Indira Gandhi National Open University
Staff Training and Research Institute of
Distance Education

MDE-416
Curriculum Development for Distance Education

Block

4

CURRICULUM EVALUATION

Print Production: Mrs. Promila Soni, Section Officer (Pub.), STRIDE, IGNOU, New Delhi

August, 2014

ISBN-

Further information about Staff Training and Research Institute of Distance Education (STRIDE) and the Indira Gandhi National Open University courses may be obtained from the University's office at Maidan Garhi, New Delhi-110068 or at www.ignou.ac.in

Printed and published on behalf of the Indira Gandhi National Open University, New Delhi, by Prof. C.R.K. Murthy, Director, STRIDE, IGNOU, New Delhi.

Laser typeset by Mctronics Printographics, 27/3 Ward No. 1, Opp. Mother Dairy Booth, Mehrauli, New Delhi-30

Printed by :

MDE-416 : CURRICULUM DEVELOPMENT FOR DISTANCE EDUCATION

Course Outline

Block 1 : The Field of Curriculum

Unit 1 : Curriculum: The Concept
Unit 2 : Foundations of Curriculum
Unit 3 : Curriculum Issues and Trends in Distance Education

Block 2 : Curriculum Development

Unit 1 : Towards Curriculum Development
Unit 2 : Curriculum Planning
Unit 3 : Curriculum Designing
Unit 4 : Curriculum Implementation and Evaluation

Block 3 : Curriculum Transactions

Unit 1 : Instructional System
Unit 2 : Instructional Techniques and Materials
Unit 3 : Learner Support Systems
Unit 4 : Role of Distance Teachers in Distance Education

Block 4 : Curriculum Evaluation

Unit 1 : Concept of Curriculum Evaluation
Unit 2 : Techniques and Tools of Evaluation
Unit 3 : Construction of Evaluation Tools
Unit 4 : Evaluation of Distance Education Sub-systems

Block 5 : Curriculum Development Experiences

Unit 1 : Tertiary Education
Unit 2 : School Education
Unit 3 : Technical and Vocational Education
Unit 4 : Non-formal and Continuing Education.
Unit 5 : Materials Production
Unit 6 : Media and ICT in Teaching Learning: IGNOU Experiences

LET US BEGIN HERE

The course on the theme of 'Curriculum Development for Distance Education' is divided into five Blocks. This is the first one. It comprises three units in all. A schematic representation of the design of the unit is given below to facilitate your access to the content presented here.

UNIT X*

X.0 Objectives

X.1 Introduction

X.2 Section 1 (Main Theme)

x.2.1 Sub-Section 1 of Section 1

x.2.2 Sub-Section 2 of Section 1

.....

.....

.....

Check Your Progress

X.3 Section 2 (Main Theme)

x.3.1 Sub-Section 1 of Section 2

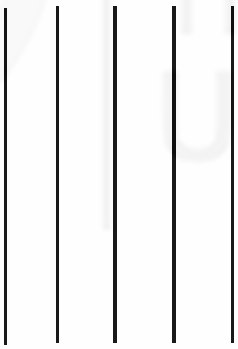
x.3.2 Sub-Section 2 of

.....

.....

.....

Check Your Progress



Answers to Check Your Progres

X.n Let Us Sum Up

* 'X' stands for the serial number of the unit required.

As the scheme suggests, we have divided the units into sections for easy reading and better comprehension. Each section is indicated distinctly by bold capitals * and each sub- section by relatively smaller but bold upper and lower typeface. The significant divisions within sub-sections are in and lower typeface. The significant division within sub-sections are in still smaller but bold** upper and lower typeface so as to make it easier for you to see their place within sub-sections. For the purpose of uniformity, we have employed the same scheme of 'partitioning' in every unit throughout the course.

We begin each unit with the section 'Objectives'. It articulates briefly what we have presented in the unit, and what we expect from you once you complete working on the units.

In the last section of each unit, under the heading, ‘Let Us Sum Up’, we summaries the whole unit for purposes of recapitulation and ready reference.

*__BOLD CAPITALS__

Relatively smaller but bold

**** still smaller but bold**

Besides, we have provided self-check exercises under the caption ‘Check Your Progress’ or ‘Self-check Exercise’ at a few places in each of these unit which invariably end with possible answers to the questions set in these exercises.

What, perhaps, you would like to do is to go through the units and jot down important points as you read in the space provided in the margin. (**Broad margins in the booklets are there for you to write your notes on.** Make your notes as you work through the materials. This will help you prepare for the examination as also in assimilating the content. Besides, you will be able to save on time. Do use these margins.) This will help you prepare for the examination as also in assimilating what you have been reading in the unit, answer the self-check exercises and the assignment questions and easily identify the item(s) to be clarified.

We hope that we have given enough space for you to work on the self-check exercises. The purpose of these exercises will be served satisfactorily if you compare your answers with the possible ones given at the end of each unit after having written your answer in the blank space. **You may be tempted to have a furtive glance at the possible answer(s)**, as soon as you come across an exercise. But we do hope that you will overcome the temptation and turn to possible answers (which are not the best answers necessarily) only after you write yours.

These exercises are not meant to be submitted to us for correction or evaluation. Instead, the exercises are to function as study tools to help you keep on the right track as you read the units.

On an average, each block will have at least one or a part of one assignment. At times an assignment may expect you to work through more than one unit to prepare you responses. You have to send your assignment responses to us for assessment and comments. In all, you may have to work on one assignment per course. Assignments are sent separately, and are changed every year.

We suggest the following norms be strictly practiced while you are working through the assignments.

- Write your roll number legible as indicated in the Programme Guide.
- Before you put down anything in words, assimilate what you have read and integrate it with what you have gathered from your experience to build you answer.
- Make the best use of the block and the additional reading materials for diligently working through the assignments.

BLOCK INTRODUCTION

The purpose of this block is to provide us with introductory knowledge in the areas of basic concept of curriculum and its implications in the curriculum. We have designed the block in such a way as to focus on skill orientation/application rather than on knowledge/theory building. The crux is that this block intends to help us acquire the basic issues in the field of curriculum and development in general open and distance education in particular. As the block addresses itself to quite diverse groups of clientele it is not unlikely that much of what has been presented might sound quite familiar to some of you. For others, the block may provide things which are new or unfamiliar. However, we should state here that we have pitched the information at a level that will cater to your needs – irrespective of the category implied above. Nevertheless this block contains such relevant themes as the concepts are the basics for curriculum development. There are three units in this block focusing on curriculum the concepts where you will be acquainted with the basic issues of curriculum. In this process you will be familiarize the issues with the other units you studied in other courses. Unit 1, we have discussed in detail the meaning of and the concepts associated with curriculum. Unit 2 attempts to give us a comprehensive picture of foundations of curriculum. In the unit 3, the historical growth and emerging issues and trends in distance education. Further, this unit specific the curriculum evaluation techniques most appropriate for determining the achievement of the objectives set. Besides, we suggest you to make a link with all MDE learning content wherever it is suitable to establish context, purpose and relevance to your understanding,

Mail us

At the end of this block, we have provided a feedback questionnaire. Please fill it after completion of this block and send it to us. Your feedback shall be highly useful for future revision and maintenance of the course. Also please take note of the time you devote in studying this block. May be you complete this block after 4-5 sittings. But for every sitting, kindly note the time separately so that you can categorically say how much time you took to read this block. You can send the feedback questionnaire by post or you can e-mail the same to: stride@ignou.ac.in. In the e-mail, please mark in the subject areas – FOR COURSE COORDINATOR-MDE-416. You may also contact for any difficulties related to the programme in general and MDE-416 in particular.

UNIT 1 CONCEPT OF CURRICULUM EVALUATION

Structure

1.0 OBJECTIVES

After carefully working through this unit, you should be able to:

- define curriculum evaluation;
- differentiate between measurement, assessment and evaluation;
- distinguish between the different types of evaluation; and
- describe the purposes and functions of evaluation.

1.1 INTRODUCTION

Education, as an institution in modern societies, has contributed much to human welfare, including intellectual, social and economic developments as well as the enlargement of knowledge and culture. The rapid changes in the socio-economic conditions of developing nations are responsible for the emergence of distance learning. Under the changing circumstances the learner can become a better learner at a distance, and all learners need help and assistance to develop better skills, knowledge and attitudes. Greater effectiveness and productivity of education can contribute still more to the advancement of nations throughout the world. But two questions arise here: 'How can we measure the effectiveness and productivity of education?' and 'How can we improve the quality of educational services?' To answer these questions we need to know the concept and role of curriculum evaluation in education. Evaluation of both the processes and products of education becomes imperative not only to (convince?) us as to what work is going on in the whole educational system, but also and more importantly, to know to what extent the goals of education have been achieved. In short, it helps to build an educational programme, assess its achievements and improve upon its effectiveness. It serves as an in-built monitor within the programme and helps to review, from time to time, the progress in learning. It also provides

valuable feedback on the design and the implementation of the programme. The curriculum has become broader based than before. Flexibility has been enhanced by providing more options. Learners are able to widen their horizons and equip themselves with a broader spectrum of knowledge in order to adjudge the various career opportunities available. Thus the role of curriculum evaluation in educational programmes is significant. Keeping this in view, we shall now look at the concept and role of curriculum evaluation. In this unit, we have given an overview of the evaluation processes and of the various types and functions of curriculum evaluation.

1.2 DEFINITIONS OF CURRICULUM EVALUATION

Evaluation can be defined in numerous ways. In one sense we evaluate all we do all the time. However, Curriculum evaluation is the process by which we attempt to gauge the value and effectiveness of any particular educational activity with regard to the student who takes part in it. For example, how much he/she knows about mathematics; the distance between your workplace and residence; a student's score in a science achievement test is 60; John got an 'A' grade in an assignment response etc. You can think of many other similar examples. All the above statements are based on some sort of evaluation. This is often relative to an implied standard of reference or normal condition. We make judgments either by observing something or by measuring it and giving it a quantitative/ qualitative value. We evaluate all we do all the time whether we are working in an institution, or watching a television programme. Most of our day to day assessments are informal and often at the subjective level but we do make attempts at being more systematic, organized and rigorous when participating in certain processes. For example, judging, marking and recording an essay competition in an institution, comparing the temperature or rainfall figures of major cities, etc. In these cases the evaluation process is more systematic and organized. So, the term **evaluation** here is used in a number of different ways. To put it succinctly, we can say that evaluation is '**to find out the value of'** or '**to judge the worth of'** or '**to estimate the magnitude and quality**' of something. There are a number of ways to define evaluation. We may define it in two distinct ways, or see it as existing in two different forms in relation to education:

i) Evaluation of learning outcomes (ELO); and

ii) Evaluation of learning systems (ELS).

Let us discuss them here.

1.2.1 Evaluation of Learning Outcomes (ELO)

Evaluation is the systematic process of determining the extent to which the specified instructional objectives previously identified and defined have been achieved. Here the process of evaluation is based on the relationship between the objectives of the content and its actual outcome. The actual outcome can measure the intended learning outcome. Ideally, the instructional objectives will clearly specify the desired changes in learners and the instruments will provide a relative measure or description of the extent to which those changes have taken place. The following figure (Fig.1.1) can give you a clear idea about the evaluation process.

Fig.1.1 should help you clarify a number of pertinent points regarding evaluation of learning outcomes.

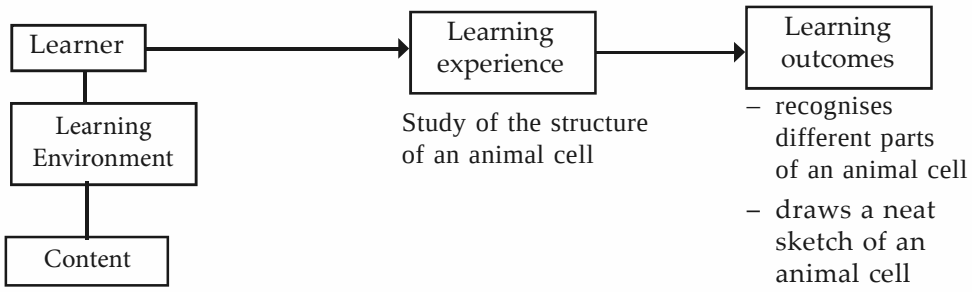


Fig. 1.1: The process of evaluation

This box-diagram should help you clarify a number of pertinent points regarding evaluation of learning outcomes.

Firstly, you can realize that objectives establish direction and that when they are stated in terms of learning outcomes they go beyond knowledge of the specific course content. You can note the difference between the '**study of**' and '**recognises**' cell structure. The process (study of an animal cell structure) includes the interaction of the teacher, learner, content and the learning environments to effect learning outcome (recognising different parts of an animal cell and so on).

Secondly, you can realize the varying degrees of dependence that the outcome, i.e., **recognises** and **draws**, has on the course content. The learning outcome **recognises** is a specific outcome of major instructional objective **knowledge**. The learning outcome **draws** is a specific outcome of the major instructional objective **skill**. From this discussion, it is clear that these specific learning outcomes, which are easily attained, can be observed and measured directly in an instructional system.

Thirdly, if the defined objectives are achieved, decisions will be taken in a particular direction to provide feedback. If they are not achieved then different decisions will have to be taken. The procedures/steps involved in ELO are:

i) Establishing objectives;

ii) Placing objectives in broad classification;

iii) Defining objectives in behavioral terms;

iv) Establishing situations in which attainment of objectives can be demonstrated;

v) Choosing or developing appropriate evaluation tools/techniques;

vi) Collecting data from the learner's performance;

vii) Comparing data with objectives;

viii) Making a decision about the learner's performance; and

ix) Providing feedback

Here, evaluation of learning outcomes (ELO) is concerned with the systematic process of determining the extent to which the instructional objectives are achieved, and also with the provision of feedback to the learners.

1.2.2 Evaluation of Learning Systems (ELS)

Evaluation, as we have maintained, is the systematic process of collecting and analyzing data in order to make decisions. Further, it refers to the process of obtaining and providing descriptive and judgmental information about the worth and merit of a programmer's goal, design, implementation and impact in

order to make informed decisions. This is otherwise known as decision-oriented evaluation.

The process of ELS involves the following steps:

i) Conducting a formal or informal survey for the assessment of needs;

ii) Defining the goals/objectives of the programme;

iii) Identifying the clients involved in the programme;

iv) Developing the programme materials, activities and administrative arrangements;

v) Choosing appropriate evaluation tools or techniques; and

vi) Collecting information related to the programme.

The basic difference between the two definitions i.e., ELO and ELS is the issue of making decisions in different situations by different persons involved in the system. In other words, we can say that ELO is a part of the daily affairs of a teacher or a learner and here we are concerned with the results or outcomes. But ELS is the concern of educational planners, administrators and researchers. ELS help in shaping, revising and replacing educational programmes to achieve better results. Let us now work out the exercise given below.

Check Your Progress 1

Notes: *a) Space is given below for your answer.*
b) Check your answer with the one given at the end of the unit.

Describe ELO and ELS.

.....

.....

.....

Let us now look at a few terms which are used synonymously but which have specific connotations.

1.3 DIFFERENCE BETWEEN MEASUREMENT, ASSESSMENT AND EVALUATION

Throughout the block you will come across terms such as ‘test’ ‘measurement’ and ‘assessment’ besides ‘evaluation’. Do these different terms mean one and the something? Very often they are assumed to do so and are used as though they are synonymous and interchangeable. But the fact is that they carry different functional meanings and as such, one cannot replace one with another without causing considerable damage to the sense conveyed. What follows is an attempt to discuss how these words are used in the field of education and training.

Let us not confuse the terms

Test is a device to obtain quantitative data. This consists of a standard set of questions to be answered or tasks to be performed. It is one among several devices (like holding interviews, sending questionnaires, assigning projects, rating scale, observation etc.) that can be used to measure the characteristics or traits present in an individual. It is the most commonly used form of measurement in education. A test usually presents a uniform set of tasks to all members of a given group at a scheduled time with due prior notice.

(Sometimes, 'unannounced" tests' are possible and useful in classroom situations.)

Measurement is an act of assigning a numerical index (i.e. quantification) to what is being assessed in some meaningful and consistent manner. For example, Suresh scored 32 out of 50 in an English test. When we assign a score 32/50 to a given test performed by Suresh, we are performing an act of **measurement**. Measurement only qualifies the data obtained through observations, rating scale, tests or any other devices.

Assessment refers to the process of obtaining information about a learner's learning, progress and achievement. That is, what and how well has the learner understood the content of a course? Upon completion of the course, a value is assigned to the performance. For example, Suresh got an 'A' grade or 75% in an assignment-response.

Evaluation is a wider and more inclusive term. This includes all the three: test, measurement and assessment. When we compare the score of a learner with those of other learners and judge whether it is good/average/ satisfactory/unsatisfactory/bad, we are performing an act of evaluation. Here is an illustration. To say Suresh has scored 32 out of 50 in a test is a report of measurement. But, to say 'Suresh is good in English' is an instance of evaluation. To be more clear let us take another example. A typist types 50 words per minute. Here, 50 is a symbol by which his/her ability is being measured. When we say that he/she types better than other typists in the office, we evaluate his/her typing ability. So in the case of evaluation, we assign a value judgement to measurement.

Measurement answers 'how much' whereas evaluation 'answers' 'how good', or 'how well'. Put differently, the latter adds meaning or value judgement to measurement. From what we have just now seen, we can note that evaluation includes three components. They are:

1) Quantitative description (e.g. Raman has scored 60 in Mathematics)

2) Qualitative description (e.g. Raman is able to solve mathematical problems systematically)

3) A value judgement (e.g. in view of the above facts, Raman is good in Mathematics).

In one sentence we can say that:

Evaluation = Quantitative description
(Measurement: facts and figures)

+

Qualitative description
(non-measurement: ranking, Weightage, value)

+

Value judgement.

The above discussion explains the comprehensive nature of evaluation and the role of measurement in the evaluation process. It goes beyond the simple quantitative description. Evaluation always includes value judgments concerning the desirability of the result.

Check Your Progress 2

Notes: *(a) Tick (✓) in the appropriate box.*

(b) Compare your answer with the one given at the end of this Unit.

