circle that is folded when it has represented an item in a given list.

Example:

The followings are the Nigeria's exports for a given year.

Table 3.5: Nigeria's Exports for a Particular Year

Export Commodities	Weight in metric tonnes	Degrees in a circle
Palm product	264	132.00
Groundnut	251	125.50
Cocoa	20	10.00
Hide	105	52.50
Rubber	5	2.50
Others	75	37.50
Total	720	360.00

Formula:

$$Q = \frac{\text{Number of each item} \times 360^0}{\text{Number of Total item}}$$

e.g. for palm product,

$$Q = \frac{264 \times 360^0}{720} = 132^0$$

Fig 3.1: Pie chart of Nigerian Export Commodities

The Pie Chart is used to illustrate a discrete form of data; hence it is used in a nominal scale.

Simple Bar Chart also is used a nominal scale since it is independent of each variable. The commodity or scores or class in Table 3.5 is on the horizontal axis while frequencies are on the vertical axis. Again, we require consistency, there must be equal spacing and drawn to scale. A graph sheet is needed. Using Table 3.5 for the pie chart above, we can draw a simple bar chart, as shown in the following example.

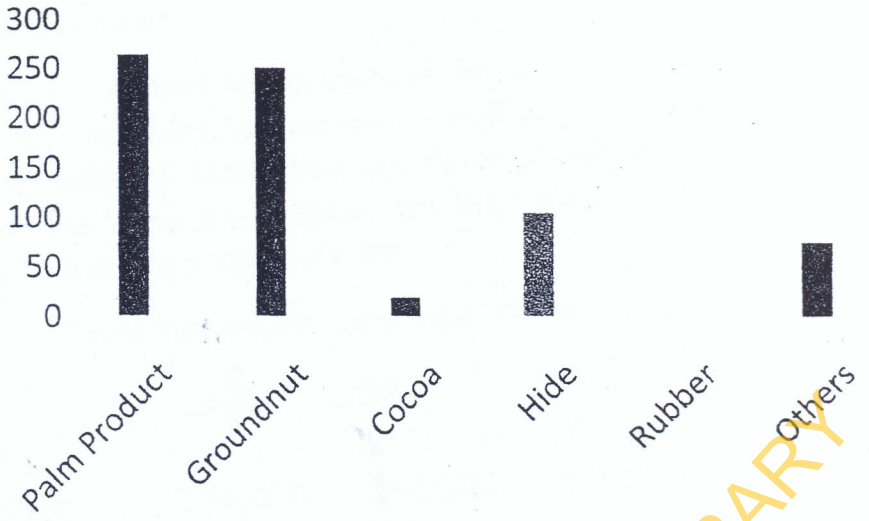


Table 3.2: Simple Bar Chart of Nigerian Export Commodities

We can also draw it with the bars being horizontal. This is done by simply tilting the chart through angle 90 in a clockwise direction as shown in the next figure.

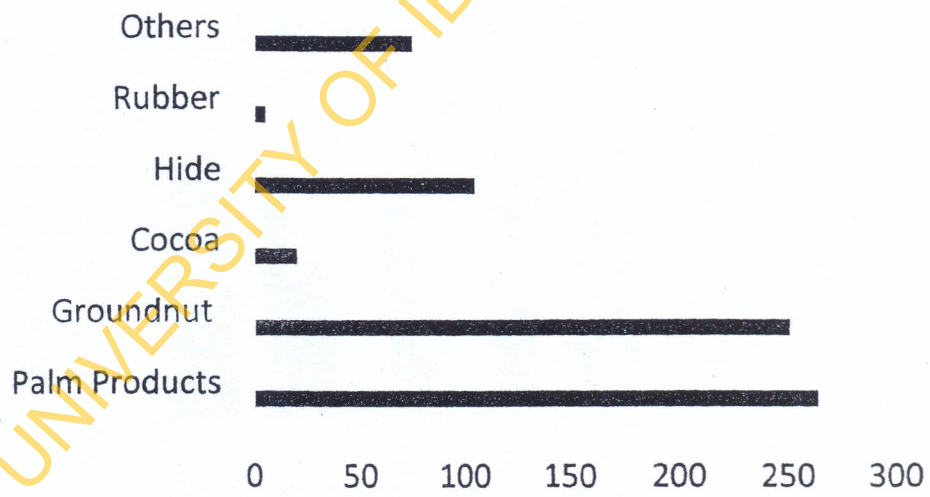


Figure 3.3: Simple Bar Chart of Nigerian Export Commodities

Multiple Bar Chart

In this case, we allow for comparison. So, when we have more than one item on our bar chart we need a multiple bar chart, it is used to compare different distribution at different times. We can therefore consider the Table 3.6. It shows the Nigeria's export crops for the different years from 1986 to 1988.

Table 3.6: Weights of products in metric tonnes

	1986	1987	1988
Palm Product	264,000	250,000	275,000
Groundnut	250,000	132,000	150,000
Cocoa	20,000	200,000	50,000
Hide	5,000	80,000	200,000
Rubber	105,000	100,000	200,000
Others	75,000	100,000	200,000

The table is illustrated with Fig 3.4.

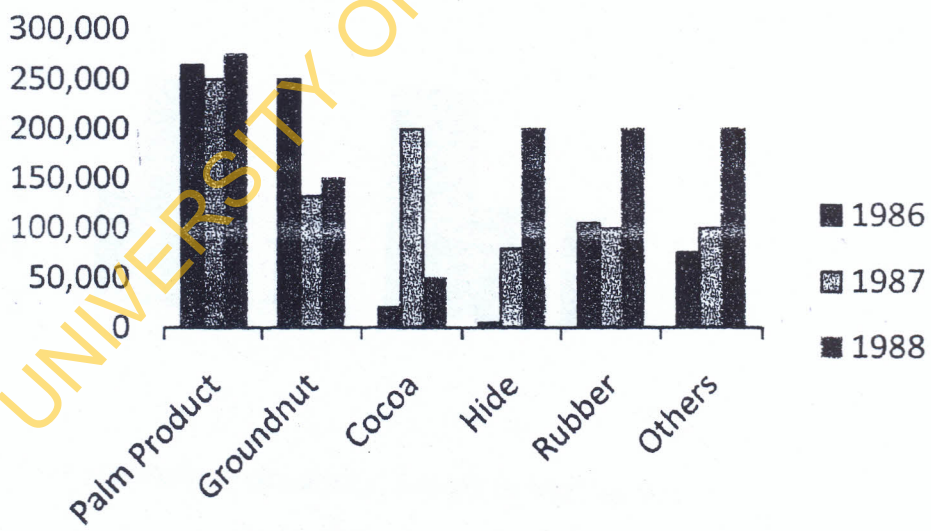


Fig. 3.4: Multiple Bar Chart of Nig. Export Commodities 1986-88

Histograms

Histograms are also bar charts but no space is left between bars thereby showing the continuity of data. Histograms are constructed with data having interval sales.

Display the data below using histogram

Table 3.7: Scores of Students in Mathematics

Scores	Frequency	Scores	Frequency
89.5 90 90.5	1	83.5 84 84.5	2
88.5 89 89.5	2	92.5 83 83.5	2
87.5 88 88.5	2	81.5 82 82.5	1
88.5 87 87.5	1	80.5 81 81.5 3	3
85.5 86 85.5	1	99.5 80 80.5	1
84.5 85 85.5	0		

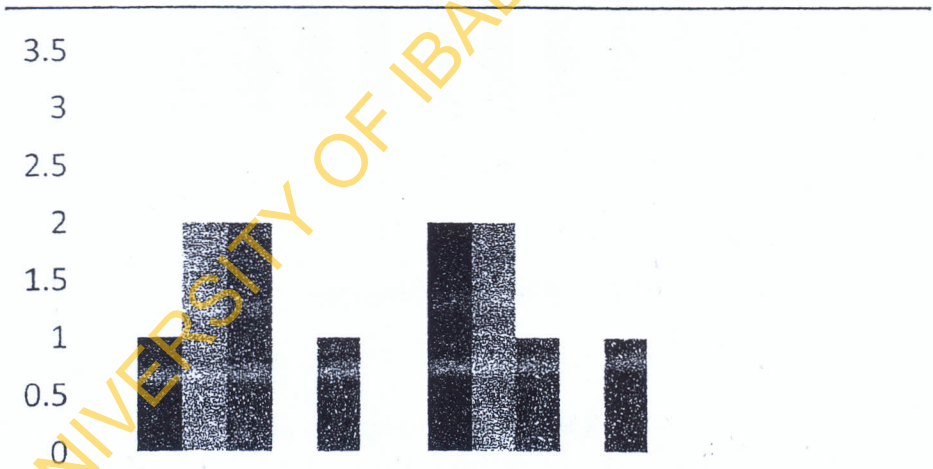


Fig. 3.5: Histogram of Students' Scores in Mathematics.

Frequency polygon

Frequency polygon can be shown on a drawn histogram. Consider the scores obtained by a group of students displayed in Table 3.8

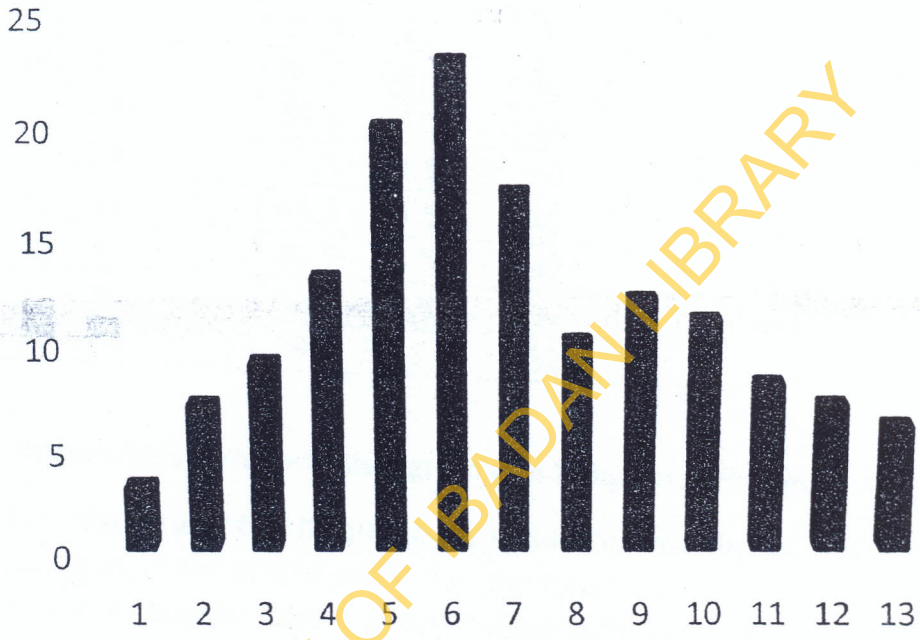


Fig. 3.6: Frequency Polygon of Students' Scores

Note the following:

i. Leave out one variable, e.g. year, item, score, etc., before you continue on the horizontal so as to accommodate the real limit.
ii. The frequency should be on the vertical axis.
iii. Frequency is spread over the real limits.
iv. The graph should form ¾ of the whole sheet.
v. Mark out the mid-point of each bar and join them together starting from the left-out space (variable).

Study the following table.

Table 3.8: Students' Scores in Basic Science

Class interval	Real limits	Mid-points	Frequency
95 - 99	94.5 – 99.5	97	3
90 - 94	89.5 – 94.5	92	7
85 - 89	84.5 – 89.5	87	9
80 - 84	79.5 - 84.5	82	13
75 - 79	74.5 – 79.5	77	20
70 - 74	69.5 – 74.5	72	23
65 - 69	64.5 – 69.5	67	17
60 - 64	59.5 – 64.5	62	10
55 - 59	49.5 – 54.5	57	12
50 - 54	44.5 – 49.5	47	11
45 – 49	39.5 – 44.5	42	8
35 – 39	34.5 – 39.5	37	7
30 – 34	29.5 - 34.5	32	6

Differences between Histogram and Frequency Polygon

For histogram, the frequency is within the real limits but with the polygon, it lies at mid-point, i.e. the frequency is evenly distributed in the histogram but it is not evenly distributed in the frequency polygon.

Measures of Central Tendency

The measures of central tendency are: (i) the mean, (ii) the median, and (iii) the mode. They are the various forms of averages that describe the group or sample concerned, e.g. the average mark of a class is the typical or overall performance of the class. Thus, the average is better than any form of data arrangement as it aids to determine the typical performance of a group. It also helps to compare individual performance groups and comparison between groups.

Mean of Group with Ungrouped Data

Assume we have the following scores for 10 student

18, 39, 21, 48, 50, 28, 41, 22, 34, 24.

Based on these figures, let us now work out the mean.

$$\text{Mean} = \overline{X} = \frac{\Sigma X}{N} = \frac{\text{Total Score Observed}}{\text{No of Observations}}$$

$$= \frac{18 + 39 + 21 + 48 + 50 + 28 + 41 + 22 + 34 + 24}{10}$$

$$= \frac{323}{10}$$

$$= 32.3$$

Sometimes, there are repetitions of scores. Consider the following:

Scores	X	F	FX
	23	1	23
	35	2	70
	46	2	92
	51	2	102
	60	3	180
	62	2	124
	68	1	68
	70	2	140
		$\Sigma f = 15$	$\Sigma fX = 799$

$$\text{Mean} = \overline{X} = \frac{\Sigma fX}{\Sigma f}$$

$$= \frac{799}{15}$$

$$= 53.3$$

When ranges are involved, the following procedure obtains.

Table 3.9: Ranges of Students' Scores in Mathematics

range	Mid-point	f	fX
55 - 59	57	2	114
50 – 54	52	4	208
45 – 49	47	10	470
40 – 44	42	16	672
35 – 39	37	23	851
30 – 34	32	10	320
25 – 29	27	10	270
20 – 24	22	5	110
		$\Sigma f = 80$	$\Sigma fX = 3015$

Mean $= \bar{X}$; $\quad \bar{X} = \frac{\Sigma fX}{\Sigma f} = \frac{3015}{80}$

$= 37.69$

Method of Assumed Mean

- We can use any of the midpoints to be the assumed mean.
- We subtract each of the scores from the assumed mean.

Here is an example, worked out.

Let the assumed mean be 37 $= Xo$

Table 3.10: Scores of Students in Mathematics

Class	X	(Xi – Xo) = d	F	F(Xi – Xo)fd
55 - 59	57	+20	2	+40
50 - 54	52	+ 15	4	+60
45 - 49	47	+10	10	+100
40 - 44	42	+5	16	+80
35 - 39	37	0	23	0
30 - 34	32	-5	10	-50
25 - 29	27	-10	10	-100
20 - 24	22	-15	5	-75

$\Sigma f = 80$

$\Sigma f(Xi - Xo) = +55$

$$\frac{\Sigma f(Xi - Xo)}{\Sigma f} = \frac{+55}{80} = 0.69$$

$$\bar{X} = \text{Assumed mean} + \frac{\Sigma f(Xi - Xo)}{\Sigma f}$$

$$= Xo + \frac{\Sigma f(Xi - Xo)}{\Sigma f}$$

$$= 37 + 0.69$$

$$= 37.69$$

Median of a Distribution

This is the middle score or point that divides the distribution into two equal parts. This score is not affected by the size of other scores in the distribution and independent of the extremes. It is concerned with the position of the score. It is not

affected by extreme scores, e.g. 4, 5, 7, 8, 11. It can be used where the distribution is skewed. In a skewed distribution, one can depend on the median rather than the mean because the mean of such a distribution gives a poor picture of a class. There are two main types of skew, the negative and the positive. Study the following two charts.

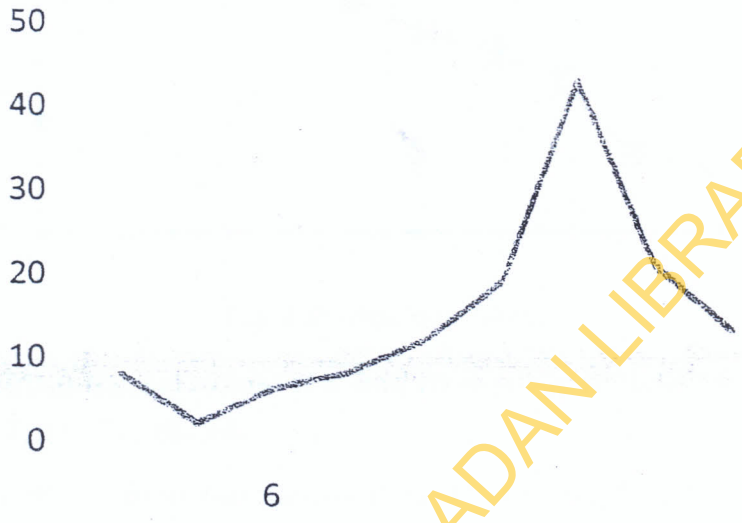


Fig. 3.7: Negative Skew

A negative skew occurs when very few students fail in any mastery learning exercise. However, oftentimes, in the school system, what we obtain most often is the positive skew where few students pass and many fail. This is well illustrated in the following chart.