Suppose after a three month course in typewriting two learners, say A & B, achieve a speed of 55 words per minute. However, when A joined the course, he had already achieved a speed of 45 words per minute, while B at the time of joining could type only 30 words per minute. Now which of the two terms - 'evaluation' and 'measurement' - will you use to describe the following statements about the achievement of the two learners?

	It is a statement of evaluation	**It is a statement of measurement**
a) The progress made by A is average	☐	☐
b) A has recorded a progress from 45 w.p.m. to 55 w.p.m.	☐	☐
c) The progress made by B is extraordinary	☐	☐
d) The performance of B has improved from 30 w.p.m. to 55 w.p.m.	☐	☐
e) At the end of the course both A & B performed well	☐	☐
f) At the beginning of the course B was a poor performer	☐	☐
g) At the end of the course, A & B performed at the rate of 55 w.p.m.	☐	☐

1.4 TYPES OF EVALUATION

In the previous sections we discussed evaluation as a process of assessing the worth of a product, a process or a programme. With reference to the processes, products and persons under consideration, different types of evaluation can be grouped as below:

- student evaluation
- institutional evaluation
- programme evaluation
- personnel evaluation.

The above four types should give us a global view of evaluation. But we can also state the types of evaluation according to functions, approaches and nature of reference because there is no one best way to classify evaluation. Every evaluation situation is unique. Evaluation needs to be situationally responsive, which is more helpful than a rigid orthodox form of evaluation. Based on functions, evaluation can be prognostic/diagnostic, according to approach, it may be formative or summative and with respect to the nature of reference, it may be criterion-referenced or norm-referenced evaluation. The above discussion provides us with a broad understanding of types of evaluation. Let us put all these into tabular form for a better understanding.

Table 1.1: Basic ways of classifying different types of evaluation

Types of evaluation	Function of the evaluation	Illustrative instruments
Placement	To Determine possession of prerequisite skills, degree of mastery of course	Reading tests; aptitude tests, pretests on course objectives; observational techniques.
Formative	Determine learning progress, provide feedback for reinforcement of learning and correct learning errors.	Teacher-made mastery tests, observational techniques.
Summative	Determine end-of-course achievement for assigning grades or certifying mastery of objectives	Teacher-made tests, performance rating scales; product scales.
Diagnostic	Determine causes (intellectual, emotional & environmental) of persistent learning difficulties	Published diagnostic tests, teacher-made diagnostic tests.
Criterion-referenced	Describe pupil performance in terms of a specified domain of instructionally relevant tasks	Teacher made mastery tests; observational techniques.
Norm-referenced	Describe pupil performance in terms of relative position in some known group	Standardised aptitude and achievement tests, teacher-made tests, interest inventories

We have already discussed above how one of the distinctive features of the evaluation process is the use of a wide variety of procedures. These may be described in many different ways, depending on the frame of reference used. We shall discuss here student, institutional and programme evaluation which will give us an overview of the subject.

1.4.1 Student Evaluation

In this case the evaluation process is addressed to each single student and to the group of students as a whole. The evaluation which concerns the individual student is based on information which defines his/her achievement, intelligence, personality, attitude, aptitude and interest. These traits can be measured with the help of individual teacher-made or standardised tests. This is otherwise known as *individual student evaluation.* The evaluation process which deals with the students as a whole has more or less the same purpose, but it concentrates on pointing out the difference between the subjects and the population. This is otherwise known as *group evaluation*.

The individual and group evaluation processes can function at three levels:

i) Individual/group initial evaluation;

ii) Individual/group intermediate evaluation; and

iii) Final evaluation.

i) Initial evaluation of an individual, or of a group, aims at evaluating the pupil entry behaviour in a sequence of instruction. This is also known as placement evaluation. For example, before the start of the proper course, the learners could be sorted out according to their abilities and placed in

a particular course. Placement evaluation is necessary for admission of students into a new course or programme. The initial evaluation can be based either on certificates, or on data resulting from objective tests or on information gathered in different ways. Collecting initial data has the purpose of starting the process of diagnostic evaluation. We find diagnostic evaluation in the case of technical education. Some institutions which want to programme their activity in a more constructive way, base their initial diagnostic evaluation on data especially obtained at the beginning of the year, by administering batteries of tests to the students. The results are defined as learning prerequisites. This has to be regarded as a useful tool in taking early decisions about the individualization of remedial work. The same situation arises in the case of distance education. In fact, it is advantageous to examine specific data obtained at the initial moment and to base diagnostic and prognostic evaluation thereon. In this case the educational decisions would be more reliable and more suitable to the individualization of the teacher activity. For example, before the beginning of the proper course, the students could be grouped according to their abilities and could be trained on the basis of specifically designed procedures to get the remedial work they actually need.

ii) Intermediate evaluation aims at the evaluation of a student's learning progress during the period of his/her learning activity. When the teacher teaches in the class and gives the students some new ideas, the teacher tries to evaluate the learning outcomes, on the basis of which he/she modifies the methods and techniques of teaching to provide better learning experiences. This helps us to find out whether the student needs some instruction to supplement or to compensate for his/her weaknesses. In the context of distance education individual intermediate evaluation aims at establishing the progress of each student in acquiring the abilities representing the objectives of the single parts in to which the course is divided.

This evaluation is called formative evaluation when its purpose is to modify the educational proposals and to make them more relevant to the individual student's needs. When the evaluation of the learning activity is based on the formative approach, it must be as analytical as possible. Suppose a student shows some weakness and is not able to reach the intermediate goals as they have been established during the planning of the course, it is necessary to know exactly in which of the objectives is the student failing and what the causes of the failure are. So intermediate evaluation does not only diagnose the difficulties but also points out the mistakes made by the student. The main tool for individual evaluation is formative tests. These tests are structured measuring instruments with specifies answers. The evaluation of the learner's progress can be measured with the help of periodical questionnaires, if the course lasts for a long time, through observation when the student personally contacts the educational structure or through informal communication such as letters.

Formative evaluation is a tool for providing feedback to the teaching/learning process. This type of instruction motivates the teacher and the student and reinforces learning by providing feedback to both of them.

iii) Final evaluation is done at the end of a course of instruction. It is the evaluation of the pupil's achievement at the end of instruction. This is also called **summative evaluation**. This helps a teacher to know how far the objectives have been accomplished and to what extent the instruction has been effective. Traditional university examinations are examples of summative evaluation. Teacher-made tests, standardised tests, rating scales are some of the widely used tools for summative evaluation. The individual

final evaluation is employed for assigning grades to students. Its chief functions are crediting and certifying the level of attainment of the pupil.

This is judgmental and terminal in character. (It is terminal in nature because it comes at the end of a course of instruction or a programme.) It views evaluation as a product. The chief concern of this type of evaluation is to point out the levels of attainment of students and to judge the efficacy of a programme or system and thereby guide us as to whether the programme or the system is to be accepted or rejected. Summative evaluation is concerned with the outcomes of learning or products of education.

Check Your Progress 3

Notes: *a) Space is given below for your answer.*
b) Check your answer with the one given at the end of this Unit.

a) Name the three levels of student evaluation.

i) .....

ii) .....

iii) .....

b) Fill in the gaps choosing appropriate words from those given after each statement.

i) Collecting initial data has the purpose of starting the process of ..... evaluation. (Summative, diagnostic, placement, formative)

ii) .....evaluation is concerned with the outcomes of learning. (Formative, placement, summative and diagnostic)

A closer look at the student evaluation in distance education in an open university

At this stage it will be useful to examine student evaluation in an Open University. For this purpose we take the case of IGNOU. In this university student evaluation takes place continuously during the course of the students studies. A distance learner while working through a course is expected to do the self assessment questions. He/she submits assignments and receives the tutor's comments and grades/marks. All the programmes end with a formal examination. From this discussion you may note that the Open Universities generally have a three-tier system of evaluating learner performance. This consists of:

1) Self-assessment;

2) Assignments (internal assessment); and

3) Term-end examination (external assessment).

Let us discuss them in the given order.

Self-assessment

The first component of student assessment comprises in-text questions given in the *course units* which are self-instructional in nature. Students are expected to work on these in-text questions on their own and having worked on a question, look for reinforcement in the answers provided in the materials themselves. These are the questions or exercises which do not carry any weight for passing the examination. The responses to these questions are not to be sent to the university or to any study centre for evaluation. This

component is called *'self-assessment'*. In other words, we can say that the process of 'self-assessment' serves the purpose *of formative evaluation*.

Assessment through assignments

The second component of student assessment is assessment through assignments. Assignments constitute the continuous evaluation component of a course and they play an important role in the two-way communication between the distance tutor and the learner. The students work on prescribed assignments. Students send written responses to the assignment questions to evaluators-on-contract at the study centres/headquarters of the university. Assignments are usually of two types:

- Computer-Marked
- Tutor-Marked.

Computer Marked Assignments (CMA) consists of objective type questions. Two sample CMA questions are given below:

l) Among the following, who could be included in the category of 'business person'?

   a) Chartered accountant

   b) Doctor working in a hospital (physician)

   c) Sales boy in a medicine (chemist) shop

   d) Owner of a grocery shop.

2) An example of an audio-visual medium of advertising is:

   i) Picture card

   ii) Poster

   iii) Film show

   iv) Radio

Students choose the correct answer and write/mark the number/alphabet against the correct answer in the box provided. Alternatively, they put a tick mark or darken the number/alphabet as the case may be.

Similarly, if a learner does not want to answer any question, he/she can put a (X) in that box. After completing the assignments, and filling in the response sheet, the students are required to mail it to the headquarters.

The tutor-marked assignments (TMA) demand written responses in the form of either essay type questions or short-answer questions or both. The learners, after going through the unit(s)/block(s) on which assignments are based, write the responses. Assignment-responses, as mentioned above, are sent to the coordinator of the study centre allotted to the student. It is compulsory for students to attempt the assignments. They will not be allowed to appear for the term-end examination of a course if they do not submit the specified number of assignments in time for that course. We should note here that for calculating the overall scores, scores of TMA or CMA or both, where mandatory, are put together.

Evaluation through assignments serves the purpose *of formative evaluation* and also the purpose of *summative evaluation* as the grades/marks scored in assignments constitute a component of the overall score a student gets in a course. This constitutes "*continuous assessment"*.

Term-End Examination

The third component of student assessment is based on the term-end examination, which in most cases is a conventional three-hour written examination. The term-end examination for various courses/programmes is ordinarily held in the months of June and December every year. Students are free to appear at any of these examinations (June/December) either for specific courses or for a whole programme provided that they complete the minimum period of study and the required assignments. This component serves the purpose of *summative evaluation*.

Grading

Grading involves expressing a distance learner's achievement/ performance in relation to a pre-specified criterion rather than in relation to the achievement of others enrolled in the programme. A learner who fulfills a certain criteria in every required task is said to have successfully completed the required tasks and so is awarded a grade/mark. The rest are considered unsuccessful.

In the computation of the overall score in a course, the weight assigned to continuous assessment may be between 25% and 30% of the total score and correspondingly the weight assigned to the term-end examination may be between 75% and 70% of the overall score. To complete a course successfully a student is expected to get at least a pass grade in both the continuous assessment and the term-end examination. The Open University follows a letter grading system on a 5 point scale (A, B, C, D and E). So the levels of performance, both in continuous evaluation as well as at the term-end examination of all programmes (degree/diploma and certificate) are indicated in letter grades (A, B, C, D and E).

The qualitative levels and grade points corresponding to various letter grades are shown in the following table:

Table 1.2: Qualitative values of letter grades and their equivalent grade points of Diploma and Certificate programmes

Letter	Qualitative Level	Grade Point
A	Excellent	5
B	Very good	4
C	Good	3
D	Satisfactory	2
E	Unsatisfactory	1

Check Your Progress 4

Notes: *a) Please write the answer in the given space below.*
b) Check your answer, with the one given at the end of this unit.

Explain the grading system of IGNOU in about 20 words.

.....

.....

.....

.....

.....

1.4.2 Institutional Evaluation

Every teaching-learning situation is different. It follows that every evaluation will be different too, as it must be tailored to meet the needs of individuals (both teachers and learners), of groups and of institutions. Many teachers are employed by large institutions and may be unable or unwilling to evaluate their own particular areas of work in isolation. An evaluation of the whole institution may be called for. Evaluation of an institution indicates some of the aspects of an institution which could be usefully evaluated and emphasises the need for effective control of the procedure of evaluation through a widely representative evaluation steering group.

It can be very beneficial to evaluate an institution in terms of its management, administration/ancillary services and the general environment. Great care must be taken to ensure that all interest-areas/groups are properly represented, unless everyone involved feels he/she has the opportunity to provide his/her opinions and concerns and contribute to decisions, the evaluation is most unlikely to be worthwhile. Much of the evaluation of an institution will centre around the work of teachers and the educational service they deliver. It can be carried out by small groups of staff. It is possible for each small group to facilitate decision making and to oversee the day to day processes.

But when an institution is to be evaluated, the process goes far beyond the examination of individual courses or programmes within it. All educational institutions may have existing data collection systems (often referred to as Management Information Systems). Many institutions are now adopting quality assurance systems including sets of questionnaires seeking students and sometimes employers' perceptions of courses. An evaluation steering group must take into account these existing systems and decide how to incorporate them into its own work to avoid duplication. It also has to examine the limitations of these systems and so avoid being restricted by them.

From the above discussion it is clear that institutional evaluation can be of two types:

i) System evaluation

ii) Course evaluation

System evaluation

System evaluation involves the evaluation of any instructional programme or instruction materials and includes evaluation of such factors as instructional strategies, textbooks, audio-visual material and physical and organisational arrangements. How many courses have been produced? How many students are there? How many applicants had to be turned away? etc.

This data can be drawn from administrative records and presented regularly, often in the form of an annual report or a 'Statistical Digest' in a system. In addition to measurements of activity come those of efficiency. How many students have successfully completed the courses? What amount of workload do they attempt? This data may come from administrative records and be produced as a part of regular monitoring procedures. So system evaluation has often moved beyond the descriptive to the examination of patterns and causes. A new programme is usually considered to be cost effective/efficient if one of the following is true:

i) it costs essentially the same as other programmes but results in a greater achievement.

ii) it costs less and results in an equal or a greater achievement.

ii) it costs more but results in a significantly greater achievement. Unless the difference is minimal, anything that costs more usually must be demonstrated to be practically better.

Course evaluation

The second major strand of institutional evaluation is 'course evaluation'. The aim of this is to improve the quality and effectiveness of the teaching and learning that takes place. The evaluation of a course or teaching materials may seek to provide information that can be used during the process of developing a course or preparing materials—these are **formative** evaluation procedures. But when we seek to provide information about how well the 'finished' instruction has worked in normal use—we use what are called **summative** evaluation procedures. Summative evaluation procedures of a course provide feedback from tutors and students. For example, tutors can give their own reactions to the teaching materials, and also accounts of the problems their students have encountered in their studies and assignments. The experience of tutors in making the course work can provide particularly useful information for subsequent modifications to or adoptions of the teaching materials and instructional arrangements. Feedback is gathered from conventional students i.e. those in the conventional system of education while they are in the classroom and responding to a teacher's questions. In the case of distance teaching, students' feedback is collected while they are taking a course or shortly after its completion. The students, feedback from one presentation in a course helps to determine revisions for subsequent presentations. The information can be collected through mail questionnaires.

Check Your Progress 5

Notes: *a) Space is given below for your answer.*
b) Check your answer with the one given at the end of this unit.

A new programme is usually considered to be cost effective/efficient, if it satisfies certain criteria. Please state the three requirements in this regard.

i) .....

ii) .....

iii) .....

1.4.3 Programme Evaluation

You may be aware that a programme or course evaluation starts with needs assessment besides gathering information regarding its marketability, utility, economic viability, adequacy and appropriateness of its content and media components. We shall here elaborate on the whole process of programme/ course evaluation. Institutions normally initiate some evaluation activities when launching new programmes. These include the need assessment of a programme on the basis of social demands, marketability, i.e., whether the programme is attracting sufficient number of students, cost-effectiveness of the programme, etc. According to the needs of the society and cost-effectiveness of a course/programme the utility of a programme in a country is also considered. For determining the adequacy and appropriateness of the content and media components of a programme, a few criteria commonly used at distance learning institutes are:

- What is the level of content-density?
- Is the language used appropriate to the level concerned?
- Is the material adequately self-instructional?

- Is the media utilisation pedagogically rational?
- Is the material socially relevant?
- Is the material easily accessible? etc.

Essentially, programme/course evaluation at Open Universities is a two tier operation. The first tier consists of those elements of evaluation which are conducted before launching a programme/course. For example, the need for a course and the rationale behind the outline of a particular course are looked into by an expert committee. The modifications brought in at this stage are again discussed and reviewed with the course writers. In the process of writing a course the materials get vetted by a programme/course editor, a language editor and a unit designer. The second tier of evaluation in the present context involves the piloting of programmes/courses. This process involves a trial of materials with the first batch of students, and revisions are brought in subsequently.

Programme evaluation comprises three components—evaluation of the planning of a programme/course, evaluation of materials that result from the plan and evaluation of the support services after the materials/courses have been launched.

i) **Evaluation of planning a programme**

This activity consists of evaluating a programme/course in terms of the need for it, its marketability, its utility, its economic viability, adequacy and appropriateness of its content and media components. For example, suppose we plan to study the economic viability of a programme or course. For this, various criteria are used to indicate how much is to be spent on preparing a course and what might be the returns. We have to see how much money needs to be spent on the various operations pertaining to a particular course such as cost of expert committee meetings, orientation programmes for course writers, paper and printing, audio and video materials, support services, and the cost of learning as far as the student is concerned. These cost-factors help in estimating whether certain courses are cheaper than others, and thus in establishing their relative cost-effectiveness. Obviously, such evaluation will yield very useful inputs for planning and decision making in an educational system.

ii) **Evaluation of materials**

Evaluation of materials can be conducted either during the actual process of developing them or after they have been developed, as a continuous evaluation. We may consider a few criteria while evaluating materials. They are content adequacy, language, media utilisation, social relevance etc. Continuous evaluation of materials can be done mainly by giving precise instructions to the authors, supervising them continually, evaluating pilot studies, releasing the finished materials, certification, etc. The most important advantage of continuous evaluation/formative evaluation is that it stands for the general possibility for improving the study material (Koul, 1991, p. 95)

iii) **Evaluation of student support services**

Evaluation of student support services is done to improve the quality of services, to allow managers/educators/tutors to take decisions about expansion and reduction of services, to share information about good practices and failures with a wider audience. This evaluation is a consideration of an aggregate of:

- tutor evaluation;
- evaluation of face-to-face session, counselling sessions;

- evaluation of continuous assessment, and
- evaluation of support system in general.

Tutor evaluation: This evaluation consists of assessing the nature of communication or interaction between the tutors and the students through letters or in face to face situations, the quality of their comments on assignment-responses, the reliability and validity of the assessment of assignment-responses and the turn around rate of assessed assignment responses.

In support services tutors/academic counsellors are essentially engaged broadly in four activities, which are:

- to inform, advise and counsel learners;
- to meet learners in face-to-face situations for a number of sessions during an academic year;
- to assess assignment-responses;
- to help learners to learn how they should learn independently.

These four activities are the bases of evaluating the students. This evaluation consists in gauging:

- the nature of communication/interaction between the counsellors and students. For example how far effective interaction was possible: through letters or in face-to-face counselling or telephone counselling or through teleconference or through some other medium
- the quality of tutor comments on assignment-responses;
- the reliability, validity and objectivity of the assessment of assignment-responses; and
- the turn around rate of assessed assignment responses.

Evaluation of face-to-face counselling sessions: Learners often need help and advice on matters that have nothing to do with the subject matter they are learning about. Sometimes he/she may need help in getting started on a course, organizing his/her time, coping with self-doubt, and deciding on the next step after completing a course/programme. We evaluate an academic-counsellor's punctuality, regularity, nature of rapport achieved with the students, quality of the conduct of sessions, students' satisfaction and tutors' motivation. This can be measured from the administration records maintained in study centres and evaluation divisions.

Evaluation of continuous assessment: Evaluation of continuous assessment consists of evaluating the validity and reliability of continuous assessment. This refers to the process of evaluation of assignments for their validity and reliability and for their correlation with the questions set in the end of term examination. Student evaluation takes place continuously during their studies. Distance students submit their assignment responses and receive feedback from their tutors in terms of specific comments, corrections, overall comments and a formal grade.

Evaluation of support system: Lastly, the evaluation of the support system itself involves evaluating the quality and quantity of support available to students in relation to various courses at various study centres. Such evaluation will have implications for the recruitment of academic-counsellors, providing support equipment, local and socio-geographical constraints and the corresponding support.

1.4.4 Personnel Evaluation

Evaluation of personnel includes evaluation of all persons responsible, either directly or indirectly, for educational outcomes. 'Personnel' includes not only teachers but also other groups such as counsellors and administrators. Personnel evaluation is generally based on the observation of the teacher/tutor in the classroom or face-to-face contact sessions by the head of institution/administrator/ coordinator and completion of a rating form by a head or an administrator or a coordinator. In some cases the teacher may be asked to complete a self-report rating form or students may be requested to rate their teacher on a number of factors. The validity of the observation approach, as typically used, is sometimes questionable on a number of dimensions. For example, the number of observations made per teacher is very small and therefore a very small sample of behaviour is observed. Sometimes people have a tendency to act differently than they usually do just because they are being observed: teachers may perform better/may feel nervous or perform poorly. So there is rarely any evidence regarding the validity and reliability of personnel evaluation if a particular method or tool such as observation is employed. From the above discussion, we can say that it is very difficult to determine what types of behaviour are to be evaluated. The solution to the personnel evaluation problem, at least for the present, involves collecting the best data possible, from as many sources as possible.

Check Your Progress 6

Notes: *a) Space is given below for your answer.*
b) Check your answer with the one given at the end of this unit.

a) What are the three components of programme evaluation?

i) .....

ii) .....

iii) .....

b) State any three criteria for evaluating instructional materials.

i) .....

ii) .....

iii) .....

c) How is continuous assessment carried out in distance teaching systems?

i) .....

ii) .....

iii) .....

1.5 PURPOSES AND FUNCTIONS OF CURRICULUM EVALUATION

Evaluation plays a key role in education. It helps to know the efficacy of different educational programmes. Through evaluation, defects and deficiencies can be identified and appropriate plans of action can be formulated. However, we may have to identify a number of specific purposes and the related functions of evaluation to make it meaningful. We have given below the most common purposes for which evaluation is used within different educational systems.

Prediction: This is the purpose served when evaluation helps us discover potential abilities and aptitudes among the students. In a secondary/high school class or public examination, one has to decide whether or not one examination often forms the basis to judge future courses of study. In higher education evaluation often forms a basis for students to continue with their chosen course of study or to find employment. Evaluation results are used as an important factor for selection for employment and for entry to higher education. In technical and commercial training, success at one level opens the possibility of training at a higher level. In all of the above cases, evaluation is used to predict future potential and it helps to assess the future performance of the student to select the right electives.

Diagnostic: The purpose of evaluation is to help us identify the learning difficulties experienced by individuals or groups. In classroom situations teachers constantly attempt to identify the weaknesses/lapses in learning. In distance education the students are not physically present in the educational situation and the study materials have been planned and carried out in advance. The collective initial evaluation during the planning and carrying out of a course, or at its beginning, is to set a diagnostic evaluation to find out the possible difficulties in the future. This might help us in producing suitable materials for a given group of learners.

Selection: This will be the purpose of evaluation when it helps us select suitable persons for a course (or career). The entrance tests to different courses are devised with this purpose in view.

Grading: This purpose is served when evaluation helps us rank the learners of a given group. Usually the term-end examination of a course serves this purpose.

Guidance: This is the purpose of evaluation when it assists a person in making decisions about courses and careers. The purpose of guidance is served when evaluation enables a learner to know his/her pace of learning or lapses made while learning. The self assessment exercises in distance education materials are meant to serve this purpose.

Decision-making: Evaluation of a course/programme helps in making decisions about the 'worth' of a course/programme. The results of programme evaluation are used to continue or discontinue a course/programme.

Feedback: Evaluation results provide feedback to students and educational programmes in order to bring about an improvement in the system.

Check Your Progress 7

Notes: *a) Space is given below for your answer.*
b) Compare your answer with the one given at the end of this Unit.

1) Fill in the blanks with appropriate words.

   a) ..... helps to discover the potential abilities and aptitudes of students.

   b) The purpose of ..... evaluation is to identify the learning difficulties experienced by individuals.

2) State the guidance function of evaluation.

   .....

   .....

   .....

1.6 LET US SUM UP

In this unit, we have discussed the types, purposes and functions of curriculum evaluation. We have stated that

- Curriculum evaluation can be defined as the systematic process of determining the extent to which the specified instructional objectives previously identified and defined have been achieved.
- There are four types of curriculum evaluation with reference to processes, products and programmes. They are student evaluation, institutional evaluation, programme evaluation and personnel evaluation.

Student evaluation: the process of obtaining and providing information for making decisions and judgments about a pupil's performance.

Institutional evaluation: the process of delineating and obtaining information useful for making decisions about the instructional programmes of an institution.

Programme evaluation: the process of the collection and use of information to make decisions about an educational programme.

Evaluation of personnel: evaluation of all persons responsible, either directly or indirectly, for educational administration (counsellors, teachers and so on).

In terms of student performance, we have said that the purpose of evaluation is to determine the current status (grading) of a student's performance, to compare that status with a set of standards for selection, to predict the potential of the studies, to identify the weaknesses and lapses in learning, to suggest any possible improvements after obtaining the feedback and to select an alternative among two or more to make a decision.

1.7 ANSWERS TO CHECK YOUR PROGRESS

Check Your Progress 1

a) Evaluation of learning outcomes is the systematic process determining the extent to which the specified instructional objectives previously identified and defined have been achieved.

b) Evaluation of learning systems is the process of collecting and analysing data in order to make decisions about the worth of a programme and its subsystems.

Check Your Progress 2

Evaluation a) c) e) f)

Measurement b) d) g)

Check Your Progress 3

a) i) individual/group initial evaluation

ii) individual/group intermediate evaluation

iii) final evaluation

b) i) diagnostic

ii) summative

Check Your Progress 4

Grading involves expressing a learners achievement or performance in relation to a pre-specified criteria. Open Universities like IGNOU generally follow a letter grading on a 5-point scale (A,B,C,D and E) for diploma and degree courses.

Check Your Progress 5

i) It costs essentially the same as other programmes but results in greater achievement.

ii) It costs less and results in equal or greater achievements, unless the difference is minimal.

iii) It costs more but results in significantly greater achievement.

Check Your Progress 6

a) i) Evaluation of the planning of a programme/course.

ii) Course materials.

iii) Student Services

b) i) Content adequacy

ii) Language

iii) Media utilisation

c) Continuous evaluation takes place in distance teaching systems through assignments. Distance learners submit their assignment responses to the study centre which are sent for tutor comments. The tutor provides feedback and awards a grade to each assignment-response.

Check Your Progress 7

i) a) Predictive

b) Diagnostic

ii) The purpose of guidance evaluation is to assist people in making decisions about the courses and careers available to them.

UNIT 2 TECHNIQUES AND TOOLS OF EVALUATION

Structure

2.0 OBJECTIVES

After going through this Unit, you should be able to:

- define the observation technique, describe the types and the steps to be taken to observe, record and interpret an observation;
- define an interview and describe the various types of interview;
- explain the uses and limitations of interviews;
- define profiles and describe the steps involved in preparing profiles;
- define the rating scale and describe the various types of rating scales;
- explain the uses and limitations of rating scales;
- define projects and describe their types, uses and limitations; and
- distinguish between the different types of tests and describe their uses and limitations.

2.1 INTRODUCTION

You may recall from the previous Unit that evaluation involves decision making. We can make rational decisions if we have necessary information or data. Various tools and techniques are used for collecting information. In order to select the most appropriate tool and/or technique for a given evaluation situation, it is

necessary to acquaint ourselves with various types of tools and techniques. Selection of tools and techniques is critical to the whole evaluation process. It does not matter how carefully you have planned an educational activity or an effort if inadequate instruments are involved, because poor tools can lead to 'poor' decisions. This means that the selection of an appropriate tool for evaluation purposes is essential, though we may be familiar with the wide variety of tools that exist. In this Unit we present to you a few criteria which are applied in selecting one tool from among alternatives. This will serve the major purpose of this Unit: to provide an overview of the various types of tools and techniques used in evaluation processes.

Let us begin the discussion now.

2.2 OBSERVATION

Observation is a technique which deals with the external behaviour in controlled or uncontrolled situations. It deals with recording the changes taking place during the process of occurrence of a phenomenon, i.e., individual, event or object. In case of human beings, there are certain traits like honesty, punctuality, persistence, truthfulness, etc., which can hardly be measured objectively through paper-pencil tests. This being the case, observation is an appropriate technique to measure the 'change'. The behaviour of the learner in the classroom, in the playground, in the institution, among his/her peer group, in social situations can be observed. For example, you could ask the students about their sportsmanship and you could ask the teachers how they handle inattentive students in their classrooms, but more objective information would probably be obtained by actually observing students at a sporting event and teachers in the classroom. In the context of distance education if the instructional system of an institution includes face-to-face tutorials, the behaviour of tutors as well as the students' behaviour can be included in evaluation (Video films of the tutorial sessions, for instance, can be made and also be used in tutor training). An observation is purposeful when it is well planned, carefully focussed and thoroughly recorded. Working within a particular environment over a period of time, an observer may become so familiar that he/she subconsciously becomes selective in what is seen. Validity increases by keeping the setting as natural as possible. By making observation more systematic, it is often easier to focus on particular aspects and to collect specific information.

Observational data especially that gained through participant observation permits the evaluator to understand a situation or a programme-setting to an extent not entirely possible using only the insights of others obtained through interviews. Of course, not everything can be directly observed or experienced, and participant observation is highly time-consuming and, relatively, an expensive evaluation strategy.

The primary purpose of observational description is to take the reader of an evaluation report into the programme-setting that was observed. This means that observational data must have depth and detail. The data must be highly descriptive so that the reader can understand what occurred and how. The evaluation observer becomes the surrogate eyes and ears of the reader. The descriptions must be factually accurate and without irrelevant matter.

To sum up, we can say that the first criterion to apply to a reported observation is the extent to which that observation permits the reader to enter into the programme situation observed. Evaluation data collection through observation is demanding work. Validity in qualitative method depends to a great extent on the skill, competence and rigour of the evaluator because the observer is the instrument.

2.2.1 Types of Observation

There is a spectrum of styles of observing: from the non-structured to the highly structured, and from the observer as participant in the activity, to the observer as non-participant, perhaps even using a video camera. Two types of observation are generally used in an evaluation effort. They are:

i) **Non-structured observation:** Non-structured or open or natural observation allows the collection of a rich variety of information. In such situations the observer, while being present at the site, does not control or manipulate anything. It is a technique of many social anthropologists. Here the observer enters the observation site with as open a mind as possible. He/she can see the things and record them in a natural setting. However, practice and skills are needed by the observer in drawing his or her experience and judgement to focus upon, and record events considered being important. It must be realised that however experienced the observer, only a fraction of the interactions and events occuring will be seen and recorded. For this reason, it is often valuable to focus upon certain aspects of the event, object, etc., being observed. For example, behaviour in a counselling session -- behaviour of the counsellor, behaviour of the learners, and the interaction between the counsellor and learners — can best be assessed through natural observation.

ii) **Structured observation:** An evaluator observes the phenomenon under structured conditions, with the knowledge of the person(s) being observed. The observation situation may also be simulated and observed. This technique allows the evaluator to observe particular behaviours. For example, a teacher, trainee observes the role play in a teacher-parent conference, etc. The major disadvantage of this type of observation is that it is not natural, and the behaviour exhibited by people may not be the behaviour that would occur in a natural setting. People may behave the way they think they *should* behave rather than the way they normally would. A checklist to be used during the observation process is valuable to have accurate records of the behaviour expressed.

2.2.2 Recording Information from Observation

Information may be recorded through a variety of methods, some of which are:

- field notes;
- observation schedules or checklists;
- audio recording; and
- video recording.

These can be used individually or collectively.

2.2.3 Uses and Limitations of Observation

Some of the uses and limitations of observation are given below:

- It helps us to get first hand information;
- Certain traits like honesty, punctuality, truthfulness, etc., can be observed;
- Structured observation yields objective and accurate data;
- The observer codes and records the overt behaviour at the time of its occurrence;
- However, the tendency of an observer to let overall feeling towards an individual or an initial impression affect subsequent observations; and
- A ‘subject’ may intentionally attempt to exhibit artificial behaviour.

Check Your Progress 1

Notes: *a) Space is given below for your answer.*
b) Check your answer with the one given at the end of this Unit.

1) List the types of observation.

a) .....

b) .....

2) State any three traits which can be assessed with the help of these observation techniques.

.....

.....

.....

2.3 INTERVIEW

We use the term 'interview' to indicate a process of communication or interaction in which an interviewee gives information verbally in a face-to-face situation. An interview is essential to assess certain inner (mental) traits. Interviews in a congenial atmosphere where personal rapport has been established can succeed in bringing out the inner feelings of the interviewee. The main objective of an interview may be the exchange of ideas and experiences, the eliciting of information pertaining to a wide range of data in which the interviewee may wish to rehearse his/her past and, present, and canvass his/her future possibilities.

Interview questions

An ideal qualitative interview is normally semi-structured. To prepare for an interview one should construct an interview protocol that allows for maximum flexibility during the interview process. One should formulate two or three lead-off questions before an interview. These leadoff questions help to open up a topic domain that one wishes a subject to address. They should be formulated very concretely and abstract questions should be avoided.

When writing these questions, we may try to anticipate the possible directions in which the conversation could go and then formulate possible questions. Thus, while formulating interview questions the method should be two or three lead-off questions for each domain, a list of questions covering the covert categories for each domain, and a set of possible follow-up questions for each domain.

Interviewer responses

It would be helpful to categorise interviewer responses into types. The different types of interviewer responses are:

i) Bland encouragement: Usually one word and/or facial expressions that show attention, interest and acceptance. These are excellent for establishing rapport and encouraging the interviewees to keep talking.

ii) Non-leading leads: Sometimes leading questions may not elicit more material on a certain topic and to indicate interest and attention one has to add something, like "Oh tell me more about that! This is interesting. Keep telling me about it."

iii) Active listening: Active listening responses are the establishing of a rapport and helping subjects to open up about certain things.

Interview analysis

Interview analysis, of course, will have many unarticulated but referenced meanings associated with it. To uncover them we need only to go through the steps and correlate the interview conversation as per the content of the subject. The final analysis developed in a qualitative way will draw upon strips of foot-notes and associated segments of the interview manuscript.

In the above discussion, we have treated observation and interview skills as two separate qualitative data collection techniques. In practice, however, they are typically fully integrated approaches.

2.3.1 Types of Interviews

Interviews may be classified according to the purpose for which they are used and according to their design and structure. For purposes of research, an interview may be used as a tool gathering data required by the researcher to test a hypothesis or solve his/her problems of historical, experimental, survey or case study type research. This type of interview is called 'research interview'. In many situations the objective of an interview is to secure information about individuals' problems, their past history, job or family adjustment. Here the major purposes of interviews are diagnosis and treatment. This type of interview is termed a 'clinical interview'. It is used by social workers and psychiatrists.

In some situations, an interviewer may interview one individual at a time. It is called an 'individual interview'. Further, telephone interviews are used when information is needed in a short span of time. In a 'group interview' a group of individuals is interviewed by an interviewer. Group interviews have been more effective with students who have completed a particular course. Interviews are classified as 'structured' and 'unstructured'. A 'structured interview' involves the interviewer asking specific predefined questions. These questions are carefully planned and the major areas of inquiry are mapped out. However, the interviewee is given considerable freedom to express his/her opinion. In this type of interview the interviewer uses a highly standardised form as a 'directive interview' because often the interviewee is directed to say 'yes/no' or give very brief answers. Unstructured interviews are also designated as 'uncontrolled', unguided, or 'non-directive'. In this type of interview, the interviewer does not follow a system or list of predetermined questions. Used with skill, unstructured interviews can yield information which may not emerge when we use any other technique. Group interviews may provide valuable insights especially in situations where people have differences of opinion.