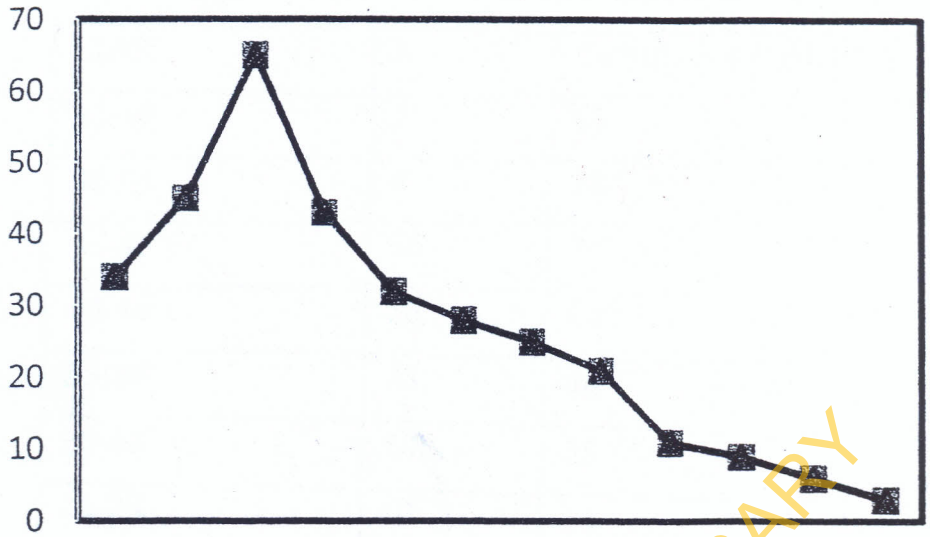


Fig.3.8: Positive Skew

To get the median when N = odd (N number of data or scores), e.g. 17, 13, 11, 18, 14.

Method: Arrange the scores in order of magnitude from the highest in the group, e.g. 18, 17, (14) 13, 11

Median = 14

To get the median when N = even, e.g. 4, 5, 7, 8, 10, 11. In this example, the median is the average of the two scores at the centre. The centre scores are 7 and 8.

$$\text{Median} = \frac{7+8}{2} = \frac{15}{2} = 7.5$$

To get the median where there is a repetition of scores, the following procedure can be used:

Table 3.9: Cumulative Frequency of Students' Scores in Math

CLASS	F	Cumulative Frequency
55-59	2	80
50-54	4	78
45-49	10	74
40-44	16	64
35-39	23	48
30-34	10	25
25-29	10	15
20-24	5	5

$$\text{Median, Md} = L + \frac{(N/2 - fb)i}{Fw}$$

Where N = Total no of scores

L = lower class limit where the median lies/falls

fb = sum of the frequencies below the lower limit.

N/2 = half of the total scores.

i = class size of class width.

fw = frequency of the class at which the median lies.

For the example being considered, the following would obtain.

N = 80

fb = 25

fw = 23

C = 5

$$Md = \frac{34.5 + (80/2 - 25) \times 5}{23}$$

$$= \frac{34.5 + (40 - 25) \times 5}{23}$$

$$= \frac{34.5 + (15 \times 5)}{23}$$

$$= 34.5 + \frac{75}{23}$$

Md = 37.8

Mode

This is the item or score with the highest frequency. In the scores given earlier, i.e. [10, 11, 12, 10, 10, 12, 10, 12, 12, 13, 13, 12], the figure 12 is the mode because it occurs 5 times as against the 4 times for figure 10 and 2 times for figure 13.

We could have a bi-modal distribution where there are two modal values. It is easy to obtain but less dependent, it fluctuates most often due to the factor of sampling.

Properties of Mean, Median and Mode of a Distribution

Example: If we have the following distribution: 4, 5, 7, 8, 11, the mean = $\frac{35}{5}$ = 7, which is also the median.

In another distribution like 4, 5, 7, 8, 19; the mean is different from 7, but the median is still 7. The mean as we know is the total of the figures, divided by the number of scores, i.e.

$$\frac{43}{5} = 8.6$$

Again, consider the following figures: 0, 5, 7, 8, 11.

The mean is $\frac{31}{5}$ = 6.2, but the median is not different.

We can therefore conclude that:

a. The mean is really the balancing point of distribution (that is the reason why it shifts) because the sum of deviation of the mean from each of the elements in the data will be zero.
b. It reflects all the scores in the distribution, i.e. a change in any one affects the mean.
c. It is most resistant to sampling fluctuation.
d. It is used in all other statistical analyses.

Measures of Variability or Dispersion

There are different measures of variability, namely:

a. The range
b. Semi-inter-quartile range
c. Variance
d. Standard deviation

The Range

This is a measure of variability and is the difference between the highest and the lowest scores in the distribution. Consider the following distribution: 8, 11, 12, 2, 9, 5.

The lowest number is 2 and the highest is 12, but the real limits are 1.5 to 2.5 and for 12, these 11.5 to 12.5. So the difference between 1.5 and 12.5, i.e. 11, is what the range truly is. Some texts would use the formula Highest – Lowest + 1, i.e. in this case 12 – 2 + 1.

Range takes into consideration only the two extreme scores and therefore it is not a true reflection of the variability of all the scores in the distribution. For instance the series, 8, 11, 1, 2, 9, 5, has the same range [of 11] as in the example above, whereas the figures are different. Again, all other scores may be changed provided the changes do not affect both the upper and the lower scores. Hence, we do not have a good reflection of the variability of all the scores in the distribution.

Semi-interquartile range

A distribution can be divided into four equal points and each point of the division as a quarter; Q1, Q2, Q3, and Q4 will represent each one.

The quarter deviation is the difference between Q3 and Q1. But the semi-quarter is the half of quarter deviation.

$$Q = \frac{Q3 - Q1}{2}$$

We use the formula for median to calculate the second quarter, thus

$$md = L + \frac{(N/2 - fb)\, i}{fw}$$

$$Q1 = L + \frac{(N/4 - fb)\, i}{fw}$$

$$Q3 = L + \frac{(3N/4 - fb)\, i}{fw}$$

Variance

Variance is the mean square deviation, i.e. taken from the mean.

The formula is $S^2 = \frac{\Sigma(Xi - \bar{X})^2}{N - 1}$

Or $S = \Sigma\frac{\{(Xi - X)^2\}^{1/2}}{N - 1}$ for $N \leq 30$

$\{N - 1\}$

Table 3.10: Deviation Squared

Xi	Xi – X = d	D ^{2}
3	-3	9
3	-3	9
3	-3	9
6	0	0
6	0	0
9	3	9
12	6	36

S^2 = variance

= 72/6

= 12

S^2 = $\sqrt{12} = 3.46$

There is computation formula for determining the standard deviation of a distribution.

$$S = \frac{\sqrt{N\Sigma X^2 - (\Sigma X)^2}}{N(N-1)} \quad \text{for } N \leq 30$$

X	X^2
3	9
3	9
3	9
6	36
6	36
9	81
12	144

$$S^2 = \frac{7 \times 324 - (42)^2}{7 \times 6}$$

$$= \frac{2268 - 1764}{42}$$

$$= \frac{594}{12}$$

SD = 3.46

Standard deviation and variance for distribution without frequencies

$$S = \frac{N\Sigma X^2 - (\Sigma X)^2}{N(N-1)}$$

For distribution with frequencies, we make use of the frequency formula, such as

$$= \frac{N\Sigma fX^2 - (\Sigma fX^2)}{N(N-1)}$$

This is illustrated in the following example.

Table 3.11: Cumulative Frequency Distribution of Maths Scores

Class	F	Mid point	Fx	Fx X Mid	Cum Freq
55 - 59	1	57	57	3249	50
50 - 54	1	52	52	2704	49
45 - 49	3	47	141	6627	48
40 - 44	4	42	168	7056	45
35 - 39	6	37	222	8214	41
30 - 34	7	32	224	7168	35
25 - 29	12	27	324	8748	28
20 - 24	6	22	132	2904	16
15 – 19	8	17	136	2312	10
10 - 14	2	12	24	288	2

$\Sigma X = 345$ $\Sigma FX = 1480$ $\Sigma fX^2 = 13965$

$$S^2 = \frac{N\Sigma fX^2 - (\Sigma fX)^2}{N^2}$$

$$= \frac{50 \times 49270 - (1480)^2}{50 \times 50}$$

$$= \frac{2463500 - 2190400}{2500}$$

$$= \frac{273100}{2500}$$

$$= 109.24$$

S.D = 10.45

Percentiles

A percentile point is a score that is arrived at by using either the ogive or the formula. In case of %, cumulative frequency curve (ogive) the frequencies are displayed as stages along the vertical axis, which means a distribution starting from 0 to 100%. For example, 10 percentile is a score that corresponds to 10% of the score, e.g. 16.5 from the graph. The rank of the score 16.5 is 10 or 0.1.

The rank can go from 0 to 1 or 1 to 100. It could also be written as

P.12 = 12^{th} percentile

Table 3.12: Cumulative Distribution Table

Score	Freq	Cum Freq	% cum freq
55 – 59	1	50	100
50 – 54	1	49	98
45 – 49	3	48	96
40 – 44	4	45	90
35 – 39	6	41	82
30 – 34	7	35	70
25 – 29	12	28	56
20 – 24	6	26	32
15 – 19	8	10	20
10 – 14	2	2	4

$$X = 1 + \frac{(N.P - Fb)\, i}{Fw}$$

Where X is the percentile point

L is the limit of the score or class

N = total no of the score

P = percentile rank from 0 to 1

Fb = sum of frequency below the lower and upper limit

Fw = frequency at no of the %tile point line

L = class size or interval

The formula for the percentile rank is

$$Pn = \frac{fw\,(X - 1) + fb}{Ni}$$

Find the score that corresponds to 60th percentile rank

$L = 29.5$

$NP = 50 \times 0.6 = 30$

$Fb = 26$

$Fw = 7$

$I = 5$

$$X = 29.5 + \frac{(30 - 28)}{7} \times 5$$

$= 29.5 + 2/7 \times 5 = 29.5 + 10/7$

$= 29.5 + 1.43$

$= 30.93$

2. Determine the percentile rank for a score 31

$X = 31$

$L = 29.5$

$Fb = 28$

$I = 5$

$$PR = \frac{fw(X-1) + ifb}{IN}$$

$= 7 \times 1.5 + 5 \times 28$

$$= \frac{10.5 + 140}{250}$$

$$= \frac{150.5}{250}$$

$= 0.602$

$= 0.6$

Application

Percentiles are ordinals in scale of measurement and can be used to:

a. find the relative position of each individual,
b. find the percentile of below or above cut-off points,
c. find the pass mark, that is if you want,
d. compare performance in different subjects over a period of time in a class.

Correlation

The word, 'correlation' can be divided into two, i.e. co and relation. Co means joint while relation means association. Hence correlation means joint association.

Height and weight are correlated, i.e. as height increases, weight should also increase. Similarly, time and temperature, I.Q. and achievement, age and fertility, area and radius of circle, etc., are positively related. However, age of a car and its resale value, force of attraction and distance, etc., are negatively related.

Examples of things that are not related are shoe size and income, weight and I.Q., height and I.Q., etc.

When two things are related, it does not mean that one causes the other. Rather, the implication is that the two have a tendency to move in the same direction.

There are two things to note about a correlation index:

1. Strength or magnitude of the correlation index.
2. Direction (sign either +ve or –ve) of the correlation index.

Uses of Correlation Index

a. Correlation index helps to predict performance of students. Thus, correlation is used as an instrument of prediction, e.g. if a JAMB examination for selection of students

for Engineering happens to correlate with their performance in the University examination in Engineering. Thus, JAMB is said to be valid for selecting students for Engineering. This sort of validity is called *predictive validity*.

b. It helps to ascertain the reliability of a test, i.e. how stable and consistent the student performance is with time over a test. We do this by Test-Retest method. This determines stability coefficient.

c. We use it also to test for equivalence. By this we are determining the coefficient of equivalence. Preparing two different tests, one as original and the second as parallel, to the old test. They should not differ in content but in structure and they should be measuring the same thing. Coefficient of equivalence tells us how comparable the two tests are.

d. We use it to determine *internal consistency* of a test. This could be done by split half reliability. This tells us how each item in a part is related to another item in another part of the test.

We will discuss two types of correlation, one non-parametric (Spearman Rank Order Correlation (r_s)) and the other parametric (Pearson Product Moment Correlation 'r').

A parametric test is used when we are sure that the data are normally distributed but non-parametric is used when we are sure that the data are not normally distributed.

Spearman Rank Order Correlation (r_s)

Condition for Use

The two sets of data must be in ordinal form or at least one of them should be in ordinal form. The formula is given as:

$$r_s = 1 \quad \frac{-6\Sigma D^2}{N(N^2 - 1)}$$

Where D = difference between pairs of ranks

and N = total number of pairs.

The order in which 15 students in a class submitted their test papers and the scores obtained by these students are as follows.

Table 3.13: Order of Submission and scores obtained

Student	Order of submission	Scores	Rank	D	D^2
A	1	28	9.5	8.5	72.25
B	2	21	14	12	144
C	3	22	12.5	9.5	90.25
D	4	22	12.5	8.5	72.25
E	5	32	6	1	1
F	6	36	3	3	9
G	7	33	5	2	4
H	8	39	1	7	49
I	9	25	11	2	4
J	10	30	8	2	4
K	11	20	15	4	16
L	12	28	9.5	2.5	6.25
M	13	31	7	6	36
N	14	38	2	12	144
O	15	34	4	11	121

$$r_s = 1 - \frac{6\Sigma D^2}{N(N^2 - 1)}$$

$$= 1 - \frac{6 \times 773}{15(225-1)}$$

$$= 1 - \frac{6 \times 77}{15 \times 224}$$

$= 1 - \frac{4638}{3360}$

$= 1 - 1.38$

$= -0.38$

The meaning of negative correlation is that as one variable is increasing, the other variable is decreasing. It means that those who submitted later perform better than those who submitted earlier. The correlation is not a perfect one because it is less than 1.

Pearson Product Moment Correlation 'r'

The two sets of data should be at least in the interval form; it then means that we cannot use it for ordinal and nominal forms. The association between the variables should be linear, e.g. time spent on a journey depends on the distance.

The formula is

$$r_{xy} = \frac{N\Sigma xy - \Sigma x\Sigma y}{\sqrt{N\Sigma x^2 - (\Sigma x)^2\ N\Sigma y^2 - (\Sigma y)^2}}$$

Consider the following scores obtained on two subjects.

[X = Physics; Y = Mathematics]

	X	Y	X^2	Y^2	XY
	3	8	9	64	24
	2	7	4	49	14
	4	7	16	49	28
	5	10	25	100	50
	1	5	1	25	5
	7	11	49	121	77
Total	22	48	104	408	198

$$r_{xy} = \frac{N\Sigma xy - \Sigma x\Sigma y}{\sqrt{N\Sigma x^2 - (\Sigma x)^2}\ N\Sigma y^2 - (\Sigma y)^2}$$

$$= \frac{6 \text{ X } 198 - 22 \text{ X } 48}{\sqrt{(6 \text{ X } 104 - 484)\ (6 \text{ X } 408 - 2304)}}$$

$$= \frac{1188 - 1056}{\sqrt{(624 - 484)\ (2448 - 2304)}}$$

$$= \frac{1188 - 1056}{\sqrt{(140)\ (144)}}$$

$$= \frac{132}{\sqrt{(20160)}}$$

$$= \frac{132}{141.99} = 0.9296$$

$$= 0.93$$

The correlation is very high, since it is positive. It shows that the 2 sets of scores are linearly related, i.e. the increase of scores in Physics is as a result of the students' knowledge of Mathematics.

Relationship between r and r^2

'r' determines the relation between the 2 variables while r^2 is known as coefficient of determination which gives the amount of variance in the dependent variables explained by the independent variable.

e.g. $r_{xy} = 0.92$, $r^2 = (0.92)^2 = 0.8464 = 0.85$

If $r^2 = 0.85$, then 85% of the variances in the Maths score is due to students' ability to spell in any situation.

e.g.

r	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
r^2	0.01	0.02	0.09	0.16	0.25	0.36	0.49	0.64	0.81	1.00

When r ranges from 0.1 to 1, then r^2 ranges from 0.01 to 1.

From the above, it is clear that the amount of variance explained by a correlation of 0.7 is below 50%.

i.e. if $r < 0.7$, then $r^2 < 50\%$

Thus, whether *r* is –ve or +ve, it does not affect the value of r^2.

CHAPTER FOUR

EVALUATION OF STUDENT ACHIEVEMENT TESTS

Nature

It should, first of all, be pointed out that the purpose of evaluation is to make a sound judgment about the worth or quality of an educational programme. Of course, this can be said also of other types of programmes, including workers' or students' performance. That is what most evaluators attempt to do when students' achievements are to be evaluated. However, the goal is not only to describe what the students or employees can do but also to seek answers to such questions as "How good is the level of students' achievement?" "How good is the educational programme?" "Is it worth it?" These are called value questions that require the exercise of judgment.

It would rather be professionally unethical to teach without evaluating the extent of learning. So, achievement tests are designed to measure the effects of a particular programme of instruction or training. They are described as terminal evaluation of students' status on the completion of a course of study. Achievement tests are used as a predictor of future learning.

It is better to say that tests usually work better than subjective judgments. Since teachers are the greatest users and producers of tests, if tests were not available, the only way would be to adopt the subjective evaluation of teachers and students.

Some educators, who might suggest achievement tests are not necessary in schools probably because they have become frustrated by the disadvantages of tests, are not likely to argue that evaluation is not needed. From the foregoing, it stands to reason

that if achievement tests were abandoned, other means of assessing students would have to be used. Thus far, no replacement has been discovered. This, by implication, shows the prominent position of achievement tests in the evaluation process. Therefore, the functions of achievement tests include the following:

a. Achievement tests measure students' achievement. They therefore contribute to the evaluation of educational progress. They help teachers and instructors to assign meaningful and accurate grades. In assigning these grades accurately, teachers and instructors should take up the responsibilities very seriously because the grades are reported to students and their parents in order to show how effective they have been in the classroom.
b. The above function now leads us to the next relevant one, which is the provision of information on the teacher's quality of instruction. Achievement tests have a telling effect on the teacher's quality of instruction if his tests are badly constructed or if he does not teach what he is testing, or if he fails to sample adequately what he is supposed to sample.
c. Another important function of achievement tests is to motivate and direct student learning. Experience has shown that students tend to study very hard when an examination is drawing near, and emphasis is always on those areas that are likely to be tested if they are aware of the fact that they will be tested. If they are aware of the fact that they will be tested and the tests also measure the essential course objective, certainly the students will be highly motivated. A well kept progress chart in the class can also motivate the students especially if the information can be used to counsel or direct individual students concerning their performance. It can also be a source from which the learner can get direct information about how he is fairing in the class.
d. Achievement tests also serve other useful educational functions. One of these is in the process of building them,

it should cause instructors to think carefully objectively about the objectives of instruction in a course. The building process should enable the instructors to define their objectives operationally most especially in an achievable or measurable way.

e. Achievement test results also serve as useful information for administrative judgments. For instance, at the end of each academic year, students are normally promoted to the next class while those who do not attain the expected standard are retained in the same class. Some of the students may even be asked to withdraw from the school. All these decisions cannot be made without test results. Administrative judgments are made about students on the basis of students' performance in various tests.

Characteristics of a Good Test

A committed and dedicated teacher is supposed to have established that the test he wants to administer on the students possesses the characteristics expected of a good test. He should do this before any attempt is made at administering a test on his students. The characteristics are discussed under the following headings:

i. Validity
ii. Reliability
iii. Discrimination
iv. Difficulty level
v. Efficiency or economy of the test

Validity

Validity refers to how accurate a test measures what it purports to measure. To put it more learnedly, it refers to how accurate a test measuring what is supposed to measure. In other words, we ask – does the test measure all we want it to measure and nothing but what we want it to measure? If for example a test is designed to measure intelligence, that test should measure intelligence only. However, if the test measures knowledge of Maths instead of in-

telligence, the test would not be valid. But if the test measures intelligence and nothing more, then it is said to be valid. It should be pointed out that validity refers to the test result and not the test instrument itself.

Another important factor relating to validity is that it is specific to some specific purposes. And this should explain why we have types of validity such as content validity, predictive validity and construct validity.

Reliability

Reliability is another characteristic of a good classroom test. It refers to the precision and the consistency with which a test instrument measures a sample of behaviour. Learnedly put, it is used to describe one of the most significant properties of a set of test scores, that is how consistent or error free the measurements are.

The foregoing implies that when a test instrument is used, it may not be possible to have exactly the same test result all the time but an individual should remain nearly the same in repeated measurements if the test produces a reliable result.

A very good example of this is a situation where a teacher gives a test to his students in English Language this week. He scores them and keeps their scores. He may again give the same test material or a similar one to the same set of students the following week. Again, he scores the students on the test. He can now compare the results obtained from the tests at the two instances, in order to establish whether or not the patterns of distribution of scores are the same for the two tests. He may want to know if the two tests agree since scores by some students are very good, good, average, poor and so on. The relationship between the two sets of test results is often discussed in terms of 'correlation coefficient'. Like validity, reliability is a matter of degree, and thus we talk of high or low coefficient or reliability.

Method of estimating reliability

The following are the different methods of estimating reliability:

i. Test-retest method
ii. Equivalence / parallel form method
iii. Split-half method, and
iv. Kuder-Richardson method

i. *Test-retest method*

This method is a measure of stability. In test-retest method, the same test is repeated on the students on two different occasions. The time lag between the two administrations may be some hours, days or weeks. The resulting test scores are then correlated. The correlation coefficient obtained provides a measure of stability of the test.

ii. *Equivalence / parallel form method*

This is another method of estimating reliability and it involves giving the two forms of the test to the same set of students within a given time. A teacher or an evaluator may decide to correlate the two forms of the test and the correlation co-efficient obtained gives a measure of equivalence. This shows the degree to which both forms of the test measure the sample of behaviour.

It should be pointed out that if the correlation coefficient is high then it implies that both forms of the test measure the same content equivalently. If, however, the correlation coefficient is low, it implies that the two forms of the test do not sample the content in the same manner with equal difficulty.

iii. *Split-half procedure*

In the split-half method of estimating reliability, test is administered to the students once, and then divided into two equal halves for *scoring* purposes. This therefore affords the test administrator the opportunity of getting two scores for each student and ultimately two sets of scores from a single test. The test administrator splits the test into halves in order to get the even and odd numbered test items, which are scored as if they are from two separate tests and the results are correlated to get a measure of internal

consistency. This shows the extent to which each half of the test represents one another.

However, if the correlation between the two sets is high, it is concluded that the test as a whole possesses a high internal consistency. On the other hand, a low correlation between the two halves implies that the test possesses a low internal consistency. Now, if an evaluator decides to estimate the reliability of the scores based on the whole test (full length test) he may adopt the Spearman-Brown formula. This is shown below:

$$\text{Reliability on full test} = \frac{2 \times \text{Reliability on } \frac{1}{2} \text{ test}}{1 + \text{Reliability on } \frac{1}{2} \text{ test}}$$

It is necessary to reinstate at this juncture that, like equivalent form procedure, split-half procedure does not take into account changes in individual from time to time.

iv. *Kuder-Richardson method*

This is a method, which makes use of formulas developed by Kuder and Richardson. The formula includes Kuder-Richardson formula 20 and Kuder-Richardson formula 21. This procedure basically assumes that the items in the test are homogenous. In other words, the items resemble one another in quality and characteristics. The basic thing done here is to find the correlation between one item and every other item in the test.

However, if it is discovered that an item is not consistent with the others, such an item may be discarded. Kuder-Richardon formula is applied to the result of any test which has been scored on the basis of the number of correct answers. The formula is indicated below:

$$\text{Reliability Estimate (KR21)} = \frac{K}{K-1}\left[1 - \frac{MK-M}{K5^2}\right]$$

Where

K = the number of items in the test

M = the means of the test scores
S = the standard deviation (a measure of the spread of scores) of the test scores

Discrimination

It is a well known fact that most of the classroom achievement tests are norm-referenced. Norm-referenced tests are those that measure differences among students. In other words, it involves comparing a person's score with the average score of some relevant group of people. Such tests would be expected to separate the students into various groups according to their respective abilities. Such tests should be able to discriminate between those students who are above average, just average, and below average.

However, it should be stressed that all the students cannot reach the same mastery level at the same time irrespective of the teaching adopted by the teacher. This probably should further explain why achievement test must possess the quality of being able to discriminate among the students.

Difficulty Level

This is another quality of a good achievement test. It refers to how appropriate is the test in the sense of its difficulty with respect to the testees. With respect to the quality of a good test, the test administrator may want to know how many of the students can answer the test items correctly. Test experts consider such factors as age of the students, their level and purpose of the test in fixing the difficulty level of an achievement test.

Ideally, we expect that a test should not be too difficult or too easy for the students. It should be such that most students would be able to score about 50% if they have been thoroughly and adequately prepared. If however the difficulty level is high, very few students will be able to score up to 50%.

Efficiency (Economy)

This is another important characteristic quality of a good achievement test. Under this, we consider the time taken, availability of equivalent form of the test, use of interpretation and application of the test result, use of administration of the test, and the cost of planning, constructing, administering and reporting the test results.

These are some of the qualities which an efficient and economic achievement test is expected to possess.

Types of Achievement Test

Test experts have classified achievement test using different parameters. Whereas some classify tests on the basis of the behaviour that is being measured, others classify them considering the types of items contained in the test, the purpose of tests, etc. However, achievement tests may be classified on the basis of the essay-type and the objective type.

Essay type tests

The essay test has been a very popular type of achievement test. Its usage started earlier than 2300 BC in ancient China. It is described as an attempt at answering questions in the form of continuous connected writing in which the candidate is free to express himself in his own way. The two main forms of essay include the following.

The extended response

In this type of essay, the students answer a small number of items. The student is expected to organise his answers and express in his own words – in a continuous connected form. A good example of this is when candidates are asked to answer 4 questions out of 6 in 2½ or 3 hours. In this case, the examiner carefully reads the essay and awards a mark on the basis of the impression he has about what the testee writes. An example of the kind of question we have under this is:

a. Discuss in detail the factors contributing to the downfall of Okonkwo.

b. Examine the advantages and disadvantages of the civilian regime in Nigeria.

The Short-answer Essay

In this kind of test, the candidate is given a number of topics or questions and he is asked to write briefly on them. It limits both the content and the type of learner's response. The following is an example of such a question.

a. Mention any four factors affecting learning.
b. Write about four sentences on any 2 of them.

Here, learners' opportunity to demonstrate such behaviour demanding greater freedom of response is extremely limited.

Advantages

i. Essay tests are very easy to prepare. An essay test may contain up to two to twelve items as against what obtains in objective tests, especially multiple choice items, each of which is likely to have four alternative answers.
ii. An essay test requires the testees to express himself freely in his own words.
iii. It can be used to measure higher levels of cognition. In other words, students can be asked to analyse, synthesise, apply and evaluate. This implies that when a teacher intends to measure the higher levels of cognition, an essay test is the best to use.
iv. It encourages students to learn how to organise their ideas logically and coherently.

Disadvantages

i. Essay tests are not easy to mark even when marking schemes are used. The major problem a scorer encounters in this regard is subjectivity and/or objectivity in scoring.
ii. It is time consuming to score, unlike objective tests that are very easy to score.
iii. It does not cover a large field of knowledge.

iv. Essay tests are subject to bias in scoring. The teacher is influenced by some factors, which include good handwriting, fatigue and halo effects.

Construction of Essay Tests

Let it be pointed out that the procedures for constructing essay tests are almost the same as those used in the construction of objective tests. The procedures include the following.

a. The instructional objectives should be stated in specific terms
b. The course content should be well outlined.
c. Prepare the table of specification (test blueprint).
d. The test items should now be written according to the table of specification.
e. The essay question should be written in such a way that the task is well defined instead of throwing the students into the ocean of confusion.

Again, it should be stressed that the difference between the procedures for constructing essay and objective tests is that the table of specification for the essay type of test does not contain many items like what obtains for objective testing. Another major difference is in the area of the writing of the essay test items that are different from that of the objective test.

General Guidelines for the construction of essay questions

i. Questions that can easily be answered within a limited time should be considered.
ii. Options or choices could be allowed, but this should be within reasonable limit so that comparability of students' achievements would be possible.
iii. While the test items can be increased, the length of answers expected for each item can be reduced.
iv. As much as possible, asking for reproduction of facts from textbooks or lecture notes should be avoided. This is especially when students in the upper levels ar involved.

v. Essay items should not be ambiguous. The testees should be able to read and understand the items. They should not have to resort to guess work as to what the questions demand of them.

vi. Items or questions that are too general should be avoided. A good example of an item that is too general is – 'Discuss the essay type of testing'; or 'Discuss the concept of demand and supply'.

vii. Items that are too difficult or too easy should not be considered at all. The difficulty level of the items should be in tune with the age and the level of the students.

viii. Testees' response to the items should be marked using a well prepared marking scheme.

Scoring the Essay Test

In this section, we shall consider two different procedures for scoring essay tests. This exercise should minimise the problem of subjectivity and unreliability in the scoring of essay answers. The two procedures are discussed below.

(i) The scoring key procedure

With this procedure, a teacher is expected to prepare a marking scheme or a scoring key. This is an outline of the expected answers to the items, indicating the marks or points allotted to each item. Such marking guide is drawn before any of the scripts is marked. The teacher answers the questions himself and with this, he draws out the possible points for which scores are then allotted. He should make sure that all possible/alternative answers are thought of, and that the relative weight of each point is taken into consideration in allotting marks to the points. The scorer should adhere strictly to the marking points.

(ii) The impressionistic marking

In this procedure, the scorer is expected to mark an overall judgment of the quality of a student's attempt. What he does is to read through, and taking all factors into consideration, decide what

score to award. He should not be influenced by any halo effect. He is expected to score very objectively and thoroughly.

There is a pertinent question of removing or seriously minimising the problems of subjectivity associated with the scoring of essay scripts. For this, consider the following options.

a. The number of markers can be increased, and the average of their ratings is established to represent the testee's score. However, this is possible only where there are more than one expert handling the course and both or all of them can competently score the scripts.
b. If only a marker is involved in the scoring of the essay scripts, then he should do the scoring twice. Or putting it another way, he should mark first and then, a few days later, he should revise the scoring.
c. If the essay test consists of several items, the scorer may score the answers to a particular item throughout all the scripts, before then proceeding to the next item again through all the scripts, and so on.
d. Each answer script should be scored without bothering to see the name or number of the testees.

The Objective Type

The scoring procedure for objective items is clearly stated in advance of testing and this is why it is called objective. The student is required to select only one response out of about four alternatives or supply a missing fact or term, which is an answer to the question. However, there are two main reasons why the objective test is really objective. First is that the guidelines or rules for scoring are very clear to both the teacher and the testees. The second is that, because only one answer is correct in each item, scoring is thus not heavily prone to the teacher's subjective judgment. The testee is therefore required to select only that answer which he considers is correct while others are wrong.

Types of Objective Tests

Objective tests are of various types. It may then be necessary to examine the major ones that can be useful in the area of classroom testing. So, the prominent types of objective tests include:

a. The true-false type
b. The fill-in type
c. The matching type
d. The multiple choice type

Now, let us examine each of these briefly.

The True-False Type

In this type of test, the testee is given some statements to which he should respond. The statements are framed such that some are correct while others are wrong. The statements have to be marked as either 'True' or 'False'. Consider the following examples:

Mark 'T' or 'F'	Statements
.....	i. Ibadan is the capital of Osun State of Nigeria.
.....	ii: Liberia is a country in Africa.

From the examples, you would see that teachers would find this type of questions easy to construct. For the student too it is fairly easy to answer them. However, there are several shortcomings. Chief among these is that this type of questions allows for a lot of guess work. In fact, if a student who doesn't have any idea of the topic decides to guess his way through all the items, he has a 50% chance of securing a pass.

The Fill-in Type

This is another type of objective test, which is most useful especially when it is necessary that testees must learn some dates, names, items figures or facts. The teacher should first get several statements based on the issues he has taught students. He then

deliberately removes some key words from each statement and requests that students should provide such missing words. Care should be taken to ensure that the words removed are those that carry meaning relevant to the topic being examined. So, we are not talking of such functional words like 'of', 'for', 'under', etc., but key words like, for instance in Geography, 'valley', 'hill', 'mountain range', etc. The principles guiding the construction of the items include the following:

i. The best thing is to use at most one or two blank spaces within each item.
ii. Only one term should sensibly complete the statement.
iii. Only the important terms should be left blank.
iv. The blank space should be placed near the end of the sentence.

The Matching Type

Though the matching type is fading out in the evaluation of students' achievement, it is still useful especially at the pre-primary and primary levels. In this type of objective testing, the testees are presented with two columns, each consisting of a list of names, facts, places, etc. Each one in the first column can logically be matched with another in the second column. The student's task is to identify which item in Column A goes with which in Column B. If it is well designed, it is a good test of establishing the testee's mastery of the subject matter, for without a sound mastery of the content, a testee would not be able to identify which item corresponds with which. Let us consider the following example which is set to test a learner's knowledge of the capitals of the various states in Nigeria.

Capital cities	States
A. Ilorin	1. Niger State
B. Ibadan	2. Rivers State
C. Lagos	3. Kwara State
D. Minna	4. Ekiti State
	5. Kano State
	6. Oyo State
	7. Lagos State

For this type of test, the following should be noted carefully:

a. The items in the right column should relate to those in the left column. In the present case, the relationship is geographical, those on the left being capital cities of states listed in the right column.
b. The test instructions should clearly state how the matching exercise is to be performed.
c. The matching should be done neatly such that it will not be clumsy.
d. The list containing the answers [i.e. the right column] should contain a few more items than those on the left. This is because it is a little more difficult to guess one's way to a pass if one does not master the content.

The Multiple Choice Type

Of the four major types of objective tests, the most widely applicable is the multiple choice type. This type of test is used to measure simple learning outcomes and learning tasks involving knowledge, understanding and application. In this type, the testee is required to choose one alternative response to a stated problem. Consider the following examples:

1. Chief Obafemi Awolowo, the Nigerian stateman, died in _____

   a. 1960 b. 1980 c. 1987 d. 1991

2. In which year did the elder statesman Chief Obafemi Awolowo die?

   a. 1960 b. 1980 c. 1987 d. 1991

It would be noticed that for the first example, a gap is given and this is to be filled in by choosing one of the options in alternatives 'a' to 'd'. In the second example, the same content is tested, but this time by giving a full statement question which can be answered by choosing one of the alternatives in 'a' to 'd'.