Sometimes, the interviewees are encouraged to relate their concrete experiences with no or little direction from the interviewer, to dwell on whatever events seem significant to them, to provide their own definition of their social situations, or reveal their opinions and views as they feel fit. Although the unstructured interview is conducted through an informal discussion, a series of questions to be asked and the procedure to be followed are decided upon in advance. The interviewer is largely free to arrange the form and timing of the questions. He/she can rephrase the questions, modify them, and add some new questions to his/her list.

This technique is very useful for collecting information in the context of improving learning. The interviewer takes advantage of the flexibility of non-structured techniques, to maintain the desired focus. For example, few staff members in an institution of higher education attend an 'induction course'. Formal evaluation of the effectiveness of the course is undertaken half way through and/or at the end. Results from these, help to incorporate changes into the present course and help in the design of future courses. One technique used to collect information is the unstructured interview. If it is in a counselling session,

a schedule of issues to be raised during interviews is drawn up by the counsellors in consultation with the learners. Questions like 'Do you consider the course to have been useful to you?' are open-ended. Questions like "How long have you spent discussing the course with your learner?" are specific in nature. The latter types of questions generate more information than the former.

2.3.2 Techniques of Interviewing

We have discussed a few techniques of interviewing here.

Preparing for an interview

It is necessary to plan carefully for an interview. The interviewer must decide what kind of data the interview should yield, whether the structured or unstructured type of interview will be more useful, and how the results should be recorded.

Conduct of interview

An interview is a stressful occasion on which every effort should be made to generate a calm atmosphere. You may not get a true picture of the 'subjects' if they find the situation intimidating. So, ensuring a cordial environment is an important task. Interview processes will be more effective if:

- the interviewee is continuously reassured;
- interruptions during the interview are avoided; and
- seating arrangement should be given a thought, so that the candidate is not at a physical disadvantage.

Some general rules for conducting interviews:

i) Ask only one question at a time;

ii) Repeat a question if necessary;

iii) Try to make sure that the interviewee understands the questions;

iv) Be a good listener, remembering that smiles of encouragement and a friendly gaze show that you are interested;

v) Make sure the questions and answers stick to what is relevant. Allow the interviewee sufficient time to answer the question;

vi) Avoid suggesting answers to questions; and

vii) Do not show signs of surprise, shock, anger, if unexpected answers are given.

Closing the interview

After you have worked your way through your plan, the interviewee should be given an opportunity to ask questions. It is important to summarise and outline the next step. For example, approximately how long it will take to make a decision and how will it be informed. The interviewee should be finally thanked warmly for his/her co-operation and for attending the interview.

During the interview, information should be properly gathered. It is easy to record information arising out of a highly structured interview. The use of a tape recorder during the conduct of the interview not only eliminates the omissions, distortions, elaboration and other modifications of data usually found in a written interview, but it also provides an objective basis for evaluating the adequacy of the interview data in relation to the interview. If the tape recorder is not available, the use of a schedule, a structured

formal questionnaire or a rating scale may be used. An open or unstructured interview is less easy to record. It is so easy to hear what one wants or expects to hear and to forget the not so desirable points made.

Objectivity can be aided by having more than one person recording the information and discussing the recorded information with the respondent.

2.3.3 Uses and Limitations of Interviews

i) An interviewee provides an opportunity to the interviewer to question the interview thoroughly in various areas of inquiry.

ii) An interview is not an entirely independent tool of research for gathering information pertaining to feelings, attitudes or emotions. It is supplementary to other tools and techniques.

iii) It is an effective tool for a social scientist in the study of human behaviour.

Check Your Progress 2

Notes: *a) Space is given below for your answer.*
b) Check your answer with the one given at the end of this Unit.

Explain the techniques of interviewing in about 30 words only.

.....

.....

.....

.....

.....

2.4 PROFILES

A profile is a portrayal of the characteristics of a person or institutions in relation to some kind of activity or another. Profiling is the task of recording information which has been commonly used in industries for many years for purposes of staff appraisal. Within the last 20 years profiles have acquired some importance in education also, especially in the schools and further education sectors in Britain. A single grade as a measurement of a person's overall performance is regarded as unsatisfactory or 'hazardous'. Profiles may therefore provide a means of overcoming this difficulty, and may also be used to give information about personal qualities and interests.

Profiles can provide a record of what has been achieved at any point of a course. Profile of the learner includes information about the learner's educational background, socio-economic status, the cultural and environmental organisation etc. In distance education these aspects of a learner should be considered for specific evaluation activity by specialised services. The individual initial evaluation is aimed at defining the profile of each learner enrolled in a distance course. This gives an idea of and information about the learners' cognitive framework at the beginning of the learning activity (placement evaluation). It also gives you an idea of the changes in the attitudes, skills, knowledge and experience after the learner engages herself/himself in a course during the intermediate period (formative evaluation), and helps you to know about the learners' achievement at the end of a course (summative evaluation).

Profiles provide a focus for guidance and counselling and continuous feedback to the learners. They give us the end statement or a quantitative value. They are not a method of assessment but a tool for recording information. So profiles are considered a prerequisite for evaluation activities.

In our context, a profile is a panoramic representation alpha-numerical, graphical or verbal — of how a student seems to his/her assessors across a range of assessment methods. The use of profiles carries no stipulation about what should be assessed or how. They can be used in reporting any assessment. In other words, they are the means of recording information. Profiles can record the following information about a learner:

- enrollment in specifications;
- year of admission;
- experiences/job/employment;
- achievements; and
- Personal bio-data.

One example of an academic profile in use in higher education in Birmingham is the Graduate Profile used by the Birmingham School of Architecture. Profiles are useful in a number of ways but chiefly, they help identify aspects of a student's work and ability which would not otherwise be systematically recorded. They contribute to the students' knowledge of themselves and help future employers and colleagues know them better. Profiles do not claim to be predictive, nor do they mark the limits of a student's potential.

Profiles explicitly state that they are internal records and will not be released outside without the written agreement of a student. Each student receives a copy of his/her own profile document. Thus the profile is private.

2.4.1 Types of Profiles

Grade profiles

These relate to examinations. Examinations in higher education are split into a number of sub-tests. Generally, it is argued that a better description of the candidate is possible if the results of the individual sub-tests are included in the score. In the 1970s this possibility was examined for GCE Engineering Science at A level in Britain and it was A Grade in 7 point scaling: ultimately it was concluded that it would cause more confusion than the single grade.

Criterion profiles

These have long been used in technical and professional education. One of the intentions of these student-profiles is to record student-progress. In that sense they are diagnostic. They help provide meaningful information to students in such a way that they can see realistic hurdles which have to be jumped if they want to improve their performance.

Disadvantage of the criterion form of profile is that it is scaled, and, as soon as that happens, the subjective view of the examiners comes into play. This suggests that anyone who is being assessed with forms of this kind should be assessed by two examiners at least, who should then agree as to the final score.

Learning profiles

At the University of Sydney, the Department of Anatomy devised a nine-week course using two complementary teaching techniques: self-instruction and group interaction.

During the course at the end of each of the seven modules (7 weeks) the students were given tests which measured recall, comprehension, application, and short-chain problem-solving. At the end of the course there was a summative examination designed to test the capacity to solve long-chain multi-step problems. This was continued for four years. The results represent a measure of performance or learning curve which the authors of the course call 'learning profile'. The learning profiles of each of the students were put together to obtain regression curves of the group as a whole. In addition, general ability, age and interest data were obtained. An anxiety test was administered and a learning styles test was conducted during the first and seventh week of the course.

An academic assessment profile is a multi-dimensional end-statement expressing the results of the formal assessment of a student's performance.??

This tells us that learning profile is:

i) an end-statement — it is not a method of assessment but an official *pronouncement* about a student made at the end of his/her course or at some other specified period.

ii) multi-dimensional— the end-statement comprises observations of more than one features of a student's abilities or achievements, unlike the classified degree which is undifferentiated.

iii) academic — it is not concerned with the non academic characteristics or activities of a student.

iv) assessment — it is confined to reporting the results of the (formal) academic assessment undergone by the student, and neither describes nor reproduces the work he/she has done.

2.4.2 Uses and Limitations of Profiles

Uses of a profile:

i) It is a means of recording information.

ii) It covers a wide range of experiences and skills and encourages a great variety of assessment techniques.

iii) Records and reports are meaningful for both learner and teacher.

iv) Profiles help to know the nature and sequence of student learning and assessment.

Limitations

i) It is just a process of attaching labels to learners rather than helping the learning process.

ii) Broad descriptions of skills and crude grades are open to casual judgments.

Check Your Progress 3

Notes: *a) Space is given below for your answer.*
b) Check your answer with the one given at the end of this unit.

1) Define a profile in about 25 words.

.....

.....

.....

2) Explain the two types of profiles.

i) Grade profiles (20 words)

.....

.....

.....

.....

.....

3) Explain the two types of profiles.

ii) Criterion Profiles (15 words)

.....

.....

.....

.....

.....

2.5 RATING SCALE

'Rating' is a term applied to an expression of opinion or judgement regarding some situation, object, character, or an attribute. A 'Rating scale' refers to a 'scale' with a set of points which describe varying degrees of the dimension of an attribute being observed. Rating scales can be self-report instruments or observation instruments depending upon how they are used. These are different types of rating scales such as:

i) numerical scales;

ii) graphic scales;

iii) standard scales;

iv) rating by cumulative points; and

v) forced choice ratings.

We shall discuss them below in the same order.

Numerical scales

In a typical numerical scale, a sequence of defined number is supplied to a rater or to an observer. The rater or the observer assigns to each stimulus to be rated an appropriate number in line with these definitions or descriptions. For example, the following scale may be used in obtaining ratings of the affective values of colours:

10) Most pleasant imaginable

9) Most pleasant

8) Extremely pleasant

7) Moderately pleasant

6) Mildly pleasant

5) Indifferent

4) Mildly unpleasant

3) Moderately unpleasant

2) Extremely unpleasant

1) Most unpleasant

0) Most unpleasant imaginable.

The use of negative numbers is not favoured, as those observers or raters who are not well versed in Algebra find it difficult to manage negative numbers.

Numerical rating scales are the easiest to construct and to apply. They are also the simplest in terms of handling the results. However, numerical scales are often rejected in favour of other types of scales because it is believed that they suffer from various biases and errors.

Graphic scales

The graphic scale is the most popular and the most widely used type of rating scale. In this scale a straight line is drawn, vertically or horizontally, with various clues to help the rater. The line is either segmented into units or is continuous. If the line is segmented, the number of segments can be varied from case to case. Given below is an example of such a scale.

How effective was the teacher in the class?

Very Effective	*Slightly Effective*	*Average*	*Slightly ineffective*	*Very ineffective*

There are many advantages in graphic scales. They are simple and easy to administer. Such scales are interesting to the rater and require little added motivation. However, scoring in the case of some formats of graphic scale is rather laborious.

Standard scales

In standard scales, a set of standards is presented to the rater. The standards are usually objects of the same kind to be rated with pre-established scale values. For example, scales of handwriting provide several standard specimens that have previously been spread over a common scale by some standardised procedure like regular intervals. With the help of these standards specimens, a new sample of handwriting can be equated to one of the standards judged as being between two standards. The 'man-to-man scale' and the 'portrait-matching' scale are the other two forms that conform more or less to the principles of standard scales.

Rating by cumulated points

The unique and distinctive feature of rating by cumulated points is its immense and easy utility of scoring. The rating score for an attribute object or individual is the sum or average of the weighted or unweighted points. The 'check-list method' and the 'guess-who technique' belong to this category of rating. 'Check-list methods' are applicable in the evaluation of the performance of personnel in a job. The weights of +1 and -1 are assigned respectively to every favourable and unfavourable trait and the individual's score is the algebraic sum of the weights. In the 'guess who technique' some statements in terms of some 'descriptions' like "here is one who is always doing bad things to make others sad", are constructed and each individual is asked to list all the members of his/her group who fit such a description, mentioning the same individual as many times as necessary. Each individual

scores a point for each favourable or unfavourable description applied to him/her, and the total score is the sum total of all such points.

Forced choice ratings

In 'forced-choice rating' methods, the rater is asked not to say whether the rate has a certain trait or to say how much of a trait the rate has, but to say essentially whether he has some or one trait or another of a pair.

Uses of rating scales

i) Rating methods consume much less time than other methods of scaling like 'pair comparison' and 'rank ordering'.

ii) Rating methods are quite interesting to the raters, especially if graphic methods are used.

iii) Best ratings can be obtained by presenting one stimulus to a rater at a time.

iv) Rating scales can be used with raters who have very little training in this area.

v) Rating methods can be used with a large number of stimuli.

vi) Rating scales have a much wider range of application and can be used for teacher-ratings, personality ratings, school appraisals, sociological surveys, etc.

Limitations of rating scales

Rating scales have several limitations. Some of them are discussed here:

i) **Error of leniency:** There is a constant tendency among the raters to rate those whom they know well or in whom they are involved higher than they should. Such raters are called 'easy raters'. Some raters become aware of the temptation of easy rating and consequently rate individuals lower than they should. Such raters are called 'hard raters'. The leniency error refers to a general and consistent tendency for a rater to rate too high or too low for whatever reasons.

ii) **Error of central tendency:** Most of the raters hesitate to rate the individuals on the extremes of the scale, instead they tend to rate the individual on the middle of the scale. Obviously, the results get distorted.

iii) **Halo-effect**: Halo-effect is an error which obscures the clusters of traits within an individual. The rater forms a general opinion about the person's merit and his/her ratings on specific traits are greatly influenced by this general impression. It results in a spurious positing of a correlation among the traits which are rated.

iv) **The logical error**: is due to the fact that raters are likely to give similar ratings for traits which they feel are logically related to each other.

v) **The contrast error:** The contrast error is due to the tendency of a rater to rate others in the opposite direction (contrasting) from himself/herself to each other.

The proximity error: It has been seen that adjacent traits on a rating trend to inter-correlate higher than remote ones, their degree of actual similarity being approximately equal. This error may be counteracted to some extent by placing similar traits further apart and the different ones close together.

Check Your Progress 4

Notes: *a) Space is given below for your answer.*
b) Check your answer with the one given at the end of this unit.

List the types of rating scales.

.....

.....

.....

.....

.....

2.6 PROJECTS

Projects are a feature of education in certain subject areas. The dissertation in the field of humanities is at times considered equivalent to a project in science. Projects are advocated for interdisciplinary studies because real-life situations require a variety of areas of knowledge and resources for their solution and also because people have had to work in teams. Projects have been used in mathematics to open up real situations for investigation and have also been employed to encourage students to read literature. For example, a study conducted in Napier College, Edinburgh, a librarian and a biologist have collaborated to help students to prepare three projects of increasing complexity involving detailed literature searches. The study claimed that this technique increased the student's ability and willingness to read scientific literature, and that these attitudes persisted throughout the course.

Project work done by a student has not been confined to full time study. It has also been used in open education, covering varieties of problem areas, including evaluating a programme. Interestingly, a study shows that project work in the context of open/distance education found favour among staff and students. Some of the problems found (during the study mentioned above) seem to be general, and these relate to the amount of time required by tutor and student for guidance and work. The role of the supervisor at both graduate and undergraduate levels has been identified as an issue. Open University students need guidance in choosing a viable topic and in identifying, locating, and collecting information. This is also true of many students in undergraduate and school courses.
The British Open University finds, as do many others, that the amount and organization of individual support is problematic. For some departments it could be too costly. There is no doubt that some project work can be expensive. The general impression of project work is that it is favored by many students who are motivated and it provides 'independence'. Some, however, suggested that project work is not all plain sailing.

Advantages

- It provides motivation to the students.
- It develops independent thinking in the learner.
- The student develops the ability to formulate a problem and solve a problem while working in a project. The ability to formulate a problem is a crucial skill: it separates those who can do projects from those who cannot.

Limitations

i) "Own-tutor effect". Tutors may give their own students higher marks.

ii) The criteria for rating may vary from tutor to tutor and often reflects their specialization /research interest.

iii) The advice offered may be non-specific and characterisation of grades may be couched in general phrases.

iv) Rank ordering is difficult if the projects are dissimilar.

v) Unreliability of grades does exist.

vi) It is difficult to discriminate between a structured, well defined project area and wide-ranging unstructured projects.

vii) Tutors' specialisms may influence their perception of particular projects.

viii) The criteria used by motors? When marking projects will affect the reliability of the grades awarded.

Check Your Progress 5

Notes: *a) Space is given below for your answer.*
b) Check your answer with the one given at the end of this unit.

List the limitations of 'project work' in about 20 words.

.....

.....

.....

.....

.....

.....

.....

2.7 TESTS

Tests as tools of measurement are concerned with the product of learning behaviour. Different types of tests are in vogue to facilitate the realisation of the different purposes of education in the varying contexts of use. They may be categorised along three lines of approach. The approaches may be:

i) purpose-specific categorisation of test-types;

ii) mode-specific categorisation of test-types; and

iii) process-specific categorisation of test-types.

2.7.1 Types of Tests

We should note here that tests of different types may require the very same kind of tasks and sometimes even the repetition of the very same item. That is to say that it is quite possible to find an item recurring in tests which otherwise belong to a different type, or even a different category. What distinguishes the test-types, then, is not what is obvious in them but what guided them into being what they are. It is neither in the choice of test-tasks,

nor in the realisation of these tasks in the form of test-items that we are to find the difference. We are to notice the difference in the overall design of the tests, the purpose that guided them in their construction and, sometimes, in the nature and extent of coverage of a given area of learning. With this understanding we can now discuss the features of design of different test-types.

Purpose-specific category

Purpose-specific category includes tests designed to achieve a specific purpose of evaluation. Generally four test-types are identified in this category:

- diagnostic tests;
- aptitude tests;
- achievement tests; and
- proficiency tests.

Let us briefly present the features of each of these.

Diagnostic tests: These help us identify the areas of learning in which a learner needs a remedial course. They give us a profile of what the learner knows and does not know in a given area of learning. To present such a profile, a diagnostic test has to consist of a battery of a number of sub-tests each covering one area fairly thoroughly.

Aptitude tests: These tests serve a predictive function. They help us identify potential talents. They identify the prerequisite characteristics which are essential for one to be competent to perform a given task. Presenting items on such sub-skills as may eventually be developed into expert complex skills, these tests identify those who can do well in the field of study or a profession and those who cannot. These tests are generally used while selecting people for special courses/careers.

Achievement tests: These tests aim to measure the extent to which the objectives of a course have been achieved. The scope of these tests is governed by the objectives of the given course and they cover only the areas of learning demarcated by the given syllabus.

Proficiency tests: These tests aim to assess the general ability of a person at a given time. Their scope is governed by a reasonable exception of what abilities learners of a given status (say, matriculates or graduates) should possess. It is not restricted by considerations of the areas covered in any specific course-objectives or syllabus as in the case of achievement tests. While the usual end-of-course examination in a school or college may be taken as a typical example of an achievement test, a national level selection or admission test for candidates coming from different states and/or university jurisdictions can be taken as a typical example of a proficiency test.

Mode-specific category

Under the mode-specific category, we identify test-types on the basis of the mode/attitude that governs the construction and use of tests. Under this category, we present six pairs of test-types along six dimensions.

Formal assessment vs. Informal assessment: Formal assessment is applicable to a situation where a body answerable to the public is holding a test for a selection or an award. Assessment in such a situation has to ensure objectivity, credibility and relevance. To ensure these, it will have to follow a set of standardised norms/ procedures of test construction, administration and interpretation. Informal

assessment is applicable to situations where an individual or a voluntary body is holding a test to obtain some information to fulfill some personal requirements. The informal assessment also needs to be objective and reliable, but the valuator is not bound to satisfy the public about these qualities of his/her assessment. Hence the process of assessment need not follow very strictly the set procedures of evaluation.

Formative assessment vs. Summative assessment: Formative assessment is concerned with identifying learner weakness in attainment with a view to helping the learner and the teacher overcome/remedy those, while summative aims at certifying the grading the attainment of the learner at the end of a given course. Tests for formative assessment are given at regular and frequent intervals during a course, while the tests for summative assessment are given at the end of a course (or at the end of a fairly long period, say a term or a semester or a year). In a course that extends over six months, a test at the end of, say, every fortnight, will be a formative test, while the test at the end of the sixth month will be summative.

Moreover, the level of generalisations sought by the items of a summative test will be much higher compared to that sought by the items of a formative test. For instance if the items of a formative test check the ability to apply a given rule or principle to a given unfamiliar situation, the items in a summative test may check the ability to apply one or more of the appropriate rules/principles from among the many given in a variety of situations.

We would like to include two notes here.

i) The account of formative assessment and summative assessment given here belongs to the context of EIEP. Yet the terms are also applied to the context of EOEP. However, in the context of EOEP, the functions of formative and summative assessments are different. Formative assessment here includes tests and other forms of measurement which are intended to give a measure of success of the parts of a course - even as the course is in the process of development. Summative evaluation includes such forms of measurement that would give a measure of success of the course as a whole.

ii) In distance education yet another term is in use, besides formative assessment and summative assessment. It is ‘developmental assessment’. It is used in the context of course development and refers to the evaluation of the preliminary versions of courses with a representative sample of learners. It is treated generally as a part of the course development schedule. Formative assessment in this context refers to the evaluation made of a course (when it is produced) with a larger group (actual, not sample) of learners. The purpose of such an assessment is not to help the process of course development (as with the developmental assessment) but to help the activities of maintenance and revision of courses already developed.

Continuous assessment vs. Terminal assessment: While progress or achievement in learning is the concern of formative vs. summative mode of assessment, it is the purpose of grading learner achievement which guides the continuous vs. terminal mode of assessment. Continuous assessment seeks to spread the basis of grading on a number of tests with regular, even intervals, instead of placing it on one end-of-the course test (terminal test). Continuous assessment, thus, allows for more intense accommodation of the learning-content in the test process than the terminal assessment normally does. Scores on a series of continuous assessment tests, taken together, can serve for summative assessment. Taken individually, a continuous assessment test may be used formatively at the time of its administration. In the same way a terminal assessment may serve the purpose of formative assessment for follow-up courses.

Course work vs. Examination: Learner assessment can be based on work(s) performed by them during or at the end of a course, or, it may be based on examination(s) taken by them during or at the end of the course. Evaluation of course work or examination at different points of time during a course can be compiled at the end of a course to serve the purpose of summative evaluation.

Process vs. Product assessment: The basis for evaluation may be either the final product or the result of a given task, or the performance at different stages leading to the accomplishment of the task (as in a research work). While evaluating a learner, one may look for the correct solution to a given problem or take into consideration the correctness of the successive stages followed to solve the given problem (as in problem-solving tasks). If we do the former, we are supposed to be engaged in product assessment, if we do the latter, we are supposed to be making a process assessment.

Internal assessment vs. External assessment: The mode of assessment is external when the evaluation of a learner's ability is made by an outsider—a person who is not related to the actual process of teaching. The evaluator and the learner are anonymous and unknown to each other in this case. When the assessment is made by a person, responsible for affecting the learning being measured, it becomes internal assessment. Formative and summative assessment of both scholastic and non-scholastic abilities is possible in the case of internal assessment. External assessment serves only summative evaluation of scholastic abilities.

These are actually different perspectives along which assessment of learner ability can be thought about and planned accordingly. It is possible to practically combine two or more of these perspectives in one's approach to assessment. For example, one may include both course work and examination as the basis for learner-assessment and these two may constitute the units of continuous assessment. Or, one may include both internal and external assessment to serve the purposes of formative and summative evaluation.

Process-specific category

Sometimes test-types are identified on the basis of the process of test-construction. We can talk of two pairs of contrasting test-types here:

- teacher made test vs. standardised test, and
- norm-reference test vs. criterion-reference test.

Teacher made test vs. Standardised test: Standardised tests are commercially produced tests adhering meticulously to certain procedures to meet the demands of objectivity and accuracy. They are finalised through the construction procedures of formulating objectives, designing test-blueprints, employing item trials, item-analysis and item-revisions. The teacher made tests, on the other hand, are not governed rigidly by such processes. The teacher who makes the tests uses his/her discretion in matters of the scope of test area and choice of task-types and items. Standardised tests derive their name from the fact that they ensure standardisation of the procedures of administration, scoring and interpretation through elaborate specific instructions.

While a teacher made test is designed to operate within the restricted situation of a given classroom (in terms of test purpose, construction and use) a standardised test is designed for a larger operational situation crossing the barriers of a classroom, and institution or even a region. A standardised test may be chosen for use by different teachers/institutions in different classrooms, on different occasions and in different regions.

Norm-referenced test vs. Criterion-referenced test: In a norm-referenced test (NRT) the purpose is to discriminate between the high-achievers and the low-achievers. Its focus is not on what one has learnt or how much one has learnt of a given chunk of learning. Its focus is on where one stands in relation to the others. It assesses the ability of one against the standard 'norm' of achievement of one's fellow testees.

A criterion-referenced test (CRT), instead, assesses one's ability against the standard 'criterion' of what has been set as an acceptable level of ability-demonstration. That is while a CRT compares the testees' performance with a set standard of performance; a NRT shows the relative position (of attainment) of a testee with regard to the other testees who took the test with him/her. This might help administrative purposes of selection through rank-ordering. But a CRT can specify in behavioural terms the ability of a testee.

2.7.2 Uses and Limitation of Tests

Tests help in:

a) Providing knowledge concerning a student's entry behaviour;

b) Setting, refining, and clarifying realistic goals for each student;

c) Evaluating the degree to which the objectives have been achieved; and

d) Determining, evaluating, and refining the instructional techniques.

For example, aptitude and intelligence tests provide information concerning the speed and ease with which a student can be expected to learn. Achievement tests provide information as to whether a student is weak or strong in a particular discipline/subject. For more specific information regarding the deficiency, diagnostic tests are used.

Tests are also used to confirm a student's ideas about the skills, abilities or personality characteristics.

Tests of intelligence or special aptitudes should not be considered the measures of pure intelligence or creative thinking because performance in such tests is partly determined by one's background and schooling.

Tests measuring cognitive processes can hardly be measured as higher mental processes such as the ability to discover scientific laws and principle.

Check Your Progress 6

Notes: *a) Space is given below for your answer.*
b) Check your answer with the one given at the end of this unit.

i) Name of four types of purpose specific category tests

a) .....

b) .....

c) .....

d) .....

ii) Explain formative and summative assessments in about 15 words each.

.....

.....

.....

.....

.....

.....

2.8 LET US SUM UP

In this Unit we have discussed a few techniques and tools of evaluation such as observation, interview, profiles, rating scales, projects and tests.

Observation refers to a technique in which one or more persons observe what is happening in some real-life situation. It is used to evaluate the overt behavior of individuals in controlled and uncontrolled situations. As a good evaluation tool, observation needs proper planning, expert execution and adequate recording and interpretation. Observation may be either participant or non-participant, and structured or unstructured.

Interview is a process of interaction in which the subject provides the information verbally in a face-to-face situation. Preparation, conduct and recording are the main steps of the interview technique.

A profile is a portrayal of the characteristics of a person or institution in relation to some kind of activity. There are three types of profiles. They are grade profile, criterion profile, and learning profile.

A rating scale refers to a 'scale' with a set of points which describe varying degrees of the dimension of an attribute being observed. There are five types of rating scales.

A project is a discrete activity. The dissertation in the area of humanities is at times considered equivalent to a project.

Tests are tools of measurements and they guide us in evaluation. There are three different types of tests. They are process-specific, mode-specific and purpose specific.

2.9 ANSWERS TO CHECK YOUR PROGRESS

Check Your Progress 1

i) a) Non-structured observation

b) Simulated observation

ii) Honesty, punctuality, and truthfulness.

Check Your Progress 2

The techniques involved in interviewing are preparing for the interview, the conducting of the interview and closing the interview. The interviewer should plan carefully for an interview and for the collecting and recording of the information. There are some general rules for conducting interviews. These are: asking only one question at a time and not showing signs of surprise, shock or anger if unexpected answers are given.

Check Your Progress 3

i) A profile is a portrayal of the characteristics of a person or an institution in relation to some kind of activity or another. It is a panoramic representation—alpha-numerical, graphical or verbal—of how a student appears to his/her assessors.

ii) a) Grade profiles are related to examinations. Examinations in higher education are split into a number of sub-tests. Generally, it is argued that a better description of the candidate is possible if the results of the individual sub-tests are included in the score.

b) Criteria profiles have been used in technical and professional education. One of the intentions of these profiles is to record student progress. In that senses they are diagnostic. They provide meaningful information to students seeking to improve their performance.

Check Your Progress 4

a) Numerical scale

b) Graphic scale

c) Standard scale

d) Forced scale

Check Your Progress 5

a) Own tutor effect

b) Criteria effect

c) Rank-ordering

The criteria used by tutors when marking projects will affect the reliability of the grades awarded.

Check Your Progress 6

i) Diagnostic

Aptitude

Achievement

Proficiency

ii) Formative assessment is concerned with identifying learner weaknesses in attainment with a view to helping the learner and the teacher overcome/ remedy those, while summative assessment aims at certifying and grading the attainment of the learner at the end of a given course.

UNIT 3 CONSTRUCTION OF EVALUATION TOOLS

Structure

3.0 OBJECTIVES

- After going through this unit, you should be able to:
- identify the principles of test construction;
- explain the processes involved in test construction;
- describe and differentiate item formats; and
- list the quality of a test and identify the different approaches to use a test.

3.1 INTRODUCTION

Educational testing involves four stages of activity: (i) planning the test structure; (ii) constructing the test; (iii) administering the test; and (iv) assessing and interpreting the learners' performance. At the planning stage decisions have to be taken with regard to the objectives, choice of content area, choice of skills/abilities, the length/duration of the test, etc. At the construction stage the choice of item formats for the proposed test-points, the nature of sampling, the sequencing and grouping of items, drafting of instructions, etc. are to be decided upon. At the administration stage the main concern is to provide the appropriate condition, facilities, accessories, etc., uniformly to all learners who take the test.

At the stage of interpretation, conclusions with regard to the testees' performance, ability, their achievement in terms of the standards set, their relative standing in the group, etc., are to be arrived at.

These stages of activities are not independent of one another. But rather they are closely interdependent. Test construction is not only guided by a test plan but also governed by considerations of available conditions of administration. The test-plan conceives not only the areas of test-construction activities but also the norms by which the learner-performance is to be interpreted.

Beyond these four stages is the stage of test validation—a stage where we undertake the evaluation of a test. This activity serves two important purposes:

i) the purpose of immediate concern—ensuring whether the test procedure and the test are dependable in terms of the objectives and measurement;

ii) the purpose of long-term concern—envisaging and directing the reformative efforts that are required to improve the examination procedures (through consistent and systematic efforts over a period of years of feeding the findings of the evaluation of the work in one examination into the planning operations of the subsequent examinations).

There are different criteria to be taken into consideration while evaluating a test or determining the worth or quality of a test. Chief among these are provided by the concepts of 'validity' and 'reliability'. There are a host of other features to consider which pertain, in general, to the question of 'usability'. We shall discuss these in the following sections. But before we take them up for discussion, let us look into some of the general questions concerning evaluation of educational tests.

3.2 PRINCIPLES OF EVALUATION TOOL CONSTRUCTION

A test in order to be good must follow certain principles. Usually, test construction passes through four distinct phases i.e., planning, preparing, trial and evaluation. The test contains a carefully prepared and fixed set of items and procedures for administering and scoring. In other words, we can say that any test tool is first to be carefully planned, secondly items are to be prepared, then the test is to be tried out and lastly it must be evaluated from different angles before it is used.

3.2.1 Planning

It is the first step in test construction that whatever we do we plan it in advance, so that our act will be systematic. Obviously so many questions come to our mind before we prepare a test. What content area is to be covered by the test? What types of items are to be asked and what are the objectives that are going to be tested? Are the objectives that are going to be tested clear? etc. Suppose we want to prepare a good test in General Science for grade X. If most of the questions are asked from Physics and Chemistry and a few from Zoology, will it be a good test? Anybody would offer the criticism that the test is defective because it would fail to measure the achievement of pupils in Geology, Astronomy, Physiology, Hygiene and Botany. So it is desirable that all the content areas be duly represented in the test. Moreover, due Weightage should be given to different contents. Whenever we want to prepare a test, the weight to be given to different content areas must be decided at the beginning.

If we are preparing a test on Physics we must decide the weight to be given to different chapters taught in Physics. Thus, at the planning stage, it is the first task

to decide the weight to be given to different content areas. Usually the weight is decided by taking expert advice. It must be noted that the Weightage must be in conformity with the amount of content taught under each content area. To draw up an objective list of such weight, the average of the available expert opinion may be taken.

Even though we give due weight to content areas, if all the questions (test items) aim at testing only the memory of the pupil, can it be called a good test? No. It must not only cover the area of knowledge (recall or memory comes under knowledge), but also other objectives like understanding, application, skill etc. A good test must aim at measuring all the behavioural areas. But can all the objectives be tested through one test? At this stage expert opinion may be sought as to which objectives are to be covered through our test. Usually in achievement tests four major objectives viz., knowledge, understanding, application and skill are tested. Now another question comes to our mind. What weights are to be given to different objectives chosen? Here again expert opinion can be taken.

Suppose that we are going to construct a test of Mathematics for students of grade IX. After due expert opinion we decide that the weight to be given to the different content areas viz. Arithmetic, Algebra, Mensuration and Geometry are 20%, 30%, 20% and 30% respectively and the weight to be given to the objectives of knowledge, understanding, application and skill are 40%, 30%, 20% and 10% respectively. After the weight to be given to the different content areas and different objectives is decided upon, a blue-print can be prepared. A blue-print for the proposed test on Mathematics is presented below.

Table 3.1: Blue-print (two dimensional chart)

Objectives	Knowledge	Understanding	Application	Skill	Total
Arithmetic	8	6	4	2	20
Algebra	12	9	6	3	30
Mensuration	8	6	4	2	20
Geometry	12	9	6	3	30
Total	40	30	20	10	100

The chart shown above is a content-behaviour chart. It is called the blue-print. Here, we have shown the weight to be given in two dimensions, viz., content and objectives. Thus, it can be called a 'Two-dimensional chart'. The blue print is just a design or a plan of the test to be prepared.

Another thing comes to mind—whether we would ask 'essay type' or short-Answer type or 'objective type questions'? If we decide to include only objective type questions the two dimensional chart will serve the purpose. If we want to include all the three types of questions, here again we have to seek expert opinion to decide the weight to be given to different forms of questions. Now the type of question (or form of questions) will be another dimension and as such a three-dimensional chart may be prepared.

Now another question arises. Should all the questions be easy or difficult or average standard? The usual practice is to include all the three categories. So a decision also has to be taken concerning the distribution of questions of different difficulty level. However, for a normal group the percentage of difficult, average and easy items to be included are 15%, 70% and 15% respectively.

We should also decide whether there would be provision for options or not. Whether we should have overall options (as, 'answer any ten') or internal options

(each question has an alternative)? Provisions for options tend to lower the validity of questions. However, we may use internal options if the two questions are comparable in most respects (i.e., they test the same objective based on the same content, are equally difficult and would require same time to complete). So, whether or not options be introduced is to be decided at the planning stage.

Further we must decide the time within which an average student can answer the test. The test will have to be accordingly planned. Moreover, the total marks of the test is be decided and according to the weight fixed the marks are to be divided. The conditions under which the testing will be done should also be thought of in advance.