When constructing multiple choice items, the following rules should be carefully adhered to.

i. There should be a very clear problem in the stem of the item. Note that the stem of an item refers to the part containing the problem either as a question or an incomplete statement.
ii. The problem should clearly point to the theme of the correct alternative.
iii. The alternatives that are incorrect should be reasonably related to the problem.
iv. The alternatives should be grammatically consistent with the stem of the item.
v. Efforts should be made in making correct alternatives not to be consistently different in appearance from correct alternatives.
vi. The correct alternatives should be so worded that they are thoroughly and clearly plausible.
vii. The alternatives for each item should be randomly ordered.
viii. The incorrect alternatives should be framed such that they are thoroughly wrong but reasonable enough to catch the interest of the weak testees.
ix. The items should be worded in very simple, clear and correct language.

x. Alternatives such as 'None of the above', 'Both A and B', or 'All the above', should be strictly avoided.
xi. Grammatical cues and sentence structures that give away the correct alternatives should be avoided.
xii. In the problem statements, negatives should be sparingly used.
xiii. The alternatives should be homogenous in relation to the knowledge that is being measured.

It should be noticed from the examples given earlier on that each multiple choice item consists of three parts, viz:

a. The stem; b. The key; and c. The distracters.

The stem contains the question to the testees, e.g.

In which year did the elder statesman Chief Obafemi Awolowo die?

That contains, really, the kernel of the issue. It is the content that is being tested. Sometimes, this question part is also called the stem or the rubric. Even when this part of the item is stated as an incomplete statement, which then has to be filled in by choosing one of the alternatives, it is still correct to describe it as the question part, or the stem or rubric.

Next, the key is the correct alternative, to be chosen from among 'a' to 'd'. It is thus one of the alternatives.

The distracters are the alternative responses other than the correct answer. The function of the alternatives is to distract the attention of the testee from the correct answer. So, they should be almost as correct as, or as attractive as, the key itself. Now, it requires some experience and expertise to draft credible alternatives that would carefully tuck away the key answer among the distracters, without giving away the key.

Advantages of Objective Test Items

i. They are 'objective' because they require the pupil to select correct answers from given options. Besides, in scor-

ing them, there cannot be any room for subjectivity from the part of the teacher.

ii. They efficiently measure knowledge of facts.

iii. The teacher can use this method to cover a very wide field of knowledge, something which is not very possible with essay type of questions.

iv. They are very easy to score.

v. It is very easy to determine its content validity and the possibility of a wider coverage because they require the writing of many questions.

Disadvantages of Objective Test Items

i. They are not easy to prepare.

ii. They are time consuming to prepare.

iii. They cannot be used to measure how students organise their thoughts.

iv. They cannot be used to measure the higher order behaviours such as analysis, synthesis and evaluation.

Test Administration, Scoring and Item Analysis

Administering the Test

It has been pointed out that multiple choice items are very difficult to construct. If therefore we have a well-constructed test, this should be administered under perfect conditions in order not to render it invalid. The teacher must provide a favourable atmosphere for obtaining testees' good responses. The testees must be given a fair chance to demonstrate their ability to achieve significantly.

The test constructor, invigilator or the teacher administering the test should note that provision is made for the following:

a. Clear and unambiguous instructions;

b. Removal of all things that can distract the testees' attention;

c. Adequate work space for all;

d. A room with proper light and ventilation;

e. A quiet environment;

f. Reduction of anxiety;
g. Comfortable seats and tables;
h. Adequate motivation to perform;
i. Lack of threat from the teacher or invigilator;
j. Avoidance of unnecessary talking before or during the test;
k. Giving hints to testees who ask about individual items should be avoided; and
l. Cheating must not be allowed.

Scoring the test

It has been pointed out earlier that one of the advantages of the multiple choice test is that it is very easy to score, most especially if adequate instructions had been given on the question paper. For each item, only one possible answer is provided. If it is truly a very good objective test, a scoring key is available indicating the answer for each item. The most plausible way of obtaining total score for each student is to simply count the number of correct answers.

The common practice in Nigeria is that teachers assign marks to the number of correct answers. For example, in an objective test of 100 items, if a testee attempts all of them but gets 60 right, he then has 60 out of 100 marks. However, a controversial aspect of this issue is a situation where a testee might have some correct answers through guess work, especially in True/False items, with only two options, and in which one is correct and the other is wrong. Experience has shown that testees also do guess work in multiple choice items. In the light of these, it has been argued that testees should be penalised for such guesswork.

If a teacher wants to penalise students for guessing, there are two formulae that can be used. The first is for True/False items while the second is for the multiple choice test where there are usually 5 options. Let us consider the following example:

i. If there are 80 items of True/False with only two options and a student has 60 correct. In order to correct for the possibility of guessing, the following formula can be used:

$$R - \frac{W}{2-1} = 60 - \frac{20}{2-1} = 40$$

This means that the student who got 60 out of the possible maximum of 80 in the True/False test will now have 40 after the correction for guessing.

ii. Again, if a multiple choice test has 80 items and the student scores 60, the same formula applies; but then study the final result this time.

$$R - \frac{W}{5-1} = 60 - \frac{20}{5-1} = 60 - 5 = 55$$

Why different final scores? The difference in the final score is that the penalty for the student in the easier T/F test is more severe than it is for the student in the more difficult multiple objective test. The fact is that there are more chances for guessing in the T/F test, with fewer distracters, and so the penalty there is more severe.

As for the formula itself, the following is the key:

R = is the number of correct answers by the testee;
W = is the number of wrong answers;
N = is the number of alternatives or options given.

Test Analysis

On the basis of what has been advanced thus far on the multiple choice items, it is very clear that it is expedient to appraise the effectiveness of individual test items. This process is described simply as ITEM ANALYSIS. This is carried out after the test has been administered and scored. Items are analysed qualitatively and quantitatively. If this is done qualitatively, the content and form of each item are seriously considered whereas statistical properties are considered for quantitative analysis. In other words, content validity is one of the activities in qualitative analysis while item validity, item difficulty and discrimination power are done under the quantitative analysis.

Irrespective of the purpose of a test, its adequacy depends mainly on the care with which the items in the test have been selected. The foregoing implies that in order to select the items objectively, it is mandatory to go through item analysis procedure in a very careful manner.

Steps in item analysis procedure

The following steps should be considered in item analysis procedure:

i. Each of the testees should be scored on the test. This could be done in the following manner:
   a. The incorrectly answered items are marked. The marks are added up and scored.
   b. The correctly answered items are marked. The marks are added up and scored.

ii. The scripts are arranged in descending order, that is, in order of merit, putting the highest first and the least below.

iii. You can now multiply the number of pupils who took the test (N) by 27% and round up the result to the nearest whole number.

Let us consider the following example:

$$N = 50$$

$$= N \times \frac{27}{100}$$

$$= 50 \times \frac{27}{100}$$

$$= 13.5$$

iv. The total obtained from step 3 should be the number of scripts to be counted from the top, and this forms the high group denoted with R_H or H_H.

v. For this step, count the same number of scripts from the bottom which forms the low group denoted with R_L or W_L.

*Please note that the middle 46% are not used for item analysis purposes.

On the basis of the steps highlighted above, the discrimination power and the difficulty level of each item can be calculated thus:

a. Discrimination power = $R_H - R_L$.
   Note that the larger the difference between R_H and R_L, the higher the discrimination will be.
b. Item difficulty = $\frac{\text{Total number of examinees right}}{\text{Total number of examinees}} \times 100$

The value of difficulty level obtained depends on the way the scripts are scored, that is, whether you mark and record wrong or right responses.

Item Selection

For classroom achievement tests, the items to be selected must be good under discrimination and item difficulty. As a matter of fact, the type of items selected in a test depends largely on the use to be made of the test.

It is expected that good test items should have discrimination indices of 0.30 or more. On the basis of item difficulty, items around 50% difficulty level should be selected; that is, between 40% and 60%. It is imperative that classroom tests have all the items with none of them too easy or too difficult.

Interpreting the test results

Another important stage in this chapter is grading and interpreting the raw scores. After scoring the students' scripts, the results obtained are usually called raw scores. It should be pointed out that these scores are meaningless until a grade and evaluative interpretation are given to the performance of each student. This attempt is necessary in order to make the scores meaningful either for advice or administrative decisions.

The raw scores are often subjected to statistical analysis in order to make the test results serve various purposes. The foregoing should

simply explain why such terms as 'very poor', 'poor', 'fair', 'very fair', 'good', and 'very good', are used. Other evaluative terms that could be used to describe the position of each student in a test are: 'excellent', 'distinction', 'credit', 'merit', 'pass', 'fail'.

Another way of describing the students' performance is by giving each student a position, starting from the highest to the lowest. However, where two or more students have the same scores, the same position is given to such students but the next score will take into consideration the number of students that have the same score. After this, the scores are re-arranged in such a way that the student that comes first should be at the top, followed by the second, etc., until all the students are re-arranged according to their performance or standing in the test. This ranking enables each student to know how he has featured in the subject and his performance in relation to his classmates.

CHAPTER FIVE

EVALUATION INSTRUMENTS

Introduction

One of the duties of a teacher is to determine the extent to which the students have mastered what they were taught. The only way he can achieve this is through the use of some evaluation instruments. So, in this chapter, we present various types of instruments that may be used to collect assessment information on the students. These instruments are not limited to the students only; they can be used on a wide variety of people. Attempts are made to describe such instruments as *test, observational techniques, expert judgement, questionnaire, socio-metric techniques*, etc.

Test

Attempts are made here to discuss the meaning of test, types of test, classification of tests, and purposes of tests.

Meaning of Test

According to Onasanya [1990], test refers to any series of questions or exercise or other means of measuring the skills, knowledge, intelligence, capacities or aptitudes of an individual or group.

It is our belief that test can be referred to as a collection of items that can be used to determine whether or not the desired changes (or outcomes) have occurred in a learner. The desired changes could be at the level of intelligence, aptitude, skills, knowledge or attitude.

Types of test

There are a lot of different tests. Majority of them will be discussed in this chapter and some would be discussed later. The tests are

oral test, achievement test, performance test, psychological test, teacher made test, standardised test, formative test, summative test, etc.

Oral Test

This type of test involves verbal communication between testee (examinee) and the tester (examiner). The testee is expected to respond to sets of questions posed by the examiner. The information provided by the testee may indicate his level of knowledge on a particular issue. The test can be administered under a formal interview (to secure a job) or in an informal situation. This test provides an on-the-spot assessment of the testees' knowledge. The examiner could provide immediate feedback. One of the limitations of oral test is that it is prone to subjectivity. Whether the testee is right or wrong depends on his ability to convince the examiner. Again, it is possible for the testee to know all the questions but one in the list of the examiner, and that may be the only one administered on him. Oral test does not provide sufficient ground for comparability of standard in a classroom because it is difficult to maintain equal length and difficulty level of questions to each testee.

Achievement Test

An achievement test is a test that is used in determining the extent to which a testee has mastered or achieved some previously taught concepts or subject. This type of test helps the teachers to give more valid and reliable scores, which are used to summarise a comprehensive evaluation of students' success. The results obtained using achievement is a function of such factors as the characteristics of testees, teachers, home and the school.

Achievement tests are good in evaluating students and the curriculum. They also provide information about the level of testees' academic ability. Achievement tests are useful for selection purposes (job placement, promotion, and certification, etc.). Achievement tests are useful in evaluating teaching and learning processes. For

example, poor achievement level of students may result from bad or wrong teaching style or method used by the teacher.

Performance Test

This is a type of test designed to measure some skills – manipulative ability of the testees who are expected to manipulate some concrete materials. An example is a type of test given to some secretarial students in order to determine how good they are on the typewriter. The strength of the test includes the fact that it is useful in evaluating the technical know-how of the testees. However, a weakness of the test is that it is time-consuming. For example, the authors tested students' ability in the use of the computer for data analysis. There were 40 students and we had only 10 computer sets. So we had to divide the class into four groups. If the test is for 3 hours, the examiners would have to invigilate for 12 hours. Alternatively, we should have bought more computers to go round. And that leads to the second weakness of this type of test: that it is expensive.

Teacher made Tests

This is a locally constructed instrument by the teacher, based on his teaching. It is a type of an achievement test, where a teacher wants to determine the level of mastery by his students. This test could be valid and reliable but without possessing a table of norms. It has the strengths and weaknesses of an achievement test as well as the uses of an achievement test.

Standardised Test

Standardisation implies uniformity, unlike what we have with the teacher made tests. A standardised test is both valid and reliable. Moreover, it has a table of norms. For a teacher made test to be standardised, a sample greater than the teachers' class students or even school students will be involved in establishing its validity, reliability and table of norms. Like the teacher made test, it is a

paper and pencil test. Standardised test's strengths, weaknesses and levels of usability are similar to those of the achievement test.

Formative Test

This is an example of an achievement test that is administered on the students during the course of instruction. The purpose of the test is to use its result to improve learning. When a teacher administers such a test, he is able to identify the weaknesses of the students and so provide remedial measures.

Summative Test

Another example of an achievement test that is administered at the end of the course of programme is known as summative test. The purpose is to take final decision as regards the course or programme, to decide whether to terminate it or continue it. Decision could also be on the candidates' performance, i.e. as to whether they pass or fail.

Psychological Tests

Psychological tests are an essentially objective measure of sample behaviour. These are like tests in any other fields, e.g. science, since observations are made on a small but carefully chosen sample of an individual's behaviour. Examples of these tests are personality, aptitude, intelligence, diagnostic tests, etc. We discuss some of these below.

a. Aptitude Tests

Most organisations recruit their staff members through the use of aptitude tests because these tests are primarily designed to predict success in some future activities. For example, an applicant into a bank who is considered fit using aptitude test to be employed as a manager is believed to have possessed some qualities which would help him succeed as a manager. Aptitude tests are employed to estimate the extent to which the individual will profit from a specified course of training or to forecast the quality of the

achievement in a new situation. An example of such tests is the *Differential Aptitude Test.*

b. Intelligence Tests

The term intelligence assumes different meanings under different situations and circumstances.

Intelligence tests in this context refer to tests of general ability. These tests are designed to assess the level of general cognitive functioning; such tests measure verbal ability of an individual.

Classification of tests

Tests are classified using many criteria. The most common of these criteria are:

i. Area of measurement, that is aspect of behaviour that is to be measured, e.g. skills, aptitude, intelligence, value, interest, etc.
ii. Ordered (time when the test is administered). For example, we have pre-test (i.e. before instruction), and post-test (i.e. after the course of instruction).
iii. Speed or time allocated to students' responses, e.g. power or speed test.
iv. Scoring pattern, e.g. subjective (essay test) or objective tests.
v. Purpose of the test. The result from the test could be used for selection, placement, certification, diagnosis, etc.
vi. The general discipline (content of subject matter) as in English Language achievement test, mathematics achievement test, etc.
vii. Norm or criterion referenced test.
viii. Individual or group test.

Uses of Tests

There are many reasons why tests are carried out. Some of these are outlined below:

1. Identification of individual differences. Tests, for example, achievement tests, should be able to discriminate between weak and strong students, hence it should be easy for us to identify the weak student from a set of students.
2. To determine the extent to which the behavioural objectives stated before the lesson began have been achieved.
3. To award certificate to learners at the end of a training programme.
4. To assess a teaching methodology.
5. To enable the school to report to students and parents or guardians on the students' progress.

Observational Techniques

One may begin by asking what observation does mean. Observation is looking at and recording the behaviours being observed (Chacko, 1999). Observation ensures first hand and thorough objective information gathering about an event or behaviour being observed. Yet another could be how should we observe? It is known that observation deals with such non-cognitive behaviours that cannot be evaluated using paper and pencil. Hence, we observe non-cognitive behaviour such as interaction between teacher and students, a student and other students, mannerism, habits and so on. We may want also to ask the question why do we observe? We observe for different reasons. One of such reasons could be to make decision about a process, e.g. teaching – learning process. A teacher could observe his students when he had given them a piece of tasks to perform, he may want to find those who have difficulties with the task so as to find ways of helping them. Students too could observe their teacher solving a particular mathematical problem so as to gain the skill required in solving such a mathematical problem.

Observation could be done by anybody depending on what or who is to be observed. Categorically, there are two types of observers: those are the trained (experts) and untrained (laymen) observers. The difference between the two is that the trained observer makes systematic observations and records the findings on paper while the untrained observer may observe haphazardly and record his

findings in his mind, which makes it difficult to reproduce and use the findings after a short or long time due to forgetfulness, among other factors.

The use of specially designed evaluation instruments to collect observational information is referred to as observational technique (Okpala, et al, 1993). And this is the main focus of this section. The following observational techniques will be discussed here:

i. Rating scale
ii. Checklist
iii. Anecdotal records
iv. Systematic observation

Rating Scale

A rating scale is an unstructured description of behaviour, hence, features subjectivity. It is made up of description of a set of traits, attitudes, qualities or characteristics to be rated and some kind of scale for indicating the degree to which each trait or quality is present in the person being rated. The rater then places the ratee at some point along a continuum or in one of the ordered series of categories. A numerical value may be attached to the point or the category in the continuum. Types of rating scales include the graphical and numerical rating scales.

The graphical rating scale has a description of a set of traits or attributes to be rated and a horizontal line representing the continuum where the rater can be placed.

Consider the rating scale on teachers' evaluation of what makes school effective shown below.

The under listed are some of the items considered to make schooling effective. Read through each of the items and rate on its importance. 1 – less important; 2 – hardly important; 3 – important; 4 – more important; 5 – most important. Your responses shall be treated in confidence.

1	Qualified and adequate teachers should be employed	1	2	3	4	5
2	Conducive and clean environment for learning should be provided	1	2	3	4	5
3	Well furnished laboratory for practical should be provided					
4	Well equipped library with recent books					
5	School rules and regulations should be kept					
6	Sports and games materials should be provided					
7	Clinic for students should be provided					
8	Students must be disciplined					
9	Number of students in a class should be minimized					
10	Social activities should be encouraged in schools					
11	Sport activities should be encouraged in schools					
12	Competition among students like inter-class debate should be encouraged					
13	There should be punctuality among students					
14	Students should do assignments					
15	There should be counselling sec-					

	tion for students					
16	Freedom of expression should be encouraged among students					
17	The classrooms must have enough space for passage					
18	The school compound should be beautified with flowers					
19	There should be adequate security for students					
20	Self study habit should be encouraged among students					
21	Admission of students should be on merit					
22	Students should participate in major competitions at both local and national levels					
23	High moral standard should be promoted among students					
24	Roll of honour and prizes should be given to good and brilliant students					
25	Parents should feed their children well					
26	Parents should support the children when in school					
27	Guidance should be provided from parents to students					
28	Social amenities like water, electricity are to be provided in schools					

29	The classrooms should be well ventilated					
30	There should be a suggestion box					
31	Effective communication among students should be encouraged					
32	There should be a good relationship between students and teachers					
33	Government policy on education should be implemented					
34	The syllabus must be covered					
35	The recommendation textbooks must be provided and used					
36	The late comers should be disciplined					
37	The uniform must be attractive					
38	Corporal punishment should be cancelled					
39	Students must be dedicated to their studies					
40	Teachers should not victimize students					
41	Examination should be seen as a way of bringing the best out of students					
42	Mastery learning should be encouraged					
43	School prefects must see to the					

	welfare of students					
44	Cultism and waywardness should be seriously checked					
45	Open day should be organised					
46	Teaching aids should be used to make learning effective					
47	Teaching methodology by teachers must be efficient					
48	Valedictory service should be conducted for outgoing students					
49	Examinations in school should be well conducted					
50	Computer education should be made compulsory					
51	School should go on internet					
52	Students should make use of the library to enhance knowledge					
53	There should be provision of school bus for easy transportation					
54	Students' goals in life should be sharpened through career talks and seminars					
55	Students should be ready to learn					
56	Interview should be conducted for prospective students					
57	Students' progress should be motivated					

Using graphical rating scale, the rater checks a portion along the continuum to represent the ratee's evaluation as in the example above. A rater's score is obtained by measuring the distance between the end of the line and the checked positions.

Like any numerical rating scale, it has descriptions of some traits on one side and numerical values are assigned to each trait. Note that instruction is always provided so as to help the rater decide the numerical value to be assigned to each trait.

Among the rating scale's strengths is its flexibility of usage; the fact that observations are directed to some specific attributes to be rated and finally, the fact that little time is spent in coding observable behaviours. Hence, a rater can do many ratings within a limited time. However, the use of rating scale is limited due to some errors in the use of the rating scale. Among these is that it is possible for a construct to have more than one meaning; and the possibility of rater's halo effect (rating due to general impression of the rater, it could be positive or negative impression).

Checklist

Usually, a list is provided and the rater checks whether the traits or attributes to be measured are there. Some people believe that a checklist is a 'Yes' or 'No' response. It is a method of data collection to determine whether an action has taken place or not. The checklist is useful for measuring performance skills. For example, in evaluating the technical know-how of a student teacher during teaching practice, a list which can be divided into series of clearly defined specific actions (of what the supervisor considers appropriate) and each student teacher is checked based on the list.

Anecdotal Records

This is a factual written description of the meaningful events, which the teacher has observed over time about a particular student's behaviours. These behaviours are recorded immediately they occur. The records should be concise and clear. Each record should be for only one person and about only one behaviour. It gives a longi-

tudinal report of a student if it is done over a long period of time. It provides an informal (not official) and qualitative description of certain aspects of the student's behaviour or traits, which can be used to monitor his growth and development. These types of traits could not be measured using other measuring instruments like test. Examples of what could be in an anecdotal record are behaviours like lateness to class and reasons for it, willingness to take or not to take up responsibilities, etc. One limitation of anecdotal records is that it is time consuming. Again, the interpretation could be cumbersome.

Systematic Observation

A more reliable and objective way of observing is through the use of a systematic observation in which a specific behaviour is observed using an instrument prepared for that behaviour. There are two types of instruments in a systematic observation; these are the signs and the category systems. These work like two types of cameras – still and moving.

This sign system works like a still camera (it produces a still picture). It takes the picture of specific and first occurring behaviour. This type of sign system is characterised by frequency of different behaviours. There could be many behaviours to be observed in the sign system.

The category system on the other hand works like a moving camera – as in a video recording. The frequency and sequence of behaviours are the major characteristics of this system. Each of the systems is equally useful in observation. A good example of sign system is illustrated below, using an effective net learning time observation schedule.

The category system has fewer numbers of behaviours than the sign system. In most cases, category system serves more purposes than the sign systems. Nowadays, there is a thin distinction between the two systems as no observational instrument has sign system or category system attached to them.

One of the earliest observational instruments is the Flander's Interaction Analysis Category System (FIACS). This is an instrument, which is still useful even now, either in its original or modified form. There are other category system observational instruments such as the Yoloye African Primary Science Programme Observation Instrument, the Verbal Interaction Category System (VICs) by Amidan and Hanter; International Association for the Evaluation of Educational Achievement Classroom Observation System (IEACOS) and the Observational System for Instructional Analysis (OSIA).

Expert Judgement

Let us assume that there is a vacant position for about 10 to 15 applicants. After a series of aptitude tests have been administered, three of these applicants got the best and the same score. Experts may be brought in to further assess these applicants. Expert judgement, therefore, is a technique, which involves experts or specialists in certain aspects of an educational or organisational programme. When these experts are called in to assess the applicants, they can now select a candidate based on their judgement. Normally, what the experts do is to further test these candidates. So, we can say that they use the interview technique for their judgement.

Interview

Interview is a face-to-face communication between the interviewer(s) and the person being interviewed. An interview schedule, which is a written document serving as a guide for asking the person being interviewed a series of questions, is used to gather information from him about the subject of interest. Normally, information about the person being interviewed is first sought. interview schedules are not usually used to elicit information on such simple questions as 'Yes', 'Agree', 'True', 'No', 'Disagree', or 'False'. Elaborate information should be provided when the interview schedule is properly used. The interview schedule may or may not be structured. If it is structured, the persons being interviewed are expected to give the same responses to the questions given. Un-

structured interview schedule does not elicit uniform responses from the people being interviewed. An interview can be conducted through the telephone, through e-mail or through personal (physical) contact.

The following are some of the strengths of the interview technique:

a. Interview can be used on both literate and illiterate persons.
b. People being interviewed can ask for explanations if the questions given to them are not clear or are ambiguous.
c. Responses from interview are quickly obtained.
d. The interviewer has close contact with the person(s) being interviewed.
e. True information is likely to be obtained through an interview.
f. Some people do not want to write but prefer to talk, so an interview is best for such people.

Some limitations of interview include:

a. It is time consuming, since each person is interviewed separately.
b. The analysis of such information obtained from an interview schedule is difficult to analyse.
c. The result obtained from an interview may be subjective since each person being interviewed differs in a number of ways.

Questionnaire

This is the most common of evaluation instruments. It is useful for both teachers, students, school administration, and research officers in gathering information about subject of concern/interest. A questionnaire, unlike an interview, is a printed list of some questions, which respondents could respond to on their own with little or no assistance of the research officer. It can be used on a fairly large sample. A questionnaire can either be open ended or close ended or a combination of the two. An open ended questionnaire

allows the respondents to express themselves on certain issues without restrictions. In a close-ended questionnaire, the respondents are restricted to pick from the options provided. An example of an open-ended question in a questionnaire is:

How old are you (that is, age last birthday)? Supply the information in the space provided.

_____

Senatorial zone ----------

School location ----------

Local Government Area----------

An example of a close-ended question in a questionnaire is:

What is your qualification? (pick one)

a. No formal education ----------
b. Primary school ----------
c. Secondary school ----------
d. OND/NCE ----------
e. HND/First Degree ----------
f. Masters degree ----------
g. Professional ----------

Other examples are:

School type 1: Public [ ]; Private [ ]

School type 2: Male only [ ]; Female only [ ]; Mixed [ ]

Gender: Male [ ]; Female [ ]

Age range: 12 – 14 [ ]; 15 – 17 [ ]; 18 and above [ ]

	Father's occupation	Mother's occupation
Civil servant		
Self-employed		
Business		
Pastor/church worker		

	Father's Qualification	Mother's Qualification
Secondary Sch		
NCE		
B.Sc/B.Ed		
M.Sc/M.A./M.Ed		
Others (specify)		

Advantages of Questionnaire

a. Questionnaires can be administered on a fairly large sample of respondents simultaneously.
b. Skill required to administer questionnaire is less as it could be posted to respondents with little or no explanation.
c. Since the identity of the respondents is not always required, the respondents feel free to express their mind without reservation.
d. Respondents are not particularly in a hurry to respond to the questionnaire, wordings of some items that look confusing can be re-examined so that the respondents will have an approximate understanding of what the questionnaire designer had in mind.

Disadvantages of the Questionnaire

a. Although, there is time for respondents to respond to a questionnaire, there could be some misleading words, which the respondents think they understood but which they may not have understood as such.
b. Questionnaire cannot be administered on illiterate persons, most especially when there is nobody to interpret the items.
c. Rate of return of questionnaires by the respondents may pose a serious problem to the researcher.

Hints about Questionnaire Design

Whenever a questionnaire is being designed, the following must be avoided.

1. *Any vague question.* This is a question that generates confusion. For example, 'What do you think about religious leaders?'
2. *Any leading question.* This type of question biases the mind of respondents. The question suggests the answer to be given. For example, a question such as 'The masses are supposed to be kind; what do you think about of the police?'
3. *Too intimate a question.* That is, a question that tends to focus on one's privacy or family affair. The following would be a poor item: 'Do you have HIV/AIDS?'
4. *Too wide a question, for which no specific response is possible:* e.g. 'What is going to happen to Nigeria?'
5. *Too narrow a question, designed for just a specific person.* E.g. 'Akpan, are you from Akwa-Ibom State?'
6. *Too technical a question.* E.g. 'What do you think about lingo conflict?'
7. *Too abstract a question.* E.g. 'Explain the activities of witches and wizards.'

CHAPTER SIX

CONTINUOUS ASSESSMENT AND EVALUATION

Introduction

The process of education is not complete without the component of assessment. Distasteful as this component might be to learners [who do not like to be tested anyway], it is indeed as necessary as the component of teaching itself. This assertion brings to mind the argument between two young learners over the need for tests and examinations. The first had asserted that it is not really necessary to subject learners to the 'ordeal' of examination since, according to him, apprentices to vocations such as carpentry, bricklaying, sculpture, etc., are not normally subjected to the 'degradation' of being examined before they are declared competent to practise their profession. The other student, who had started off from the plank that examinations are unavoidable, ended up agreeing with his friend, since in any case he too was not much in love with examinations. Both regarded the 'inventor' of examinations as a wicked and malicious soul.

Poor young chaps! If they had carried the argument of the 'detester' of examinations a little further, they would have inevitably reached the conclusion that assessments were necessary, since indeed the apprentices in the argument were all the while going through a process of testing. Think of the young apprentice right from the day he takes up the vocation. His master expects him to watch every movement he makes on the job, and then expects him to imitate him. The longer the apprentice stays on the job, the more tasks he is expected to be able to handle, and if he is a competent learner, he soon is able to undertake tasks on his own. At each stage, his master keeps him under watch and each time he strays off from the expected path, he is promptly corrected. This

way, continuous assessment goes on. A constant examination of some sort is progressively on. This is the very reason why some learners are 'freed' at exactly the expected duration of the course whereas some others are not so favoured; they are told they have not sufficiently mastered the vocation to warrant being 'freed' and a few more years are added to their period of apprenticeship.

The point being made above is simple: assessment is necessary, even in the most mundane task a human being undertakes, and even more so is continuous assessment. In academics, continuous assessment is even more essential. In the next sections, we spell out what continuous assessment really is, what makes it slightly different from other forms of evaluation, and why it is necessary.

What is continuous assessment?

The term continuous assessment might connote the idea that an action is going on all the time, that there is no breathing space to do any other thing, and that all that matters is the assessment. After all, the Pocket Oxford Dictionary defines *continuous* thus:

> *...uninterrupted, connected throughout, in space or time...*

This is the sense in which some educators have suggested that a better term might have been *continual assessment*. The dictionary definition of *continual* reads thus:

> *...constantly or frequently recurring, always happening...*

Happily the same dictionary explains:

> *Continual is often confused with continuous. Continual is used of something that happens very frequently (e.g. there were continual interruptions), while continuous is used of something that happens without a pause (e.g. continuous rain all day).*

These distinctions notwithstanding, the fact is that in educational circles, the specific term *continuous assessment* is used. We go back for the final time to spell out what this term means:

> *...(regular) evaluation of a pupil's progress throughout a course of study.*

With that, we should accept that the terminology is not a blatant anomaly. This is even more acceptable if it is remembered that the teacher does not have to use tests all the time to assess his learners, that a good teacher can 'feel' the mood of the class and thereby assess whether or not they are learning what he is teaching, we might accept that assessment is going on all the time as the teaching is progressing.

Besides, the good teacher uses occasional short, impromptu quizzes, which can sufficiently elicit sufficiently reliable information about the level of the learners' mastery of the subject matter. So, in a way, describing assessment as *continuous* might not be totally wrong.

However, *continuous assessment* entails much more than what we have described above. A formal description of the term would establish it as involving all the performances of a learner during the duration of the whole of his course, and in all the areas of learning [i.e. cognitive, affective and psychomotor]. This is indeed the import of the Federal Ministry of Education's [1979] definition of the term, thus:

> *Continuous assessment is a method of ascertaining what a pupil gains from schooling in terms of knowledge, skills, industry, and character development, taking account of all his/her performances in tests, assignments, projects and other school activities during a given school period [term, year or entire period of educational level].*

The same source goes further to stress that continuous assessment is a method of using the recorded performances of each learner to help him/her improve on his/her achievements through individual

guidance, while it also enables the teacher improve his/her teaching by identifying and remedying areas of difficulties in the learners' performances.

Main Features and Characteristics

The main features and major strengths of an assessment procedure such as has been spelt out above can be more elaborately outlined thus:

a. It is systematic in the sense that a clearly defined procedure is followed in settling and administering the tests, much especially because the various units [or different doses] of the tests are so carefully designed to ensure that each one takes cognisance of those coming before and those coming after it. Besides, under this system, the test would not 'happen by chance', since all assessment units must have been pre-planned as integral parts of the course of work for the period [term, semester, or session].

b. It is objective in that the weaknesses of the 'one shot' system of examination are reduced to the barest minimum since the teacher himself learns from the weaknesses of earlier exercises and improves in setting and marking subsequent units, thus ensuring that the assessment procedures are, on the whole, fair to all. [Objectivity, as a term, is used to imply 'fairness'. In the present context, it is not used in the usual sense in which one and only one of the provided alternative answers is correct. Rather, objectivity has been used here to imply utmost reduction of discrepancies in setting and marking, and optimum areas of coverage, all of which ensure fairness to all candidates].

c. It is cumulative in that the results are compiled over a long period of time, all such results being combined to define the learners' progress.

d. It is comprehensive in the sense that all the various units of the tests, taken together, would cover all the different facets of the teaching done during the period in question [term, semester or session], in a manner which no 'single

shot' examination exercise can do. This is especially so when different examination procedures are used, e.g. the essay type, the multiple-choice type, the oral interview, etc.

e. It is guidance-oriented in that the tests lead to results, which the teacher can use to guide the learner's activities, help the weaker ones, and structure the energies of the stronger ones.

f. It makes the teacher more effective in contributing to the decisions leading to the award of the final certificate than would have been the case under a different assessment process. For instance, under the terminal and periodic assessment systems, apart from preparing his/her students for the certificate examination, the classroom teacher has no say whatever in the decisions concerning the award of the certificate. Under the continuous assessment practice, however, since a substantial part of the final grade would come from the cumulative school record, the class teacher has an influence in decisions concerning the award of the certificate, a privilege he hitherto did not enjoy.

g. It makes both the teacher and the taught more diligent in the sense that while teachers must constantly assess, award grades, and try to improve on their own performances, the learners similarly would invariably find themselves constantly on their toes if they are to keep on performing satisfactorily in the frequently administered tests.