If a test is to be successful, a careful planning must precede its construction. From the foregoing discussions we feel that the planning of a test is not so easy. To sum up, planning of a test involves the following:

i) A detailed study of the text books, reference books, journals, test manuals, old questions, other reports, etc. is to be made.

ii) Weight to be given to different content areas is to be decided.

iii) Weight to be given to different objectives is to be decided.

iv) Weight to be given to different forms of questions is to be decided.

v) Whether or not provision for options is to be made.

vi) Weightage given to different categories of difficult level of questions is to be fixed.

vii) Total marks of the test along with the time required for its administration, the conditions of administration, etc. are to be planned in advance.

After all these considerations, a blue print of the test is to be prepared. It would not only give us a picture of the question of the test, but also serve as a guide for the preparation of the test.

3.2.2 Preparation

The second step in test construction is the preparation of the test itself. At this stage we have to prepare:

i) the test items

ii) the directions to test items

iii) the directions for administration

iv) the directions for scoring

v) a question wise analysis chart.

i) **Preparation of the test items**

Items must be prepared in conformity with the blue print. We have to choose appropriate items (test situations) which would test the specified objectives in the specific content area. Construction of test items is not so easy. It is the task of test-specialists and experts. An experienced teacher who is sufficiently trained in test-construction can prepare appropriate test items. There are certain rules and guidelines for construction of test items. Separate guidelines are there for construction of 'essay type', 'short-answer type' and 'objective type' tests. Even for construction of different types of objective-type tests, specific guidelines are prescribed. One must have access to all

these guidelines and also access into the taxonomy of objectives before constructing test items. In general, the test items must be clear, comprehensive and free from ambiguity. They must be aimed at measuring the desired pupil-behaviour. They must fulfill their functions to ensure validity.

After the test items are framed they must be arranged properly and assembled into a test. If different forms of test items are being used, they should preferably be grouped form-wise. Moreover, easy items are to be given a place in the beginning, the difficult items at the end. The test items may be arranged in the order of difficulty. Of course, there are various ways of assembling the questions and we may assemble the questions according to our purpose and convenience of interpretation.

ii) **Preparation of directions to test items**

Appropriate directions to test items should be prepared. The directions must be clear and concise so that the students will understand them easily. The students should know as to whether he/she has to write the response or put a tick against the right response or to mark his/her response in some squares provided on the right side of the question or to mark his/her response on a separate answer sheet etc. Sometimes the directions to test items are so ambiguous that the students cannot follow them and as such he/she responds to the items in a manner which he/she thinks fit at that instant or simply passes on to the next item leaving it unanswered. Due to lack of clarity or directions students will respond differently at different times which would lower the reliability of the test. It is essential that the directions to the test items must be carefully prepared and they must be as clear and simple as possible. If necessary, full guidelines (even demonstration) for responding on item may be given.

iii) **Preparation of directions for administration**

A clear and detailed direction as to how the test is to be administered is to be provided. The conditions under which the test is to be administered, when the test is to be administered (whether in the middle of the session or at the end of the session etc.), within what time limit it is to be administered etc. are to be stated clearly. If the test has separate sections, time limits to cover each section must be mentioned. The materials required (if any) for the test such as graph papers, logarithm tables etc. must be mentioned. The directions must state clearly what precautions the administrator should take at the time of administration. So it is important that appropriate and clear directions for test-administration be prepared.

Preparation of direction for scoring

To facilitate objectivity in scoring, 'scoring keys' are to be provided. Scoring key is a prepared list of answers to a given set of objective-type questions. Suppose there are 10 multiple-choice objective type questions (each having four options, A, B, C, D) in a section of the test, the scoring-key will be as follows:

Section-I Scoring key

Q.N.	1	2	3	4	5	6	7	8	9	10
Key	D	C	C	A	B	D	B	A	C	B

A scoring key is prepared by listing serially the key (or right answer) to each question against each item.

For short answer type questions and essay type questions, marking schemes are to be prepared (i.e., marks allotted to different parts of the answer or to different important points etc. are to be mentioned). Such scoring keys and

marking schemes must be carefully prepared. They serve as guides at the time of scoring the test and they ensure objectivity in scoring.

Moreover, it must be clearly stated as to how scoring is to be done. For example, if a strip-key or a window stencil is used, appropriate directions for using them are to be provided.

In certain cases corrections for guessing is necessary. Under the directions, it must be clearly stated for which type of items such corrections are to bee made. The formula used for 'correction for guessing' is given below:

$S = R - \frac{W}{N-1}$ where S = the corrected score

R = No. of right responses

W = No. of wrong responses

N = Total no. of options

Thus, such specific directions for scoring as are likely to be necessary must be prepared. Of course, these may vary from test to test.

v) **Preparation of a question-wise analysis chart**

A question-wise analysis chart is given here. In this chart every question is analysed. This chart shows the content area (topic) the question covers, the objectives (with specifications) that it intends to measure, its type, the marks allotted to it, expected difficulty level and time taken to answer it. This chart not only analyses the items, but also gives us a picture of coverage of contents, objectives, type of questions and coverage of different difficulty levels etc. Moreover this gives us some idea about the total time to be taken for taking the test. This chart further helps us check whether the test has been prepared as per the blue print or not.

Table 3.2: Question-wise analysis chart

Topic	Question	Objectives with specifications	type	Mark
Solitary Reaper	1. Give a tick(✓) mark against the right answer			
(A)	Q. When the poet saw the girl. She was: a) sitting alone b) passing gently c) listening to the songs of the nightingale d) reaping grains and singing	Instructional objective (*Knowledge*) Specific/ behavioural objective (*recall*)	Multiple choice	01
(B)	Q. "The music in my heart I bore, long after it was heard no more." What was the effect of music on the poet later? (Answer in three sentences only.)	Understanding or *Comprehension* Specific or behavioural objective (*interprets)*	Short answer	03

Note: Preparation of such a chart is a necessity for teacher made tests; but for standardised tests we may or may not do it at the preparation stage. However, such analysis may be necessary at the time of editing the final form of the tests.

3.2.3 Try-out

The questions may be carefully constructed, but there is no guarantee that they will operate in the same manner as planned. So before the final form of the test is prepared it is necessary to have a try out.

For the trial of the items, a preliminary form of the test is generally prepared. This contains more number of items than are actually required for the final form. Usually the number of items included in the trial form should nearly be double of the number of items required for the final form. The lesson is that at the item-analysis stage many items will be discarded. A detailed scoring key of the trial form should therefore be prepared.

i) **Preliminary try-out:** After the test items, directions for response, administration and scoring are prepared it is tried out on a few 'sample' students just to ascertain how it works. At this stage 10 to 15 students of different abilities are selected and the test is administered. The aim of doing is to detect the omissions or mistakes if any, to examine whether the directions to items are actually being followed by students, to examine whether the time allowed is sufficient etc. Although a test is constructed with caution it may have some errors or ambiguity of directions here and there. The preliminary tryout will be to bring these to light. This helps us to modify or revise the items or directions whenever necessary. After due corrections the test is edited.

ii) **Final try-out:** At this stage the test is administered to a representative sample. The sample may not be too large. As it is just a pilot study a sample of 200 to 300 will do. But it must be borne in mind that this sample must be a representative sample of poor, average and brilliant students. The aim of such a tryout is to identify the defects and deficiencies of the test and to provide data for evaluating the test.

The purpose of the tryout can be summed up as follows:

a) to identify the defective or ambiguous items.

b) to discover the weaknesses in the mechanism of test administration.

c) to identify the non-functioning or implausible distracters in case of multiple choice tests.

d) to provide data for determining the discriminating value of items.

e) to determine the number of items to be included in the final form of the test.

f) to determine the time limit for the final form.

At the tryout stage the directions must be strictly followed. Conditions for test administration must be normal. The atmosphere should be calm and quiet. There should be proper seating arrangements, light, ventilation and water arrangements. Proper investigation and supervision must be ensured. A wrong administration of the test will give us wrong data for its evaluation.

iii) **Scoring:** After the try-out form of the test is administered the answer sheets are scored as per the scoring key and scoring directions. 'Corrections for guessing' are also done if required under scoring directions. Now the scores are ready for item analysis and evaluation of the test.

3.2.4 Evaluation

After scoring is complete, the test must be evaluated to examine whether the test items are good and whether the test is reliable and valid. For this purpose we:

i) analyse the items to examine their worth of inclusion in the test (Item-analysis);

ii) determine the validity of the test;

iii) determine the reliability of the test;

iv) assess the usability of the test.

i) **Item analysis**

Item analysis is a procedure by which we analyse the items to judge their suitability or unsuitability for inclusion in the test. As we know, the quality or merit of a test depends upon the individual items which constitute it. So only those items which suit our purpose are to be retained. Item analysis is an integral part of the reliability and validity of a test. The worth of an item is judged from three main angles viz.

a) Difficulty index of the item

b) Discriminating power of the item

c) Its internal consistency with the whole test.

a) ***Item difficulty***

When an item is too easy, all the students would answer it. If it is too hard, nobody would answer it. What is the use of having such items in a test? If all the students get equal scores, the very purpose of the test (i.e. to assess the ability of students) is defeated. So it is clear that too easy and too difficult items are to be totally discarded. It is desirable that items of a medium difficulty level must be included in a test. Item difficulty is calculated by different methods.

Method 1: Item difficulty (I.D.) is calculated by using the formula. ID = X 100 where R = no. of testees answering correctly, and N = Total no. of testees.

If in a test administered to 50 pupils an item is passed by (i.e. correctly marked by) 35 students the I.D = X 100 = 70

Here, we understand that the item is easy.

In essence, if the I.D. value is more, the item is easy and if the I.D. value is less than the item is considered to be difficult.

N.B. : Usually I.D. values in between 16 and 84 (or 15 to 85) are retained.

Method 2: Item variance and the difficulty level.

The proportion of passing an item is an index of difficulty. If 90% of a group pass an item, it is easy and when only 10% pass the item, it is difficult. If 'p' is the % of testees passing an item and 'q' is the % of testees failing in it.

S.D. = $\sqrt{Pq}$ or variance = pq

If p = .50, q is .50 and its variance .25

If p = .60, q is .40 and its variance .24

If p = .90, q is .10 and its variance .09

Items with more variance must be included in the test.

Method 3: Difficulty level of items can also be given in terms of standard deviation of the normal curve. For example, when 84% of pupils pass an item, it means that only 16% face the difficulty and its difficulty index in terms of S.D. of the normal curve will be - 1 . Other examples are given below.

Table 3.3: Item difficulty index

Item passed by	Difficulty index in terms of S.D. of the normal curve
16%	+ 1
84%	- 1
31%	+ .5
69%	-.5 etc.

Note: The table of areas under the *normal curve* may be referred to. Items with difficulty values in between $\pm$ 1 are usually retained

Method 5: The item analysis procedures used to obtain a reliable ranking of learners; indices of item difficulty and item discriminating power include the following:

i) Arrange the answer papers after scoring on the basis of merit (Highest mark at the top and lowest mark at the bottom).

ii) Select the 27% of the answer papers from the top and 27% of the answer papers from the bottom. The top 27% who have secured better marks constitute the higher group (H-group) and the bottom 27% who have secured poor marks constitute the lower group (L-group).

iii) Calculate W_H for each item i.e., for each item to determine the number of persons from the H-group who have wrongly answered an item or who have omitted it.

iv) Calculate W_L i.e., for each item calculates the number of persons in the L-group who have wrongly answered the item or omitted the item.

v) Calculate $W_H + W_L$

vi) I.D. = $\frac{W_H + W_1}{2n} \times 100$ where

n = number of persons in either lower group or higher group (n = 27% of N)

For multiple choice tests (where the options may be three or four) the following formula is used.

I.D. = $\frac{W_H + W_1}{2n} \times \frac{100 x option}{option - 1}$

Usually items in the range 16% to 84% of difficulty level are retained.

We can calculate the desired $W_H + W_L$ values from the following table.

Table 3.4: Calculation of item difficulty levels for multiple-choice questions

Difficulty level	$W_H + W_L$ values / No. of options each item has			
	2	3	4	5
16%	.160n	.213n	.240n	.256n
84%	.840n	1.120n	1.260n	1.344

Suppose in a test consisting of all multiple-choice questions with 4 options and the number of persons in the H-group or the L-group is 120 (i.e. n = 120), what would be the $W_H + W_L$ values at 16%, 84% difficulty level?

Referring to the above table, $W_H + W_L$ value at 16% difficulty level = .240n = .24 x 120 = 28.8 = 29 (nearly).

$W_H + W_L$ value at 84% difficulty level = 1.260n = 1.260 × 120 = 151.2 or 151 (nearly).

Thus all items whose $W_H + W_L$ values are in between 29 and 151 are to be retained.

Any item whose $W_H + W_L$ value is less than 29 or more than 251 is rejected.

b) ***Discriminating index:*** To be considered good, an item must have discriminating power. For example, if an item is too easy or too difficult to all the testees, it can't discriminate between individuals. Logically, it is expected that a majority of students of a better standard and a few students of lower standard will answer an item correctly. Thus, an item must discriminate between persons of the high group and the low group. In other words,

W_L = Number of persons in the *lower group* (i.e. 27% of N) who have wrongly answer an item or omitted it.

W_H = Number of persons in the *higher group* who have wrongly answered an item or omitted it.

It is expected that W_L will be always more than W_H i.e., $W_L - W_H$ will always be positive. If W_H is more than W_L the item is either ambiguous??? and it is to be totally rejected.

We need to calculate the $W_L - W_H$ value for each item. Representative minimum $W_L - W_H$ values for an item with different options for different 'n' (27% of N) have been provided in Table 5.

Table 3.5: Representative minimum Values

Total no. Tested	**No. in Low or High group (.27 N)**	**$W_L - W_H$ at or above which an item can be considered sufficiently discriminating**			
		True or false/ Two options)	3 options	4 options	5 options
350-353	95	13	14	14	14
443-446	120	14	15	16	16
1110-1112	300	22	24	24	25

By referring to the table we can find that for a'n' of 120, the minimum $W_L - W_H$, value for an item with 4 options should be 16. So, all the items whose $W_L - W_H$. value is 16 or above are considered to be sufficiently discriminating. If $W_L - W_H$ value of an item is less than 16, it is to be rejected-

c) ***Internal consistency of items with the whole test***

Statistical methods are used to determine the internal consistency of items. Biserial correlation gives the correlation of an item with its sub-test scores and with total test-scores. This is the process of establishing internal validity. There are also other methods of assessing internal consistency of items and as they are beyond the scope of our present purpose, we have not discussed **them** here.

3.2.5 Finalisation

After **item** analysis, only good items with appropriate difficulty level and with satisfactory discriminating power are retained and these items form the final test. Time required for the test is determined by taking the average time taken by three students who represent three groups: bright, average and below average. Now the test is administered to a large representative sample and the test-papers are scored.

3.3 ITEM ANALYSIS

In Sub-section 3.2.4 of this unit we discussed that an **item** analysis is a procedure by which we analyse the items to judge their suitability or usability for inclusion in the test. In this section we shall talk about the characteristics of a good item (i.e. a question).

The concern about the quality of an item becomes immediate when we attempt to develop a test or an item bank or when we attempt to evaluate a test being put to use. To determine how sound or good an item is we ought to know the features that go into its constitution and the qualities which contribute to its soundness.

3.3.1 Mechanics of an Item

An item or a question is an 'instrument' that we use to measure learning-outcome. One's learning is measurable by another only when it is demonstrated in observable behavioural patterns. An item, intended to be used as an instrument to measure learning, should make a learner 'act' or 'behave' or 'respond' so as to demonstrate his/her mastery (or the extent of mastery) or otherwise with regard to the select 'bit' of learning. 'The stimulus may be in the form of a task. The task may require the learner to do a 'descriptive' and/or a 'practical' activity. Descriptive activities may be oral or graphic (involving linguistic, semiotic and other features of communication). Practical activities may involve the use of some tools and materials and they may be performed in realistic or stimulated conditions.

3.3.2 Required Functional Conditions of an Item

An item or a question is primarily the specification of a task, the response to which is expected to put a desired bit of learning to demonstration. This involves two distinct activities on the part of the item-writer:

i) devising a task to meet the specific objective of the test, and

ii) specifying the task precisely and adequately.

The devising of the task has to be done carefully so as to ascertain that the performance of the given task requires the learner to display the desired quantum of knowledge of a chosen content or the ability to use a skill in a desired way. The specification of task also has to be done with great care. The specification may defeat its purpose:

i) when it is not adequate, and

ii) When is not 'communicated' clearly.

Let us elaborate these conditions further.

The specification will not be adequate if it does not point out the conditions under which the task is to be performed (say, for instance, tools, materials,

guidelines, facts and information, etc. to be provided and the stages at which they are to be provided to the learner). It will not be adequate also when the level of accomplishment to which the task is to be performed is not mentioned in clear terms.

And even when an adequate specification is conceived, its purpose may not be served if is not 'conveyed' or 'presented' to the learner properly. This means that simple language or some graphic signs or gestures or some other mode of communication (sometimes, more than one of these at a time) has to be employed to make the specification clear to the learner. If the mode of communication employed to present the specification is beyond the comprehension of the learner, then the item, however, adequately conceived, may not be of any use.

To summarise, an item, to be an effective instrument of measurement of learning, should meet the following requirements adequately:

i) The task that an item specifies should, in the process of learner-response, demand and reflect only those specific aspects of skills or bits of learning that are being tested.

ii) It should specify precisely:

a) what the learner is to do,

b) the conditions under which it is to be done, and

c) to what level/standard it is to be accomplished.

The medium (linguistic, graphic semiotic, etc.) used to present the task specification should be such that there may not be any gap in its communication to the prospective testee (i.e. the learner should be able to follow the medium without any misunderstanding.

3.3.3 Checking the Functional Conditions of an Item

These general requirements that make an item sound are all concerned with qualitative features. To determine whether these requirements are satisfied by a question or an item, there are no objective measures available. One will have to depend on one's own experience and subjective assessment and the conclusions will be essentially empirical in nature. Collective effort, in this regard, therefore, may result in more reliable conclusions. A group of content experts, in collaboration with an evaluation expert, may do better than an individual. The conclusions may be still more reliable if a comprehensive check list of criteria is prepared in advance for each type of item in every subject area and used while validating a question or an item.

This act of checking whether a question or an item meets the qualitative requirements to make it an effective instrument is known as 'pre-validation in the process of question/item bank development.