Demands of Continuous Assessment on the Teacher

From the foregoing sections, it should be clear by now that the practice of continuous assessment invariably places a much heavier demand on the class teacher. Although these demands might vary from subject to subject, there are very many common areas of demands; and these we discuss in the following paragraphs.

In the first place, a teacher operating this system has to go on updating learners' scores and in this way maintain an up-to-date picture of the rate of improvement of the learners.

Secondly, a consequence of the last point is that the teacher would find himself keeping a record of learners' performances. This record has to be updated from day to day. So, he must cultivate a habit of keeping and maintaining records. This, we daresay, is a great challenge to those teachers who find it extremely difficult to keep proper records.

Thirdly, the teacher has to continuously update his teaching techniques, procedures, materials and objectives in line with the performances of his learners. Teaching reflects testing, and vice versa. If one has to assess learners on a regular basis, it stands to reason that one has to be up and doing in one's teaching. This is one of the most important demands of the continuous assessment practice, because the main essence of the practice is to generate the next unit of teaching from the learners' performance in the last one. Each new unit of teaching must be closely guided by the information gathered from the last assessment exercise. We will have to expatiate further on this in the paragraphs following the next.

Fourthly, the teacher must carefully design the teaching/learning/testing procedure in such a way that the pressure on learners is eased as considerably as possible. If the procedure is not carefully designed and meticulously handled, it is easy to allow the emphasis to shift from the teaching to the testing. Such a situation might make the learners get attuned to the testing procedure without really learning much. It is the teacher's task to make the learners see the assessment procedure as an integral part of the teaching and that the testing is indeed one of the avenues for helping and guiding them.

Implications on Linear and Cyclical Curricula

One way of delineating the types of curriculum that there are, is by describing them as either linear or cyclical. According to several leading authorities, under a linear curriculum are subjects in which are generally independent of one another, whereas with subjects belonging to the cyclical genre, concepts to be taught are very much dependent on one another and so one cannot teach one

thing entirely independent of a previously taught concept. Let us consider the nature of languages and the sciences to illustrate our points more clearly.

In teaching a science subject, e.g. Biology, it is generally possible to teach a concept almost entirely independently of other concepts. For instance, one could teach the human skeleton without recourse to the knowledge of blood circulation, the digestive system, the reproductive system, etc. Similarly, one could teach blood circulation without demanding that students should use their knowledge of skeletal system, the nervous system, the reproductive system, etc. However, in order to teach a concept in language, it is very often necessary to build on the previously taught concept. For instance, to teach verbs denoting futurity, it is essential to first establish that students master the present tense, the present continuous tense, and even the past tense, as well as the use of modals that are necessary in expressing futurity. No teacher of language worth his salt would teach futurity before teaching these other concepts. Similarly, to teach plurality one must first teach singularity; to teach the way plural nouns affect the verbs they take demands that one should first establish that students understand the way singular nouns affect the verbs that follow them. Virtually every concept in language learning is intricately interwoven in some other concepts. So, language teaching falls within the cyclical [or spiral] curriculum whereas the sciences fall within the linear curriculum.

The implication of all this on the teacher in designing his continuous assessment tests is that, by and large, the teacher has to understand the nature of the concepts he is teaching in order to assess them judiciously. If he is handling a subject that is linear in nature, he might be a little more flexible in his choice of the order of which topics to teach first. A teacher in this type of subject area needs not follow the order of arrangement adopted by the text book in use; rather, he should decide which topics he could most conveniently handle first, [e.g. those for which he has the essential tools] and which ones should come later. When it comes to the regular assessment exercises, he might decide to be guided in his

choice of which topics to teach in the various weeks by the result from the previous tests. The net implication of what we are stressing here is that the teacher in a linear subject area can exercise more flexibility than the teacher in a cyclical subject area.

A teacher in a subject area that is cyclical in nature hasn't got that much free hand. Essentially, he has to accurately measure the extent to which the present unit has been mastered before finally deciding on his procedure for the next concept. In planning his teaching of any new concept, he has to meticulously assess how well the learners have mastered the last concept taught. This of course makes it the more necessary for him to measure the achievement of any concept taught before going on to teach the next concept. If the results of an assessment exercise indicate that complete mastery has not taken place, it would be essential that he should re-teach the concept; otherwise the next concept would equally not be mastered.

This again demands that he should vary his methods of teaching. Varying the method ensures some freshness, some ingenuity in the teaching process, and therefore a higher chance for the learners to understand. Similarly, variety is also demanded in handling the tests. Using the same one test format and test procedure week after week will soon lead to boredom with its inevitable consequences. But when the teacher designs tests for the different weeks, with different formats and adopts different procedures, the students are more likely to be stimulated for the tasks before them. Of course, one clear implication of this is that the teacher must master the different types of tests. It could be a short quiz this week, while for the next week there could be a multiple-choice test, and for the following week, it could be the matching test or the true-false test, etc.

Of course, there is yet another reason why the test types must be varied over time. To probe meaningfully into the learners' level of mastery of the concepts taught, one method of assessment used over and again cannot produce the desired accurate measurement. Some learners excel when essay types of questions are used, some

do better when the stimulus is an objective form of assessment, while others do better in oral examinations. The teacher should use as many of these different devices as he can muster from his arsenal of evaluation machinery. Short quizzes, short comprehension passages, elaborate composition exercises, objective tests, etc., all should be used at various times.

Indeed, even within each testing mode, variety and flexibility should prevail. Pictures, photographs, diagrams, drawings, newspaper and magazine articles, the plan of the school compound, etc., could be used as stimuli for testing learners' competence in the various topics taught over time. It is saddening that a biology teacher could be testing a topic like the alimentary canal or the life cycle of a cockroach in Junior Secondary School without using a diagram of any sort. Similarly, it would have been more cheering to find a teacher of English introducing newspaper cuttings of good articles in the tests.

Essential Steps for Continuous Assessment Practice

In order to execute the continuous assessment practice, the teacher should ensure that the test items are able to discriminate sufficiently between the various ranges of abilities in the class. This is to say that the test should be able to produce results showing the very best students, the next best and then down the line until we have the poor and the very worst ones. This is putting it in the layman's language. The best way to have this is to ensure that the test is valid, that it is reliable, that it has sufficiently wide coverage, and it is practical to administer and easy to score under the environment for which it is designed. We have already discussed these at length earlier on in chapters 3 and 4. So, we need not overstress these here. Specifically, for the continuous assessment practice, the teacher should take cognisance of the following.

First, the teacher should realise that a continuous assessment test should focus on a small unit of knowledge or a specific concept taught during the week, fortnight, month, or part thereof. The ideal is that the continuous assessment test should form the natural

culmination of the lessons or lesson taught in recent time. Thus, he should plan his series of lessons right from the beginning so that after the lesson or few lessons which deal with a specific topic or concept, the test comes up as a logical end.

Second, the shorter the assessment test is, the better. An assessment test could cover as short a time as five minutes, although it could equally be as long as a whole lesson. The essential point is that it should not be too tedious or too demanding. In fact, the less tedious it is the less the students would be scared of such exercises.

Third, the teacher should create a relaxed atmosphere for a continuous assessment exercise. Let the classroom atmosphere be lively. This way, the myth of an examination or a test as an ogre or a tyrant would be removed, and students would be able to learn from it. Remember, this is the very essence of continuous assessment – that the students should learn from and through it.

Moreover, the test should be as adequately planned as the lesson itself is. That we have recommended that it should hold in a relaxed, fear-free atmosphere should not mean that it should be haphazardly handled. Right from the beginning, when the lesson is being planned, the teacher should decide what type of test he would use, how many items there would be, how much time to devote to it, how it should be scored, and how the result would be disseminated and used for further teaching. Indeed, even where oral test is to be employed, the items should be carefully designed in advance and written down. This is true whether the test is for five minutes or for a lesson of forty minutes.

Not less important in this regard is that the items should be the type that could be scored within a very short time. In many cases, it should be possible to score the students' attempts right there in class, sometimes even by the students themselves after exchanging their papers. Scoring the students' attempts in class swerves two important purposes. First, it ensures that students have an immediate feedback or knowledge of result, a concept which psychologically help to reinforce learning. [Remember, psychological experi-

ments have shown that when knowledge of result or feedback is prompt, learning is reinforced; whereas when there is so much delay before results are known, learning tends to be impeded].

Secondly, scoring the attempts right there in class helps to serve as a revision. As the teacher stands before the class, calling out the items and leading the class to suggest the correct answers, revision of the concepts earlier learnt is going on. Where they go wrong, students note the correct answers and learn from them. Where they are correct, they will rejoice and thus reinforce their learning.

We may need to examine here the earlier statement that very many modes and types of tests should be used. The ideal is a situation in which test types are varied from time to time. If a teacher resorts always to the essay type of testing, the ordeal of scoring students' attempts would soon discourage him from assigning further tests. On the other hand, if he should use quiz in one week, the oral test in another, the multiple-choice objective the next week and the matching type the following week, no one would be bored and scoring would be relatively easy.

Before leaving this issue, we need to clarify the recommendation on the use of the quiz and the oral tests. A number of people do not think highly of quizzes, apparently because many people think of them as puzzles or riddles. This need not be so, however. Rather, the teacher may conduct an oral test by standing before the class and dictating the questions which students answer in their papers. Ideally, such a test should involve short questions demanding one word or one-phrase answers.

As for oral tests, we are not insisting here on the use of the interview, one candidate at a time type of exercise. Rather, if the items are so constructed as to demand just very short, one-word answers, the questions could be posed orally to students. They are asked the questions in turns, one student after another, until all the members of the class have had a taste of it. As the exercise is going on, the responses are being scored. The weakness here is that this

is time consuming, and besides some candidates may be lucky to have easier items than others.

Finally, the scores should be kept in a permanent record book. Recall that continuous assessment is supposed to be cumulative in nature, i.e. with results from several such exercises built up over a long period of time. This way, students would have scores from as many as ten tests within the term or semester, with the obvious advantage that those who found things difficult on a few occasions would have had time to recover, and one's weaknesses would have been evened out by better performances at some other occasions. The results from these several exercises would be summed up as a part of the total overall score in the end of course result.

From the foregoing, and most especially from the last point, the question arises as to what the relative weight of the continuous assessment scores should be. This is discussed in the next section.

The Relative Weighting of Continuous Assessment Scores

This section becomes necessary here in the light of the various attitudes of different people and different institutions to continuous assessment exercises and scores. These are those who claim with sufficient reasons that continuous assessment scores should not amount to more than 10 per cent of the total end of course scores; while there are those who argue that the scores should account for as much as 50% of the total. Indeed, there are some teachers who rely exclusively on their continuous assessment scores, i.e. without recourse to formal examinations. Our stand is that all these extreme practices are perhaps abnormal.

We need to ask here why some experts are wary of giving so much weight to continuous assessment scores while some prefer very generous weights. Those who prefer very low weight for continuous assessment scores often argue that a lot of malpractices often go into the exercise. This, sad indeed, is very often true, especially in public examinations involving several schools across the country for which many teachers forward abnormal scores which bear little or no correlation with the scores from the real public examinations.

On the other hand, those who advocate very high weightings to continuous assessment scores insist that it is better to judge students on the basis of what they do regularly, on a continuous basis, than on the basis of just the end of course examination at which a student may be so unfortunate to perform woefully because of ill health, emotional disturbance, domestic problems, etc.

Both sides seem to have a valid point. This is perhaps why there is a formal recommendation on this, a recommendation which appears to steer a mid-way course between the extremes. The public examining bodies in this country assign just 30% of the total score to continuous assessment. Similarly, that is the recommended weighting in most of the higher institutions. It is our stand that this is the ideal.

We now attend to the demands of record keeping for continuous assessment practice.

The Demands of Record Keeping

We have stressed earlier on that continuous assessment is cumulative in nature. This implies clearly that the scores obtained from the exercises should be kept somewhere. How should they be kept, and how should they be treated and used at the end? These are the issues we consider in this final section of this chapter.

To begin with, the teacher should have a register for all his scores. In many schools, each class has a ledger in which all scores are entered on a continuous basis. However, where this is not the practice, each teacher should keep the record of continuous assessment scores in a notebook, especially that in which all the plans for his lessons are contained.

Furthermore, it is ideal that each of the tests should have a common maximum score. For instance, each could have a maximum score of 10. This is not a rigid rule, however. The one advantage it has is that the students regard each one exercise as having equal weight and they thus give each one equal attention.

Finally, at the very end of the exercises, the various scores should be summed up and computed on the basis of the relative weighting assigned to continuous assessment by the school.

References

Ayodele, S. O. [1985] *Continuous Assessment Practice and the English Language Teacher.* Ibadan: Evans Brothers Publishers.

Obanya, Pai. [1982] *Evaluation in Modern Language Teaching.* Institute of Education, University of Ibadan Press.

Oxford Pocket Dictionary [1999] *Oxford Pocket Dictionary. [CD edition on the computer]*

CHAPTER SEVEN

CURRICULUM EVALUATION

What is Curriculum?

Most definitions of 'curriculum' refer just to the academic activities conducted within the confines of the school. For instance, the *Pocket Oxford Dictionary* says tersely that curriculum means 'subjects included in a course of study'. Even the bigger *New Oxford Dictionary of English* is hardly more helpful when it defines curriculum as the 'subjects comprising a course of study in a school or college'. Indeed, the same dictionary's definition of national curriculum is hardly more embracing as it says this means 'a common programme of study that is designed to ensure nationwide uniformity of content and standards in education.'

Much closer to the ideal is that by Martin Mclean [2002] who describes curriculum as 'systematically organised course of teaching and learning'. He adds that 'Some definitions of curriculum focus narrowly on the arrangement of subjects over a sequence of grades, others include everything that students and teachers do.'

With this, we can say that we are getting nearer the ideal position. From what we have had thus far, we can identify both the narrow and the wide perspectives of curriculum. By the narrow perspective we are referring to curriculum as just those academic activities conducted within the confines of the school, while the broad perspective of curriculum should refer to all those activities, within and outside the school, as well as all other activities and agencies that are connected, even if remotely, with the business of education.

We should be aware that education is one of those spheres of life in which it is not just the professional that has a hand. The professional in this context is of course the teacher. When we think of education of learners, we would agree that virtually all people in a

country are involved. Think of the parents, think of religious institutions, think of several educational or examination agencies, think of publishers, etc., and you would see that all these are highly interested in the business of education. Indeed, we should accept that the professionals occupy just one of the vertices of education. Think of the formulation of syllabuses in the various subjects; this is conducted not just by the classroom teacher but also by curriculum experts invited by the ministries of education. Moreover, it is the examination bodies that formulate their own examination syllabuses.

From all these, you would see clearly that curriculum is a very ambitious concept that spans far beyond the classroom. And this will continue to be so for as long as the learners are regarded as held in trust for parents, for the society and for the nation.

What then is curriculum evaluation?

From the foregoing, we should begin to see that curriculum evaluation would go far beyond the ordinary classroom assessment. Here we should be thinking not just of the ordinary assessment of individual subjects like English, mathematics, geography, chemistry, etc. Rather, we should conceive of curriculum evaluation as embracing the assessment of the whole educational system. We should not be interested in just how learners are performing in subjects taught during the course of a specific period of contact with the school, nor are we interested in how brilliant a pupil has been in a particular subject, or indeed, how well a particular teacher is performing his duties. All those are fundamental to education, all right, but they do not tell us the whole story of why things are as they are.

To obtain the full story, we have to conduct a thorough evaluation of the system itself. We have to do a good evaluation of the business of education. The use of the term 'business of education' here is deliberate. You are well aware that in economics, a particular business venture can be evaluated. Indeed, before setting up a business, it is just as well that a good, thorough feasibility study

should be conducted so as to determine the likely chances of success of that business in a particular place, at that particular time, and given the particular costs of investments, etc.

You are aware, aren't you, that there is virtually nothing in life that cannot be evaluated. Depending on your approach, you can evaluate the social health of a family, the attitude to religion of a people, the contributions of religion to the welfare of the people, etc. You can evaluate just anything. And so, it is for the curriculum in its broad sense.

So, then, why are we interested in evaluating the curriculum in its much broader sense? There are various reasons for this, and these we try to explain in the next section.

Reasons for Curriculum Evaluation

If you have studied the earlier chapters of this book you would understand what evaluation is all about. Thus, you would begin to see why evaluation is very necessary, even at the narrow level. Curriculum evaluation is essential first in order to enable the stakeholders in the business of education to determine if the system is performing as it is expected to perform. No sane person ever sets up a business deliberately to make a loss. No! The rational human being goes into business to make a living, and if he has to do this, he has to ensure that his business is profitable. Similarly, in the business of education, the intention is to run a system that is highly efficient, that makes sure that the children are getting the type of education which will make them useful citizens, while the parents, the state, etc., can assert that they have value for their money without experiencing unnecessary wastage.

Speaking like this often leads to the question of whether or not education is supposed to be an economic venture in which profits are expected at the end of a business quarter. Well, we are not talking here in that sense. The 'profit' in educational equations is not in monetary terms. The profits are in the production of learners who are able to perform the way they are expected to perform after a specific course of study. The 'profit' that society expects

from the school system lies in learners not turning out as highway robbers, or as miscreants, or as liabilities on society. The 'profit' lies in having learners who leave the school system to make their world better than they met it.

That being the 'profit', it has to be weighed against other factors such as time, monetary costs, etc. Given the explanation above, to establish if the educational venture has been really 'profitable', it has to be shown that the time taken to educate a learner to a given level is equitable. For instance, if we assume that it takes six years to educate a primary school learner, it has to be accepted that a system is not profitable in which a learner or some learners spend more than six years to cover the distance. On the other hand, where a learner does the journey in six years and cannot perform like a true primary school graduate, the system that produces him is not profitable. Thus, where the learner can neither read, nor write, nor perform simple computations, etc., the system that produces him has not been 'profitable'.

Besides, the issue of cost is still very germane, although we have stressed that monetary gains do not come in with educational evaluation. No matter which way we look at it, an educational system that costs twice to run as another one is definitely less efficient than the norm, if their products are merely at par with those of another system. But then, in many cases nowadays, the issue of cost is not primary to many people so long as they have results.

Issues such as these should lead us, in the final analysis, to the objectives set for the system to begin with. So, we should examine first and foremost the steps or procedure in curriculum evaluation.

Steps in Curriculum Evaluation

Before we can understand what curriculum evaluation really entails, we should consider first the steps that are crucial.

Arieh Lewy [1977] has spelt out six steps which he styles 'six facets' of curriculum evaluation. We have over the years accepted these in our definition of steps in the process of curriculum evaluation al-

though, as we shall show at the end, these are not watertight and could be adapted.

Determination of general aims

This constitutes the first essential step. This is an inevitable first phase since, as in general educational procedure, the definition of objectives is crucial for a successful exercise. What this phase entails is that the evaluators should define clearly the objectives for which the specific educational institution or educational programme was set up. It is essential to map out clearly why the institution or programme was set up, what it was set up to achieve. This is logical because before you can make a pronouncement as to whether or not a programme is operating well, you have to find out what it was set up to do. If for instance a state government sets up a training institution with the express purpose of producing learners who will design and manufacture mechanical, electrical and electronic gadgets and components, and this institution is to be evaluated ten years later, this express goal of the government should form the basis of the evaluation. On the other hand, if a religious sect establishes a college with the express aim of producing itinerant preachers, this again should form the basis of the evaluation. In each case, the crucial goal of the evaluation exercise is to establish the extent to which each of these establishments is achieving the set goal.

But that is putting it rather simplistically. In most cases, goals and objectives are not so easily defined. A state government does not usually set out in very clear terms what the objectives behind any new institution are. Similarly, a school proprietor usually does not define in very clear terms what the objectives behind his ambition are. In any case, evaluation is usually not called for until several years, indeed several decades, after the establishment. Very often, evaluation may not come up until several decades after the founding fathers have left the stage. So, what does the evaluator do about this? In most cases, the logical step to take is to resort to the generally accepted national documents that contain definitions of objectives for similar programmes. In this sense there are several

documents to resort to. The *National Policy on Education* contains several statements which can be used as valid guides. And apart from that, there are several texts, either by such international agencies as the UNESCO or similar bodies containing highly useful objective statements.

Finally, of course, seasoned evaluators can set valid and useful objectives for themselves. Experience, conventions and common sense should come in handy in helping to determine what the objectives of any institution or programme should be. Every rational being, for instance, would accept that some of the worthwhile goals for a technical training centre set up for secondary school graduates should include: (a) the production of useful citizens, (b) who are thus self reliant, (c) able to use their hands in producing materials for sale to the public, and who are therefore (d) able to settle down to earn their livelihood. Each of these could be developed into worthwhile objectives for the institution and therefore used in the course of the evaluation.

Planning:

This is the second essential step. This may sound too obvious, but the reality is that just too often many a venture is embarked upon without the participants really settling down to do a thorough planning.

This is the stage at which the evaluation team should settle down to define the task before them, the objectives of the programme to be evaluated as well as the objectives of the specific evaluation exercise to be undertaken, how to go about the various stages of the evaluation exercise, how long it is likely to take, what instruments to use, how the results would be analysed, the nature of reporting that should be considered desirable under the circumstances, who would write which report, etc. So, you would see that the planning stage entails quite a lot; and it has to be, since this is the foundation of the edifice, the foundation on which the whole structure lies. Should the foundation be weak, the result can be predicted.

The planning stage also requires that the evaluators should do a lot of reading in order to find out what has been done in similar circumstances in the past. Similarly, a lot of literature is also necessary to be mastered in order to enable them discover what types of instruments could be used, even if in the final analysis entirely fresh, novel instruments would need to be fashioned out.

Finally, under planning, the evaluation budget should be worked out and/or thoroughly reviewed and revised. Realise that for a research institute, obtaining a curriculum evaluation project is exactly like winning a contract. The good contractor has to work out the best way of executing the job with the most equitable cost so as to maximise profit and thus stay in business. This definition of the evaluation budget is quite different from establishing the financial efficiency of the project itself. This latter aspect of the exercise is a necessary part of the evaluation: that it has to be found out whether or not the project is being operated on a basis that is cost effective.

Tryout and Revision

This constitutes the third stage. This is the stage at which materials and instruments hitherto prepared will have to be taken to the field and tried out. The purpose here is to enable the evaluators discover possible areas of weaknesses in the materials and instruments so as to fine-tune and improve on them. It also enables the evaluators discover if indeed there is any possible flaw in the procedure planned for the job. You would be well aware that no matter how experienced you might be, the reality is that when preparing a document or a procedure, there might be some lapses that may not be apparent during the preparation. During the try-out, the hidden lapses should then become obvious enough to be revised.

Usually, the tryout is done, not in exactly the ultimate target institution, but in another one similar to it. Several members of the evaluation team should take part in this, so that at the end they would come together to comment on their experiences with the

instruments. Is the material a questionnaire which calls for open-ended responses? If so, are the responses obtained at this stage easy to analyse? If they are not, then perhaps it would be advisable to resort to the objective type of responses. Are the items too many with the result that respondents find it difficult to complete them? In that case, maybe it is better to reduce the items by removing the less crucial ones. In any case, it might just be possible that some items indirectly repeat one another; in which case it is sensible to identify similar items and merge them.

Where the tryout has shown that only minor modifications are needed, the revision task would follow. However, where in all honesty several problems have been detected, then the evaluation team had better settle down and redo all the materials and instruments. Indeed, the whole evaluation process may have to be revisited and modified. It is better to go back to the drawing board and plan the whole process all over than to go ahead and embark on a venture that is fated to fail.

Field trial

This constitutes the fourth stage. This stage, similar as it might be to the last, is distinct from it. This is the full-blown pilot stage. Whereas the last stage involved mainly the selection of a judgemental sample of a few subjects, this time a full random sample that is representative of the total population is necessary. Whereas the tryout stage required several instruments, mainly raw and calling for refinement, this time only a few virtually final instruments are needed, tried out on a large number of respondents. This stage entails the detailed analysis of data and results so that following this, the very next stage the final step can be confidently taken.

We should stress here that in most cases, stages three and four are often merged, especially in circumstances where funds are limited.

Implementation

This comes as the fifth stage. This is the stage at which the real fieldwork is done, entailing the collection of data with the instru-

ments earlier designed and refined for this task. This stage involves several research assistants, as well as evaluators working together to cover all the various facets of the target population.

Even where just a specific education institution is involved, the evaluators still have to combine their efforts in the data or information gathering exercised. Where several institutions constitute the focus of the study, very many hands are required. Usually, several days may be involved, and so efforts would have to be synchronised in as an integrated a manner as possible. This is the only way an equitable result may be expected.

At this stage, we should stress that many problems are associated with this phase. First, there is the problem of non-cooperation from the target population. This is there more likely among a population that does not understand or appreciate the purpose of the project. There are quite many instances when people are approached for their responses to some questionnaire items and the first question they ask is, 'How much are you going to give me?' Even some teachers do just that. And since you do not budget for this weird reaction, you may have to do without the opinion of such people.

There is also the problem of disturbances. When you plan for an exercise such as the data collection, you usually assume that everything would be smooth sailing. But you might be shocked to discover that just as you set out, everything goes berserk. The whole of the school system in a state may embark on an industrial action, the schools may be called on to observe a week holiday to mark an event not before thought of, you may enter a rural community and discover that half of the learners are absent because it is harvest time in the village, etc. Even after the period of data collection, or the tour of the institutions for observations, the analysis of data on computers could be fraught with its own problems. Many a project has been delayed by incessant power disruption or, more seriously, a total breakdown of the computer system. Problems such as these should be foreseen and made allowance for.

This should come as the logical end to all the earlier activities. The write-up should again be a team work, reflecting the inputs from all those participants in the course of the work. A good report should be based purely on the findings from the fieldwork and not tainted by one's own personal biases. An objective report should be well supported with data and figures, which are quite accessible to the understanding of the target population. Should the owners or sponsors of the project under evaluation be less sophisticated in academics than you the evaluators, then you should tailor the language, use of data and graphics to their level as well as ensure that you do not baffle them with the volume. Even for a government agency, there is clearly a limit to which your report could be academically sophisticated. In all, you should bear in mind that the clearer your report is, the better for all.

Many a time, apart from the submission of the report, it is often required that you hold a meeting with the owners or sponsors of the project. This need not be a very formal meeting, and the more relaxed the atmosphere is, the better. During such a meeting, some of the most salient findings, as well as the most pressing recommendations, could be discussed. This indeed could be as effective as the report and recommendations themselves. This way, understanding is engendered and rapport promoted. Such an exercise could in fact pave the way for a follow-up exercise in the area of quality control.

Actually, quality control should be a continuing process and not just a once-for-all event. Quality control is itself a big aspect of any economic venture, educational agencies or institutions inclusive. The ideal in most cases is to have a department that ensures quality as a part of the establishment itself. This is the norm in manufacturing establishments. In government educational establishments, the practice is to entrust quality control to the inspectorate division. This however works only to a certain limit. The problem there is that where the control of quality is from within, with time the in-

spectorate division itself soon calls for controls and indeed for serious shaking up.

And so, this takes us back to curriculum evaluation exercises on a regular basis. With one circle of curriculum evaluation exercise completed, the ideal is to have a follow-up on a not too distant a period. The end result is thus to do this regularly, virtually on a circular basis.

Evaluation of Institutional Programmes

Most of the curriculum evaluation activities have to do with whole institutions. This is precisely where professional evaluators come in. In a number of cases, proprietors of educational institutions do request that their ventures be evaluated to determine how well they are performing, what flaws there are that can be rectified, and how improvements can be made. Sometimes, though, such proprietors themselves do not make the move; it is often some others that suggest to them that they need the input of professionals. For instance, when ministry officials come calling and point out that their wards are not getting the ideal, it is then they realise that it is time to consult the experts.

When professionals are consulted, the team has to begin work. First, there should be a preliminary or investigatory visit to the institution with the express purpose of determining what exactly the proprietor wants. Quite often, he [or very often she] does not really know what the exercise would entail and what indeed he wants. There have been instances of proprietors simply saying that they would like to have regular visits from the group of experts to see how the pupils are learning and how their teachers are performing so that the latter can be guided to perform better. In many cases, the proprietor is ready to discuss the cost of this type of venture. It takes real professionalism to lead the discussion towards the need for a full evaluation of the whole institution and all the programmes connected with it.

That fixed, the team would need to recess and discuss how integrative the needed evaluation exercise should be, how long it should

take and what instruments would be needed. Of course, costing should be carefully decided and discussed with the proprietor.

In an ideal situation, the exercise should cover all facets of the institution from the teaching-learning environment, the teachers' workload, the background of the learners, the influence of the community in which it is situated, the teaching facilities available, the motivations available for learning, the learners' performances in the key subjects, etc. For each variable, a worthwhile instrument should be developed and tried out before final use. For each facet of the work, it should be very clear from the onset who would be in charge.

For most exercises of this nature, there should be an overall manager who does not necessarily have to go under that title. His duties should include planning thoroughly along with the team, managing the finances available for the project equitably, ensuring that individual members of the team perform assigned duties on time, establishing a smooth and working relationship with the awarder of the contract [for contract it really is], and ensuring that all members of the team are conversant with each step being advanced at crucial times. Ideally, there should be briefings from time to time. Ideally, there should be briefings from time to time so that the progress of the project can be discussed and adjustments made when deemed necessary.

The final stage of any project of this type is the report writing and of course the presentation of the report. Again, this should be a team work, with various members of the team addressing specific aspects of the project. Essentially, a good report should be seen as something that the awarder of the contract could fully understand and benefit from. For each finding and each problem identified, cogent solutions should be suggested in clear terms. This is the best way the whole exercise would be worthwhile to the consumer.

Thereafter, the contract awarder might decide to call back the team to help supervise the programmes on a regular basis.

Conducting any worthwhile education evaluation exercise in Nigeria is fraught with enough problems, many of which must have been addressed in this and earlier chapters. Much more serious than these are however those several problems which are the lot of curriculum evaluation exercises. Some of these are discussed here.

First, there is the problem of lack of cooperation from the populace, including those who otherwise should have been expected to understand. Even teachers and indeed heads of schools are often guilty of this. There was the case, for instance, of a principal of a school, who, during the data collection exercise for the national Education Sector Analysis project, ordered all the questionnaires brought to her school to be collected and brought to her. That done, she seized them and told the data collector to come back the following day. During the follow-up visit, she asked if the federal government was going to pay her, her staff and students, and the response not being favourable to her, she handed back all the unfilled instruments. This could be an extreme instance, but if it could happen from so enlightened a person, then the gravity of the problem could be appreciated.

Then there is the problem of statistical and other records. Ours is simply a nation with a very poor record keeping tradition. Individuals, institutions and organisations, more often than not, do not keep and maintain proper records. The situation is not helped with the occasional fire incidents in organisations, resulting in total loss of such records that have been kept, usually in one fell swoop. The ideal obtains where there are records that are from time to time updated, so that when any new exercise is to be conducted, there are enough to guide evaluators as to where to take off from.

Apart from that, there is the very pathetic phenomenon of purposeful distortion of data, even at official levels. In our country, it is common knowledge that most state ministries and agencies keep different types of data for different purposes. Thus, there are different pupils' data for different purposes, one for planning at the

state level, and another for federal government officials when they call. So, a researcher who calls for school population figures or for the numbers of teachers in schools, may not be too sure how authentic the figures given him are. To be doubly sure, he may have to get to schools himself, i.e. before going ahead to use the figures as the basis for the new research project.

Yet another problem is that of funding. Research exercises are never cheap. And the more ambitious they are, as curriculum evaluation exercises are, the more costly they inevitably would be. Unlike in most advanced countries, independent sponsors are hard to come by. A philanthropist would rather donate towards building a physical structure that bears his name than aid a research venture. Government ministries or agencies hardly ever do either. These otherwise would have helped matters in a country that does not put any premium on research and evaluation.

Reference

Lewy, A. [1977] "The Nature of Curriculum Evaluation", in Arieh Lewy (ed) *Curriculum Evaluation*. Paris: UNESCO Press.

Martin McLean (2002) Reference Library 2002. Microsoft Corporation on Computer CD.

CHAPTER EIGHT

APPLICATION OF COMPUTER TO DATA PROCESSING IN EVALUATION

A computer in an electronic device that is capable of assessing data /information by means of a set of instructions, process the information and produce a result. The information is passed into the computer through a component called the *input device,* processed by the *central processing unit (CPU),* also known as *process device* and the result is displayed on the *output device,*(visual display unit, VDU or a printer).

What makes a Computer better?

The main features of the computer are also seen as the advantages of using it as alternative to manual or any other forms of processing data include:

Speed: The computer works at the speed of electricity. Speeds of computer operations are usually measured in milliseconds, micro-seconds and pico-seconds, 10^3, 10^6, 10^9, 10^{12}, respectively.

Accuracy: Results from computers are highly accurate. Often times, we experience GICO (Garbage in, Garbage Out), this term is used to explain human source of errors.

Storage: A vital component of the computer is the *memory unit.* Memory is the capacity to store and release information when needed. This enables the computer to supply information to management timely and accurately too.

Flexibility: Modern computers can be used for a variety of purposes. For example, batch processing, on-line processing, real time, multi-programming. Computers can also operate on a variety of application packages.

Automation: the operations of a computer are automatic; there is no human intervention once a programme has been activated as a result of the stored programme concept.

Consistency and Efficiency: The computer is very consistent, i.e. it would produce the same result on the same data at any time if the same instructions were given. The computer does not suffer from human traits of being tired or losing concentration. It performs the task with the same speed and accuracy as the first task no matter the number of tasks involved.

Computer Hardware and Software

A computer is made up of two components (i.e. the hardware and the software).

The Hardware

Hardware is the name given to the physical component of a computer, for example, the visual display unit (monitor or screen), the system unit (also called central processing unit – CPU), keyboard, mouse, etc.

The Software

Software is the programme written to accomplish a specific task. There are two types of software; these are the system software and the application software. System software is a set of instructions that are written for or by the user in order to facilitate speedy operation of a specific task. Application software is also divided into two; we have the user's application programmes that are written in variety of programme languages by the user in order to accomplish a task and the software packages that are the generalised programmes for solving problems. Examples are the MS Word, MS Work, Spreadsheet, accounting packages, MS Excel, SCORBATT, SAS, etc.