Check Your Progress 1

Notes: *a) Space is given below for your answers.*
b) Check your answers with the ones given at the end of this unit.

Two English teachers, say X and Y, working under similar conditions frames the two following questions (I and II respectively for use in their classroom after an hour of instruction. Their intention is to check whether the students are able to use correctly the different forms of given vocabulary items in contexts relevant to them. Study the two questions carefully and say which of the two is better. Give reasons for your judgement. You can consider the

two questions from the points of view of, a) task-objective relationship and b) clarity, precision and adequacy of task-specification.

1) Write one sentence each of your own using the different forms of the following two vocabulary items:

   i) to relieve

   .....

   .....

   .....

   ii) to blast

   .....

   .....

   .....

2) Following are two words from the material you have just read. Change the form of these verbs to fit into the sentences given. In some cases you will have to change them into nouns and adjectives. Fill in the blank spaces with appropriate forms.

i) to relieve

   a) He was (_____) to hear that his wife had not boarded the ill-fated airbus which was hijacked.

   b) Aspirin is supposed to provide (_____) for a headache.

   c) Some people do not like to take pills as pain (_____)

ii) to blast

   a) The (_____) of the dynamite was very loud.

   b) The glass was (_____) out of the window by the explosion of dynamite.

   c) The wind was (_____) all night long because of the hurricane.

3.3.4 Behavioural Characteristics of an Item

However, good the collective effort and however well composed the expert group attempting to determine the qualitative attributes of a question or an item, it can never be predicted with any certainty as to how a question/item will actually 'behave' or perform with a likely group of testees. What an amount of difficulty an item will cause, what sources of difficulties it will develop (or with what amount of ease it will be taken up and what sources of facilitative clues it will develop), what kinds of responses it will generate, are general questions which can never be precisely and adequately answered by any test-constructor or test-evaluator.

Observations on task - objective relevance and adequacy of specifications, though good in themselves, cannot offer much guidance to determine such behaviour traits of an item with a prospective group of testees. Only the testees that take up the question/item can be the arbiters in deciding these traits. They alone can provide these bits of information which we need to determine the eligibility of a question or an item to build up a sound achievement test. An achievement test, as we have noted earlier, intends to

provide a basis for selection and grading of learners- As such, questions/ items making up an achievement test are expected

i) to be neither too difficult not too easy for the prospective testees, and

ii) to discriminate effectively 'the more able' (among the testees) from 'the less able'.

We may also need information about the behaviour characteristics of an item when we construct tests of special specifications - like tests of same **difficulty** level ('parallel tests') and tests of progressive difficulty levels ('graded tests').

3.3.5 Measuring Behavioural Characteristics

The procedure used to get information about the behaviours/characteristics of a question or an item is known as item-analysis. The information is obtained in quantitative terms and they are given in the form of numeral indices, unlike the information about the qualitative features of an item which are given in the form of empirical statements. Two indices are given after analysing the responses to a trial test paper administered to a representative group of the testees:

i) the measure of difficulty of each question, and

ii) a measure of the extent to which each question discriminates between the high scores and the low scorers on the same test.

The first measure is known as the Facility Value (FV) and the second the Discrimination Index (DI).

Different ways of calculating the two measures are available and some of them are sophisticated statistical operations which can be done only with the help of computers. The procedures given below for calculating the two measures may not give very accurate indices, but they do give satisfactory ones. The operations involved are simple and can be carried out even by those who do not have any acquaintance with statistics as a discipline.

Tabulation of scores: To facilitate the computations of facility and discrimination indices of the item comprising a test, we need to tabulate the scores properly. An illustration of score tabulation is given below (Table 2.6). The illustration may help you follow the steps given as under.

Step 1: Arrange the response sheets in order of value of the total score (Col. IV) in the test. Put the highest score on top and the lowest score at the bottom. (Response sheets of the same total score may be put one below the other.) Number the response sheets serially (Col. If).

Step 2: Divide the response sheets into three ability-groups (Col. I).

Higher Ability Group (HAG),

Middle Ability Group (MAG), and

Lower Ability Group (LAG).

If the strength of the sample (number of testees) is less than 40, the top 50% of the response sheets can be taken up to form HAG and the rest to form LAG. If the sample is between 40 and 100, then 27% or 10% respectively of the top and of the bottom scripts may be taken to form HAG and LAG. (If the percentage works out to a fraction, the fractions may be ignored and either of the groups may be allowed to be larger by one or two numbers).

FV for item 1

$$= \frac{\text{Total Score of the sample}}{\text{Total No. of candidates}}$$

$$= \frac{9+63+2}{11+88+11} \times 100$$

$$= \frac{74}{110} x100 = 67.27\%$$

FV for item 4

$$= \frac{\text{Total Score of the sample}}{\text{Total No. of candidates}}$$

$$= \frac{9+72=5}{11+88+11} \times 100$$

$$= \frac{86}{110} \times 100 = 78.18\%$$

DI for item 1

= Facility in respect of HAG

- Facility in respect of LAG

$$= \frac{9}{11} - \frac{2}{11}$$

$$= \frac{9-2}{11}$$

$$= \frac{7}{11} = 0.636$$

DI for item 4

= Facility in respect of HAG

- Facility in respect of LAG

$$= \frac{9}{11} - \frac{5}{11}$$

$$= \frac{9-5}{11}$$

$$= \frac{4}{11} = 0.363$$

Table 3.6: Illustration: tabulation of scores to facilitate FV and DI computations

Ability Group	Sl. No.	Roll No. in order of ranking	Total individual score	Item-wise score V							
I	II	III	IV	1	2	3	4	5	6	7	8
H	1	891	25	√	√	√	√	√	√	√	
I	2	702	24	√	√	x	√	√	√	√	
G	3	801	23	x	√	√	√	√	√	√	
H	4	705	23	√	√	√	√	√	√	√	
E	5	712	23	√	√	√	√	x	√	x	
R	6	813	22	√	√	x	√	√	x	x	
	7	811	22	x	√	√	x	√	√	√	
	8	737	21	√	x	√	√	x	x	x	
	9	785	21	√	√	√	x	x	x	√	
	10	721	20	√	x	√	√	√	√	x	
	11	850	20	√	√	x	√	x	x	x	
Total		11 Candidates		9	9	8	9	7	7	6	
M	12										
I	.										
D	.										
D	.										
L	.										
E	99										
Total		88 Candidates		63	66	49	72	47	15	41	
L	100	786	8	x	x	x	√	x	x	√	
O	101	809	8	x	x	√	√	x	x	x	
W	102	722	7	x	√	x	x	x	√	√	
E	103	826	7	√	x	√	x	√	x	√	
R	104	813	7	x	√	x	√	x	√	√	
	105	851	7	x	x	x	√	x	x	√	
	106	870	6	x	√	x	x	x	x	x	
	107	764	6	√	x	x	x	x	x	√	
	108	783	6	x	x	x	√	x	x	x	
	109	822	5	x	x	x	x	x	x	x	
	110	847	5	x	x	x	x	x	x	√	
Total		11 Candidates		2	3	2	5	1	2	7	

Step 3: Draw a table of vertical columns and horizontal rows. Enter the serial number of the Response Sheets which you put while carrying out Step 1, one below the other in the first column. (The corresponding Roll No. of candidates can be given in the second column, if necessary.) Leave a gap of three rows each below the HAG, the MAG and the LAG.

Step 4: Enter item-wise scores in the horizontal row against each candidate (Col. V). When the item-wise scores of all candidates are entered, the total score of the sample on each item could be calculated by adding up scores in the vertical columns and the total score of each individual on the test could be calculated by adding scores along the horizontal row.

Determining the facility value: Facility value is generally presented **as a percentage.** In the case of an objective type item, it is calculated as the number of learners answering the item correctly divided by the number of learners attempting it. The fraction is multiplied by 100 to get the figure in percentage. In the case of a supply-type question, facility value is the average mark obtained by the sample on the question divided by the maximum mark allotted for the question. Here too the fraction is converted into a percentage figure.

To summarise, FV of an objective item =

$$\frac{\text{No. of learners answering the item correctly}}{\text{No. of learners taking the test}} \times 100$$

FV of a free-response question =

$$\frac{\text{Average score obtained by the sample of the question}}{\text{Max. mark allotted for the question}} \times 100$$

The facility value ranges from 0% to 100%. FV represents the fact that none of the sample has answered the item correctly and hence the item has no 'facility' whatsoever for the given sample. 100% FV represents the fact that everyone in the sample has answered the item satisfactorily and the item has no difficulty whatsoever with the given sample.

Determining the discrimination index: Discrimination index of an item is arrived at by deducting the facility value of the LAG **(**in the item from the facility value of the HAG on the same item, The DI is always presented in the form of decimal fraction and it may range from - 1.0 to + 1. .

In the case of an objective type item:

DI = FV of the item with HAG - FV of the item with LAG

$$= \frac{\text{No. of testees answering the item correctly in HAG}}{\text{No. of testees in HAG}}$$

$$= \frac{\text{No. of testees answering the item correctly in LAG}}{\text{No. of testees in LAG}}$$

(We should mention here that multiplication by 100 which is normally required **to** present FV's in percentage is avoided here, because Dl's are to be given in decimal fraction, not in percentage.)

To calculate DI in the case of a free-response item, you have to find out the mean score (MS) i.e., average of the scores on the item in the HAG (Let us represent this as MS-HAG) and the mean score on the same item in the LAG (Let us represent this as MS-LAG). The difference between the two mean scores (obtain by deducting the LAG mean score from the HAG mean score) divided by the maximum marks allotted for the item gives the D1 of the item.

$$\text{DI of an item} = \frac{(\text{MS-HAG}) - (\text{MS} + \text{LAG})}{\text{Max. score fot the item}}$$

3.3.6 Interpreting Behavioural Characteristics

FV and DI can be helpful to us in determining:

- the quality of an item at the tryout stage in the development of a test or a question/item bank; and
- the quality of teaching/learning at the stage of actual **use** of a test.

What different values of facility and discrimination can signify with regard to the quality of an **item** in the context of item tryout are tabulated in tables 7 and 8:

Table 3.7: FV in the context of item try-out: an interpretation

FV ranging			
From	**to**	**would mean that ….**	**would require……**
0%	25%	the item **is** too hard	modification of the item (probably checking of dstracters)
25%	75%	the item is within the suitable range of facility	retention of the item
75%	100%	the item is too easy	rejection of the item (perhaps, checking of clues sometimes can help improve items and retain).

Table 3.8: DI in the context of item try-out: an interpretation

DI ranging			
from	**to**	**may mean that ...**	**may require**
- 1,00	+ 0.20	the upper sample is not doing better than the lower	suitable modification or rejection of item
+ 0.20	+ 1.00	the sample item makes satisfactory discrimination	the item is to be retained.

How different values of FV and DI of objective type items can be interpreted in relation to the quality of teaching/learning is presented in tables 9 and 10:

Table 3.9: Interpretation of FV in the context of actual test-use

FV ranging		
from	**to**	**would mean that**
0%	25%	the topic has not been taught/learnt well and that the teaching techniques are to be reviewed to ascertain what has gone wrong.
25%	75%	the topic has been taught/learnt reasonably well.
75%	100%	the learners have gained an exceptionally good knowledge of the topic.

Table 3.10: Interpretation of DI in the context of actual test-use

DI ranging		
from	**to**	**would mean that**
-1.00	- 0.25	the weaker students have a better grasp of the topic than the 'good' students.
- 0.25	+ 0.25	all the students have an equal grasp of the topic.
+ 0.25	+ 1.00	there is too great a gap between the lower ability and the higher ability groups (perhaps, some remedial work for lower learners must be planned).

FV and DI of subjective items: Interpretation of FV and DI in the case of subjective items is slightly complicated. 'Md FV in the case of an essay-type or short answer type question need not be the index of the facility of the question alone. Since subjective assessment is involved in scoring and since it is the scores that we take as the data, the FV in such cases may not be reliable. In the absence of precise guidelines for scoring, the FV in these cases may reflect the lenient/severe attitude of the scorer, or perhaps the scorer's attitude and the item's facility (if the scorer's attitude remains constant).

The interpretation of DI is also difficult in the case of free-response tests. In such cases the number of questions that make up the test are fewer than the number of questions that make up an objective type test. Consequently the weight that each question carries in a subjective test is relatively high and each question contributes in considerably large chunks to make up the total score. As a result the correlation between the score on a question and the score on the test turns out to be undependable. A DI in such a case is likely to be spuriously high. The 'satisfactory' level of DI in such cases, therefore, has to be higher. It should be, say, more than + 0.50.

Check Your Progress 2

Notes: *a) You can work out your answer in the space given below.*
b) Compare your answer with those given at the end of this unit.

i) Distinguish between 'facility value' and 'discrimination index' of an item.

.....

.....

.....

ii) What should be the range of 'facility value' and 'discrimination index' for an item to be retained in a test?

.....

.....

.....

3.3.7 Use of Behavioural Indices

The facility and discrimination indices will be helpful to both test constructors and teachers. To test constructors they will help identify faulty items which they can modify and use, or reject. They will also help them in developing

effective multiple choice items by providing feedback on the working of the 'key' and the 'distracters'. They could also help test constructors to develop test tools designed in order of progressive difficulty with items of weak discrimination followed by items of gradually increasing discrimination in favour of better students. The facility value obtained from different groups of learners taking the same test can provide teachers with a basis for comparison and can also help in defining and maintaining 'standards'. The discrimination indices locate topics to be addressed to all the learners and topics to which learners of lower ability are to be restricted.

3.4 GUIDELINES FOR THE USE OF AN EVALUATION TOOL

Most learning is a complex mix of physical and psychological activities. The proportion of the mix varies from one area of learning to another. But, invariably, in all areas of learning, the involved psychological processes make precise assessment of accurate learning. For the same reason, the evaluation of an educational test cannot be very accurate and will mostly be a subjective assessment.

3.4.1 Quality of a test: Some Focal Points

If we are to evaluate an educational test, with all its inherent complexities disallowing total precision and objectivity, how are we to go about it? Though the assessment is subjective, it can be regulated and guided to a certain extent by a careful consideration of the following features of a test-content. (Table3.11).

Table 3.11: Characteristics expected of a test-content

Feature of test content	Aspects to check
Presentation of item numbering sequencing, wording, punctuation	Clarity in expression and specificity in task requirement
Presentation of general and specific instructions—wording, punctuation, etc.	Clarity in expression and precision in the description of requirements
Sampling of items	Appropriacy in terms of coverage and allocation of due weight
Choice of item-format	Suitability to the chosen test-point
Choice of the form of measurement (written test, oral test, field work, laboratory work	Suitability to the general objective of the test
Scheme of evaluation (given to examiners)	Maintenance of concurrence with the set objective of the items and the test

These features are the focal operational points with regard to which even any little amount of slackening of care might impair an educational test meeting its stated or intended objective. The chief attributes of a good test—validity, reliability and usability—should be verified with these features to ascertain the quality of a given test. We shall now take up each one of these attributes and discuss them in greater detail.

3.4.2 Validity of Tests

What is validity?

The concept of the validity of a test is primarily a concern for the 'basic honesty' of the test 'honesty' in the sense of 'doing' what an item promises to do. It is a concern for the relationship , on the one hand, between the purpose to be achieved and on the other hand, between the efforts taken, the means employed and what those efforts and means actually achieve. There is always a gap, whatever be the size, between the purpose of a test and the extent of realisation of the purpose in practice. Hence absolute validity is ideal in educational testing. Perfection in terms of validity—a perfect match between the purpose and practice—is hard to achieve. This is due to:

i) the nature of 'learning', which is the subject of measurement, and

ii) the nature of each of a number of factors that become involved in the measurement learning.

(These factors, we have discussed briefly in Section 3.4 above).

The less variant that a test turns out in practice from its stated purpose, the more valid it is. Hence validity is a measure of the degree of success with which a test accomplishes what it sets out to accomplish. It is an attempt to answer how close a test is in its operation to the purpose in its plan-design. To be precise, a test is valid to the extent to which it measures what it purports to measure.

Types of validity:

`Purpose' and 'practice' then, are the two dimensions of considerations involved in the concept of validity. 'Practice' is conditioned mainly by three operant forces—the test, the testee and the examiner. If the demands made by the test, the performance offered by the testee and the valuation (of the testees' performance) done by the examiner are all directed to be symmetrical with the given/set purpose of the test, then validity is ensured. Thus the test purpose is the constant point of reference for validity. Consequently, several types of validity are conceived of to suit the specific purposes that tests are designed to serve. We shall take up four types of them for our discussion. They are:

i) Content validity

ii) Criterion-related validity

   a) Concurrent validity, and

   b) Predictive validity

iii) Construct validity

iv) Face validity

Content validity

Content validity is the most important criterion for the usefulness of a test, especially of an achievement test. It is a measure of the match between the content of a test and the content of the 'teaching' that preceded it. The measure is represented subjectively after a careful process of inspection comparing the content of the test with the objective of the course of instruction,

The key aspect in content validity is that of **sampling.** Every achievement test has a content area and an ability range specified for its operation. Given the limited human endurance in taking a test (say three hours at a stretch) and hence a limited test-duration, no single test can ever make **a** total representation of any considerable length of a content area. A test, therefore, is always a sample of many questions that can be asked. It is a concern of content-validity to determine whether the sample is representative of the larger universe it is supposed to represent.

A table of specifications with a careful allocation of weight to different units of the content area and the several abilities, keeping in view the set objectives of the course and the relative significance of each of these, can help a test-constructor as a road-map in the construction of items and the development of a test. A careful scrutiny of the table of specification (if any has been used) and the loyalty with which it has been adhered to while developing the test can help you assess the adequacy and appropriateness of sampling. Where such a table of satisfaction is not available, you may have to develop a 'concept-mapping' of the content area to check the sampling of content represented by a test.

We should note here that a test that is content valid for one purpose may be completely inappropriate for another. We should also note that the assessment of content-validity has to be subjective basically as it depends on the assessor's estimate of the degree of correspondence between what is taught (or what should be taught) and what is tested it requires a careful examination of the stated objectives of the course in terms of course content and target abilities and a study of the size and depth of realisation of their coverage. Such examinations lead to 'estimates' and not to 'measurements'. That is; the observations of such examinations tend to be subjective statements. They cannot be expressed in terms of objective numerical indices.