History of Computers

Computer can be said to be as old as mankind. Toes, fingers, pebbles, stones and sticks were the only means of calculations in the olden days. The earlier calculating aids include Napiers Logarithm, Oughested's slide rule, Pascals Digital Counter, Leibnitz's Calculating Machine, etc. John Neumann designed what could be called the first computer in 1946. John Neumann's machine was prominent because of its stored programme concepts, a computer that will store the programme alongside the data.

Computer Generations

Many attempts were made at designing calculating machine / computer between 1946 and the late 1950s. The first attempts were the UNIVAC (Universal Automatic Computer) in 1951 and IBM (International Business and Machine) in 1953/54. One general feature of these computers is the use of vacuum tube. They are identified as first generation computers. Between 1959 and 1960, other sets were designed (they are referred to as second generation computers) using transistors instead of vacuum tubes. The transistors are smaller, less expensive, generate less heat and require little power. Examples include IBM 1402, IBM 1620, IBM 7090-7904.

The production of small scale and medium scale integrated circuit gave birth to 3rd generation computers between 1965 and the mid 1970s. This resulted in higher speed, better performance and compatibility of the computers (i.e. computers that can be used to work on the same jobs). In the 1980s, there was the production of very large-scale integrated circuits. This made it possible for microprocessors (which are the characteristics of the fourth generation computers) to be produced. A microprocessor is a central processing unit that is fabricated on a chip. Mass production of microprocessors led to the production of microcomputers, also referred to as computer on chips. Microcomputers were smaller, faster and more efficient than third generation computers. They were also affordable. This led to the development of a wide variety of software tools, such as Word processing, Spreadsheet, and Graphic Pack-

ages. Local Area Network, Electronic mail and other capabilities were implemented. The last generation of computers (i.e. the fifth generation computer) is a product of researches by world powers like USA and Japan. Super computers – computers which perform operations in the higher range of 10 billion instructions per seconds – were built. Also computers that could be made to think and act like human beings, i.e. having capacities for sight, hearing, etc., and which could simulate human thought (artificial intelligence) were designed. It is believed that we are currently in the fifth generation of computers.

How important is a Computer?

It is practically impossible to exhaust every area where computer has found its applications. Few examples will be cited here.

Business: Banks, insurance companies, manufacturing firms, stock brokerage firms, firms involved in importation, buying, selling, and production, are some of the areas in which the computer has been widely used. Also, they are widely applied for financial management, project management and human resources. Applications package and customized packages are used.

Government: The computer is widely used for census (for compiling census and survey data), licensing operations, e.g. for vehicle licence, driving licence, for tax computations, for compiling data on staff emoluments, for assessing taxpayers based on information supplied. Government can use computers for planning, i.e. to plan a project, to draw economic models and for forecasting, etc.

Police and military: Computers are used to keep records of crime and related cases. They are as well used for research and development of the military and for allocation of scarce resources for the upkeep of man, machinery and for monetary management in the forces.

Education: The computer has found wide ranges of uses in the field of education, e.g. for programmed instruction, as in computer aided instruction (CAI). It is also found very versatile for assessment

of students' performance. Information about students (e.g. name, age, height, academic ability, parents' background facts, etc.) can be stored on the computer.

Architecture and Engineering: In the design of bridges, buildings and machines, computers have been found very versatile. Virtually all the cars, trains, aeroplanes, household machines, television sets, etc., that are commonly used today are designed with the use of the computer. It might interest you to learn that it is very possible today to have a plane take off from one international airport and fly to another international airport in another continent, all without any pilot on board. In this respect too, it should be mentioned that the computer is found very useful for solving problems associated with complex calculations and precision.

Law: It is used for storing and retrieving information about previous cases. The computer is helpful to lawyers in consulting previous cases in order to argue new ones.

Medicine: Medical records of patients are kept in the computer. It can also be used for diagnosis and precision. Indeed, it is now possible to use robots (computerised machines) to perform complex but delicate surgical operations, and it has been claimed that a doctor using a computerised robot can perform an operation from afar off; e.g. in two distant cities.

In what way can we categorise computer?

Computers are classified using different parameters, some of which are listed below.

1. By function.

Based on the various functions it can perform, the computer can be categorised under three headings:

Analogue Computers: These are devices that measure in a continuous form. They are faster; good examples are the thermometer, speedometer, etc. They are commonly found in research and scientific research and scientific laboratories.

Digital computers: These operate with variables that have discrete values, they are flexible. They are widely used in commercial fields such as banking, accounting, and insurance.

Hybrid computers: These combine the features of both the digital and analogue computers. They combine the high speed of the analogue with the flexibility of digital. They are mostly found in scientific and technical applications.

2. ***By purpose***

This is describing computers by the purposes for which we want to use them. Here are some purposes.

General Purpose: This category of computers has a high level of flexibility. They can store programmes that can handle a wide variety of jobs. They can perform any kind of jobs, both business and scientific by changing its programmes.

Special Purposes: These are called dedicated computers. They are designed for a specific type of application. Programmes are in-built into this type of computers by the manufacturer so that the given task is performed quickly and efficiently. They cannot be used for any other tasks.

3. By capacity

Computers are classified by their size, that is, the amount of work they can do. Some of them are the micro-computers, mini-computers, mainframe and super computers.

Micro computers: These are the smallest in the range of computers. Microcomputers are also referred to as Personal Computers (PC). Majorities of earlier versions are called Desktop Computers but they now come in different sizes and shapes.

Today, we have different types of microcomputers, namely notebook, laptop, desktop, mini-tower and tower system. The microcomputer is built around micro-processor chips. A Microprocessor is single integrated circuit that contains arithmetic and logical (ALU) as well as control capabilities for memory and input/output access. The microcomputer has a keyboard for entry of data and monitor or screen for display purposes. Apart from its portable nature, ad-

vancement in technology has made microcomputers viable options to a more expensive and sophisticated mini and mainframe computers. Wide ranges of Software Packages are commonly run on microcomputers.

Mini Computers: These are smaller versions of mainframe computers.

Mainframe Computers: Mainframe is the most expensive of all these. It is very big in size and offers the maximum computing power. It is often called the maxi computer. It can support large numbers of peripherals that can be attached to it. It can be used in large networks of computers with the mainframe serving as the nodal point. It occupies large space and has capacities for running large programmes. It has a large main memory and performs operations very fast. They are housed in specially prepared places equipped with powerful air-conditioners. They are supported with UPS, are used for solving commercial and scientific problems. A typical application capabilities associated with mainframes is now implemented on Super Micros that now assume extreme processor power. Mainframe can be differentiated from mini computers physically because of size, processing capability, processor design, memory capability, number of peripherals, etc.

Super Computers: These are big, general purpose computers with high capabilities in terms of speed, storage capacity, etc. They are used to handle the manipulation of complex mathematical models requiring millions and trillions of operations to resolve. They are mostly used in simulations. Besides, they are used in the studies connected with the explosion of super weapons. Meteorologists use super computers to study formation of tornadoes and other weather conditions. Examples are CRAY – 2, CRAY – 3, CYBER – 810, CYBER -813.

What is the Relevance of Computer in Data Analysis?

We have already discussed the normal way of performing statistical operations manually in this book. Effort shall be made here to discuss application of computer to data processing in Education.

There are so many software packages in data analysis. Some of them are SCORBATT, SAS, SPSS, MSTART, and so on. The emphasis of this section is the application of SPSS to Educational Research. SPSS is an acronym for Statistical Package for Social Sciences. We have the DOS versions (versions 1 to 5) and the WINDOWS versions (Version 6.0 to the latest now, Version 16.0). Since the WINDOWS versions are learner friendly, attempt will be made to discuss a version in WINDOWS. A basic assumption here is that the readers are familiar with at least WINDOWS 95, since attempt will not be made to teach or discuss WINDOWS operations.

We can use computer for so many statistical tools. Explanations on how to use these statistical tools are enough to form a volume. Since this aspect of application of computer to data processing in Education is a chapter in this book, this section will be limited to such statistical tools as frequency, cross tabulations, mean and standard deviation (aspects of descriptive statistics), analysis of variance (ANOVA) and multiple regression analysis (aspects of the inferential statistics).

Before we discuss these statistical tools, there is a need to examine the commands we use in entering our data set, data definition, and data modification at the WINDOWS desktop.

File Menu Commands

What do file menu commands do? File menu contains the following:

Open: (Command allows you to open an existing or a new file).

Read ASCII Data: (This command allows you to transform ASCII data, i.e. data from DOS into the one for WINDOWS).

Close: (Command lets you close a file you are working with).

Save: (Command allows you to save a file you are currently working with).

Save As: (Command allows you to save a file you are currently working with using another name).

Print: (Command allows you to print a file you are currently working with.

Exit: (This command allows you to close down this application software).

There are other commands you will need to know in order to execute SPSS for WINDOWS, we may meet them later.

Statistical Menu Commands

There are two fields of statistics: the descriptive and the inferential. Attempt will not be made here on how to perform such functions as analysing data because of the limitation of this chapter. However, important issues on the computer itself will be discussed. Illustrations will be made from computer print outs of certain analysis in both descriptive and inferential statistics. Such illustrations will cover such aspects as frequency, descriptive t-test, cross-tabulation and analysis of variance.

The statistics commands are organised into submenu according to the type of analysis performed.

1. Descriptive Statistics

Frequency

Frequencies are used to make tables and displays that show how often different values of a variable occur in the data. You can also use it to obtain summary statistics that describe the typical value and the spread of the observations. For example, in a survey of teachers' perception of school effectiveness in Oyo State, a table can be made which shows how many of the teachers came from Oyo Central, Oyo South, Oyo North senatorial districts. From the table, one can tell that 40% of the teachers came from Oyo Central senatorial district, and so on. The information can be represented graphically. The bar chart display turns the table into a chart in which the length of the bar corresponds to the number of teachers in a particular senatorial district. For a variable like teaching experience, which can have many possible ordered values, one can use a histogram to see how often different values occur. The following printout is an example of frequency analysis.

		Valid			Cum
Value label	Value	Frequency		Percent	Percent
Oyo Central	1	362	40.2	40.2	40.2
Oyo South	2	272	30.2	30.2	70.4
Oyo North	3	266	29.6	29.6	100.0
	Total	900	100.0	100.0	
Valid cases	900	Missing cases 0			

LOC location		Valid			Cum
Value label	Value	Frequency		Percent	Percent
Rural	1	508	56.4	56.4	56.4
Urban	2	340	37.8	37.8	94.2
Semi urban	3	52	5.8	5.8	100.0
	Total	900	100.0	100.0	
Valid cases	900	Missing cases 0			

TYPE		Valid			Cum
Value label	Value	Frequency		Percent	Percent
Public	1	843	93.7	93.7	93.7
Private	2	57	6.3	6.3	100.0
	Total	900	100.0	100.0	
Valid cases	900	Missing cases 0			

Descriptive

Descriptive analysis is used to calculate statistics that summarise the values of a variable. For example, one can calculate the average teaching experience of teachers and see how much spread or variability there is around this average value. Thereby, one can calculate values above which and below which certain percentages of the cases fall. For example, we might find that 25% of our teachers have less than 10 years teaching experience, 50% have experience less than 20 years and 75% have experience less than 30 years. You should not use this procedure for variables, which are nominal. That is, don't compute average school location or average teacher gender. The following printout is an example of descriptive.

Number of valid observations (listwise) = 900.00

Variable	Mean	Std Dev	Min	Max	N. Label
TE	14.38	11.85	0	35 900.0	Teaching exp

Cross tabs

Cross tabs are used to count the number of cases that have different combinations of values of two or more variables, and to calculate summary statistics and tests. For example, if you want to know if there is a relationship between the region of the city in which a person lives and the person's willingness to buy your product, you can make a table that shows how many people say they would buy your product, and how many say they would not, for each of the regions. Or, you can look at the relationship between job satisfaction and job performance ratings separately for males and females.

2. *Inferential Statistics*

Independent samples of GENDER

Group 1	GENDER EQ 1		Group 2	GENDER EQ 2
t-test for	EFFECTIVE	Mean	Std Dev	Std Error
	No of cases			
Group 1	413	39.9395	8.387	.413
Group 2	472	39.4597	8.489	.391

[3]Pooled variance Estimate			[3]Separate Variance Estimate		
F 2-Tail [3]	t	Degree of 2-Tail [3]	t	Degrees of	2-Tail
Value Prob [3]	Value	Freedom Prob[3]	Value	Freedom	Prob
1.02 .802	.84	883 .399	.84	870.11	.399

The ANOVA Models submenu provides techniques for testing univariate and multivariate Analysis-of-Variance models. Simple Factorial performs an analysis of variance for factorial designs. You can specify covariates and choose alternate methods for partitioning sums of squares. E.g. a simple factorial design can be used to test if five different coaching methods result in the same average test

score for students in three different majors. In this example, the dependent variable is the test score, and the factors (the variables used to form the groups) are the coaching methods and the students' majors. You can test the hypothesis that the coaching methods are equally effective; that the majors respond in the same way, and that there is no interaction between the major and the coaching method.

In a simple factorial design, you test the effects of each of the factors individually, and of the interactions. The Simple Factorial ANOVA procedure allows you to specify the highest order of the interactions and then build a model containing all interactions of that order and lower.

If you want to include only some of the interactions of a particular order, you should use the General Factorial ANOVA procedure. If you have only one factor, you should use the One-Way ANOVA procedure.

General Factorial analyses more general univariate analysis of variance designs, allows great flexibility in specifying the model, and offers a great variety of statistical output. This command is in the Advanced Statistical option.

For instance, if you are interested in the effects of three incentive structures on the productivity of two classes of employees in four locations of your company, you might design a factorial experiment to test the hypotheses of interest. Since baseline productivity prior to the introduction of new incentive structures may be a strong predictor of subsequent productivity, you might want to use baseline productivity as a covariate. (That is, you might want to adjust the observed productivity for baseline values). You may also want to see whether the effect of the covariate is the same for the two classes of employees. (That is, you want to see if there is factor by covariate interaction).

The Regression Techniques

Linear is used to examine the relationship between a dependent variable and a set of independent variables. For example, you can try to predict a salesperson's total yearly sales (the dependent vari-

able) from independent variables such as age, education, years of experience, and sales territory. Or you can try to predict a student's score on the Graduate Records Exam based on undergraduate GPA, IQ score and major. Both the dependent and independent variables must be measured on an interval scale. Nominal variables such as religion, major, or region of residence must be recorded to binary (dummy) variables or other types of contrast variables. If you have collected a large number of independent variables, and want to build a regression model that includes only variables that are statistically related to the dependent variable, you can use one of the variable selection methods to select the independent variables. To see how well the regression model fits your data, you can examine the residuals and other types of diagnostics that this procedure provides. Print-outs of regression analysis are made similar to previously displayed ones.

Analysis from the computer can be highly reliable. Once the print-out has been obtained, the task of producing a high quality report becomes relatively easy. From the array of figures and data, highly attractive tables and graphs can be generated to support the findings, conclusions and recommendations. Indeed, nowadays, when a good report has been produced, it is usually not easy where to put the credit – to the computer programmer, or the analyst, or the writer. Where everything has gone well, each person connected with the work as well as the computer software programmer should have the credit.

Reference

1. Cowart, R. (2001) *How to do Everything with your PC.* Berkeley California – Osborne/ McGraw Hill.
2. Mueller, S. (1995) *Upgrading and Repairing PCS.* Indianapolis. Stone Sherif and Que Corporation.
3. SPSS *PC Statistics 4.0 for the IBM pc/X/AT and PS/2 Marija.* J. Norusis SPSS Inc.

REFERENCES

Alkin, M. C. (1970) 'Product for Improving Educational Evaluation'. In *Evaluation Comments.* Vol 2, 3

Ayodele, S. O. [1985] *Continuous Assessment Practice and the English Language Teacher.* Ibadan: Evans Brothers Publishers.

Chacko, I. (1999) Observation in Evaluation. In J. O. Obemeata, S. O. Ayodele, and M. A. Araromi (eds) *Evaluation in Africa.* Stirling Horden Publishers.

Cowart, R. (2001) *How to do Everything with your PC.* Berkeley California – Osborne/ McGraw Hill.

Ezewu, E. E. (1980) *Handbook on Planning and Teaching for Continuous Assessment.* Evans Brothers (Nigeria Publishers).

Lewy, A. [1977] "The Nature of Curriculum Evaluation", in Arieh Lewy (ed) *Curriculum Evaluation.* Paris: UNESCO Press.

Martin McLean (2002) Reference Library 2002. Microsoft Corporation on Computer CD.

Mueller, S. (1995) *Upgrading and Repairing PCS.* Indianapolis. Stone Sherif and Que Corporation.

Obanya, Pai. [1982] *Evaluation in Modern Language Teaching.* Institute of Education, University of Ibadan Press.

Okpala, P. N., C. O. Onocha & O. A. Oyedeji. (1993) *Measurement and Evaluation in Education.* Stirling Horden Publishers.

Oxford Pocket Dictionary [1999] *Oxford Pocket Dictionary. [CD edition on the computer]*

SPSS *PC Statistics 4.0 for the IBM pc/X/AT and PS/2 Marija.* J. Norusis SPSS Inc.

Tyler, R. W. (1949) *Basic Principles of Curriculum and Instruction.* Chicago: University of Chicago Press.

Yoloye, E. A. (1971) *Evaluation for Innovation.* Ibadan: I U Press.

TABLE OF INDEX