Check Your Progress 3

Notes: *a) You can work out your answer in the space given below.*
b) Check your answer with the one given at the end of this unit.

We have said that an achievement test of high content validity cannot be a content valid test for diagnostic purposes. Why?

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Criterion-related validity

Unlike content validity, criterion-related validity can be objectively measured and declared in terms of numerical indices. The concept of criterion-related

validity focuses on a set 'external' criterion as its yardstick of measurement. The 'external' criterion may be a data of 'concurrent' information or of a future performance.

The 'concurrent' criterion is provided by a data-base of learner-performance obtained on a test, whose validity has been pre-established. 'Concurrent' here implies the following characteristics;

i) the two tests—the one whose validity is being examined and the one with proven validity (which is taken as the criterion)—are supposed to cover the same content area at a given level and the same objectives;

ii) the population for both the tests remains the same and the two tests are administered in an apparently similar environment; and

iii) the performance data on both the tests are obtainable almost simultaneously (which is not possible in the case of 'predictive' criterion).

The 'predictive' criterion is provided by the performance-data of the group obtained on a course/career subsequent to the test which is administered to the group and whose validity is under scrutiny.

The validity of a given test is established when 'concurrent' criterion correlates highly (i.e. agrees closely) with its own data.

Validity established by correlation with 'concurrent criterion' yields **concurrent validity** and similarly validity established against the scale of 'predictive' criterion is called **predictive validity**. The former resolves the validity of tests serving the purpose of measuring proficiency; the latter resolves the validity of tests meant for predictive function. The 'concurrent' criterion has been widely used in the validation of psychological tests, especially tests of intelligence. The general practice is that one or two standardised tests of intelligence with proven quality are used to validate the item. It is crucial in all selection and placement tests. For example, when a Banking Recruitment Board selects candidates for the post of clerks on the basis of a clerical aptitude-cum-intelligence test, the selection will be purposeful only if a high correlation is established between the test results of the candidates and their performance ability, subsequently, in the clerical position. The higher the predictive validity, the more emphatic this assertion will be. In all cases of criterion-related validity, an index of the degree of correspondence between the tests being examined can be obtained. This index of agreement is known as correlation coefficient in the statistical parlance.

Construct validity

The word 'construct' means the ideas developed in one's mind to define, identify or explain objects/phenomena. Let us suppose that a person is interested in the study of intelligence. He/she hypothesises that the third or fourth generation learners will have a higher IQ than the first generation learners. On the basis of his/her observations he/she may build a theory specifying the degree of difference in the IQ of the two groups of learners.

If a test is constructed, then, to measure the difference in the levels of intelligence of first generation learners and third/fourth generation learners, the test would be considered to have construct validity to the extent that its scores correspond to judgements made from the observations derived by the scorer about the intelligence of the two groups of learners. If the expected level of difference is not established by the test scores, then the construct validity of the assumption that the test measures the difference in the levels of intelligence is not supported. Thus, a test will be described to have construct

validity if its scores vary in ways suggested by the theory underlying the construct. In other words construct validity is the degree to which one can infer certain constructs in a psychological theory from the test score.

Construct validity is an important concept to those who are engaged in theoretical research on various constructs.

Face validity

Before we take up a detailed scrutiny of a test for any of the above validity-types, we generally tend to make an impressionistic assessment so as to develop some propositions which may guide our approach to the assessment of validity. Such propositions are developed on a facial understanding of the extent to which a test looks like a valid test or the extent to which the test seems logically related to what is being tested. These propositions constitute what is known as face validity.

Face validity may not be dependable. A test may look right without being rational or even useful. For instance, a terminal examination in a course of 10 units may appear to have reasonable face validity until you come to realise that it contains questions on the first five units only and therefore lacks content validity. Sometimes there may be situations where a test may appear to have low face validity, but in practice it may turn out to be a sound one. In such cases, the testees too may not know what is being tested and ipso facto the measure may provide far more effective assessment. For instance, the ability to react quickly to a flash of light may be a good test of potential as a football player. Such a test of reaction time may have content validity even if it doesn't have much face validity.

(Incidentally the 'idea' that prompts using this reaction time test for identifying a potential football player is a construct. The idea or the construct perhaps may be explained as that one who is able to react speedily and correctly to the sudden darts of an object can be a potential football player).

Check Your Progress 4

Notes: *a) Write your answer in the space given below.*
b) Check your answer with the one given at the end of this unit.

We talked about four types of validity-content validity, criterion-related validity, and face validity. Given below are three situations. Identify which relates to what type of validity.

i) A test given at the end of a semester of a degree course to measure, how much the learners have achieved of the given course-

.....

ii) An entrance test to an engineering course to select suitable candidates.

.....

iii) A test designed to be parallel in structure and content to another test of proven worth.

.....

3.4.3 Reliability of Tests

The discussion of the concept of validity relates to the question of what to test. The concept of reliability which we shall discuss in this section relates to the question of accuracy' with which the 'what' is measured.

If a student were to take the same test twice, the logical expectation would be for the student to get more or less the same score both the times. But this does not happen practically on most occasions, Differences in scores do occur and they are likely with every repetition of the test.

The difference may be due to many reasons:

i) the characteristic (say, intelligence) which is being measured may change across time (we can call this 'trait instability');

ii) the particular questions that a test-constructor chooses to be representative of the quantum of knowledge which the test has to deal with may affect the score (we can call this 'sampling inconsistency);

iii) any variation in the communication of instruction or in test-timing or in the rapport with the administrator could lead to score variability. Let us call this 'administrator inconsistency');

iv) lack of objectivity in terms of one or more of the following could also affect the score:

   a) the item (it may not have the same significance to all the students or to a student, every time it is repeated),

   b) the response which the item permits (it may not be limited in number and may permit various levels of adequacy), and

   c) the scoring method used (it may leave room for free play or subjective judgement of the scores),

We can see (a) and (h) as constituting 'item inconsistency' and (c) as 'scoring inconsistency' and

v) personal characteristics of an individual like fluctuations in memory, effort, attention, fatigue, emotional strain and similar factors could cause score variability (Let **us** call this 'human inconsistency'),

The concept of error score and the concept of reliability: The sources of variation that we have talked about are called 'sources of error' and the variation in a person's score is called 'error variance'. The assumption here is that all scores given to learners are affected positively or negatively by one or more of the factors mentioned above. The 'true score' reflecting a learner's ability is hypothetical and is never determinable as the sources of error are not totally extricable from any measurement of learning. But the 'true score' is supposed to be constant (assuming the 'learning' that it reflects is stable). The 'error score' is not constant as the error-sources themselves are variable.

Thus, the 'true score' being supposed to be constant and the 'error score' inconstant, the combination of the two which we get in the form of scores on test-papers (i.e. 'observed score') varies in proportion to the difference in the 'error score'. That is, whatever variation we find in the 'observed score' of a learner on different administrations of the same test, are due to errors in assessment brought about by the variables operating on different occasions—the 'true score' of the learner remaining the same all through.

You should note here that among errors two types are possible: 'systematic error' and random error'. When a weight bridge consistently weights 50 kg. less of a lorry-load repeatedly, the error is constant and we call it systematic. We refer to those errors as random errors which do not remain the same on every occasion of measurement.

Some points of further clarifications: Before we proceed further to the methods of determining reliability it is worth observing here some points of further clarification:

i) Reliability refers to the results obtained with an evaluation instrument and not to the instrument itself. An instrument may have a number of different reliabilities depending on the groups of subjects and situations of use. Hence it is more appropriate to speak of the reliability of 'test scores' or of 'the measurement' than of 'the test' or of 'the instrument'.

ii) Test scores are not reliable in general. An estimate of reliability always refers to a particular type of consistency—say, consistency of scores over a period of time ('stability') or consistency of scores over different samples of question ('equivalence') or consistency of scores across scoring on different occasions ('scorer reliability') and the like. The scores of a given test may be consistent in one of the above respects and not in another. The appropriate types of consistency in a given case are dictated by the use to be made of the results.

iii) Reliability is a necessary but not a sufficient condition for validity. While low reliability can restrict the degree of validity that is obtained, high reliability provides no assurance for a satisfactory degree of validity. (A balance that always weighs 10 gms. in excess of the true weight of an object may be highly consistent in recording the same weight for the same object every time it is weighed. Nevertheless measurement suffers in respect of validity in as much it fails to give the real weight of the object).

iv) Reliability is primarily statistical in nature; it may be expressed in terms of shifts in relative standing (in respect of scores) of persons in the group or in terms of the amount of variation to be expected in a specific individual's score. In the former case it is reported by means of a correlation coefficient called a 'reliability efficient' and in the latter case it is reported by means of the 'standard error of measurement'.

Check Your Progress 5

Notes : *a) Indicate your answers in the boxes given against each item*
b) Check your answers with those given at the end of this unit.

i) From the sources given below identify those that are responsible for measurement.

a) the purpose of a test may not the same for all situations ☐

b) the scoring standards of test-responses may vary from person to person ☐

c) the representation of content may not be the same for any two tests on the same subject area ☐

d) the same group of testees are not always available for two consecutive test-administrations ☐

e) the situations of administration may not be the same for any two tests

f) an item may not give the same idea (about the task to be performed) to different testees ☐

g) a testee may not be equally alert and active on two occasions.

ii) What do you call the score given by an examiner on the cover-page of an answer book?

a) True score ☐

b) Observed score ☐

c) Error score ☐

d) Random score ☐

Estimates of reliability

The methods used to measure reliability differ according to the source of error under consideration. The most common approaches to estimates of reliability are:

i) Measures of stability,

ii) Measures of equivalence,

iii) Measures of stability and equivalence,

iv) Measures of internal consistency, and

v) Scorer reliability.

Measures of stability: Measures of stability are known as 'test-retest estimates of reliability'. They are obtained by administering a test twice to the same group with a considerable time-interval between the two administrations and correlating the two sets of scores thus obtained.

In this type of estimate we do not get specific information as to which of the sources of error contribute(s) to the variance in the score. It gives only the measure of consistency, over a stretch of time of a person's performances on the test.

The estimate of reliability in the case will vary according to the length of time-interval allowed between the two administrations. The intervening period can be relatively long, if the test is designed to measure relative stable traits and the testees are not subject enduring the period between the two tests administrations to experiences which tend to affect the characteristic being measured. The intervening time should be shorter when the conditions are not satisfied. But it should not be so short as to allow 'memory' or practice effects' to inflate the relationship between the two performances.

Measure of equivalence: In contrast to the test-retest estimate of reliability which measures change in performance from one time to another the estimate of reliability with equivalent forms of tests measures changes due to the specificity of knowledge within a domain. Instead of repeating the same test twice with an intervening time-gap, the latter procedure administers two forms ('parallel' in terms of content and difficulty) of a test to the same group on the same day (i.e. with negligible time-gap) and correlates the two sets of scores obtained thereon.

The two methods of estimating reliability are quite different and can yield different results. The choice between the two depends on the purpose for which you administer the test. If your purpose is long-term prediction about the reliability of the test, you can choose the procedure of retest reliability estimation. If your purpose, on the other hand, is to infer one's knowledge in a subject matter area, you will have to depend on equivalent forms of estimate of reliability.

Measures of stability and equivalence: When one is concerned with both long-range prediction and inferences to the domain of knowledge, one should obtain measures of both equivalence and stability. This could be done by administering two similar (parallel) forms of a test with considerable time-gap between the two administrations. The correlation between the two sets of scores thus obtained by the same group of individuals will give the coefficient of stability and equivalence. The estimate of reliability thus obtained will be generally lower than the one obtained in either of the two other procedures.

Measures of internal consistency: The three methods discussed above are concerned with consistency between two sets of scores obtained on two different test administrations. The methods that we are to discuss, hereafter collectively called 'measures of internal consistency, arrive at reliability estimate taking into consideration the scores obtained on a single test-administration. 'The estimate of reliability obtained through these methods is mostly indices of homogeneity of items in the test, or of the extent of overlap between the responses to an item and the total test score. The three types of measures of internal consistency are discussed below.

***Split-half estimates*:** Theoretically the split-half method of estimating reliability is the same as the equivalent forms methods. Yet the split-half method requires only one test administration but while scoring the items, a sub score for each of the two halves of the test is obtained and the two sub scores are correlated to get the reliability estimate of half the length of the test. To estimate the reliability of the scores on the full length test, the following formula is used:

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

The application of this formula assumes that the variances of the two halves are equal. That is to say that the items in one half are supposed to match in respect of content and difficulty with the corresponding items in the other. The question then is how the tests can be split into two halves. Different methods are followed, but ordinarily it is done by a preconceived plan (say, assigning the odd numbered items to one half and the even numbered items to the other) without obvious statistical measures to make them equivalent.

Kuder-Richardson estimates: This method of estimating the reliability of test scores from a single administration of a single form of a test by means of formulae KR 20 and KR 21 was developed by Kuder and Richardson. With the help of these two formulae we can estimate whether the items in the test are homogeneous, that is, whether each test item measures the same quality or characteristics as every other, In other words, these formulae provide a measure of *internal consistency* but do not require splitting the test in half for scoring purposes.

The formulae are:

$$\text{KR 20} = \frac{n}{n-1}\left(1-\frac{\sum pq}{\sigma t^2}\right)$$

Where n = number of items in the test,

$\sigma\tau$ = standard deviation of the test scores,

P = Proportion of the group answering item *correctly,*

q = 1–P = proportion of the group answering a test item *incorrectly.*

We use KR 20, we have to

1) Compute the standard deviation of that rest (i.e., σt)

2) Compute p and q for each item,

3) Multiply P and q to obtain the value of pq for each item,

4) Add the value of all the items to get $\sum pq$

5) Using the formula KR 20 we can calculate the reliability of a test.

The computation of KR 20 is more extensive but accurate. A less accurate but simpler formula to compute the reliability of a test was proposed by Kuder and Richardson Known as KR 21.

$$KR21 = \frac{n\sigma t^2 - (Mn - M)}{\sigma t2}$$

Where σt^2 = standard deviation of the test scores

n = number of test items in the test

and M = the mean of the test scores.

Crombach alpha: Kuder-Richardson estimates are possible when the scoring of items is dichotomous. When the scoring is not dichotomous as in a test consisting of essay questions, the formula developed by Cronbach can be used to get the reliability estimate. This formula known as Cronbach alpha is the same as KR 20 except for the fact that is replaced by, where is the variance of a single item. The formula is:

$$\alpha = \frac{n}{n-1}\left[1 - \frac{\sigma si^2}{\sigma st^2}\right]$$

Where n = number of items in a test

σsi^2 = variance of a single item

σst^2 = variance of total number of items in a test

[*'Variance'* is a statistical term. It measures how much the individual scores of a group of learners vary from the average score of the group. It is the mean of the squared deviation of the scores from their mean. Please refer to block 3 of ES-315 for more detail about variance]

The measure actually compares the variance for any single item with the variance for the entire test. It is, therefore, suggested that there should be at least five questions in the test to make this measure meaningful. Having given you some idea about the three types of measures of internal consistency, we shall now talk about yet another variety of reliability.

Inter-scorer-reliability: The question of estimating inter-scorer reliability does not arise in the context of objective tests. But where the scoring is subjective, it is necessary to determine the likely error component in scores which may be there due to the person(s) that scored the performance, and establish objectivity in evaluation. In determining inter-scorer-reliability, the same procedure of correlating two or more steps of scores as done in test-retest or equivalent forms estimate is followed.

This would give the reliability estimate of a single reader (i.e. the scorer who reads through the response). If we want to know the reliability of the sum or average of scores of two or more readers we could use the Spearman -Brown prophecy formula.

$$r_{xx} = \frac{nr}{1 + (n-1)r}$$

Where r_{xx} = reliability coefficient of a test

r = reliability estimate of a single reader

n = number of readers

Check Your Progress 6

Notes: *a) Space is given below for your answer.*
b) Check your answer with the one given at the end of this unit.

Write down the three types of measures of internal consistency and mention in what situation each of these are used.

.....

.....

.....

.....

.....

.....

.....

.....

Let us now sum-up the procedures of the different methods of estimating reliability in a tabular form.

Table 3.12: Different methods of estimating reliability

Methods of estimating reliability	Types of reliability measure	Procedure
1) Test-retest method	Measure of stability	Give the same test twice to same group with a considerable time-gap between the two administrations.
2) Equivalent forms method	Measure of equivalence	Given two forms of a test of the same group in succession.
3) Retest method using equivalence forms	Measure of stability and equivalence	Given two forms of a test to the same group with between the two.
4) Split-half method	Measure of internal consistency	Administer a test once. Get sub-scores for items of two equivalent halves of the test. Use Spearman-Brown formula to obtain reliability estimate for the whole test.
5) Kuder-Richardson method	Measure of internal consistency	Administer a test (of objective type) once. Score the test and apply Kuder-Richardson formula.
6) Cronbach alpha method	Measure of internal consistency	Administer a test (of subjective type) once. Score the test and apply Cronbach alpha formula.

7) Multiple scorer method	Measure of scorer reliability	Administer a test once. Let it be scored by two or more scorers. Correlate the sets of scores to measure the reliability of the scores of one scorer. Apply Spearman-Brown Prohecy formula to obtain the reliability of the sum (or average) of the scores of two or more scores.

Comparison of methods

As noted earlier each type of reliability measure represents different source(s). A summary of this information is given in Table 9. Note that more sources of error are represented by measures of equivalence and stability than by any other type of measure. Naturally, reliability estimates obtained on measures of equivalence and stability are likely to be lower. This should caution you to take into account the type of measures used to report reliability estimate, especially when you attempt to choose a test from among standardised tests guided by reliability estimates.

The 'X' mark in Table 13 indicates the sources of error represented by the reliability measures.

Table 13: Representation of sources of error

	Types of Reliability Measures			
Sources of error	**Stability Scorer reliability**	**Equivalence**	**Equivalence & stability**	**Internal consistency**
Trait instability	X		X	
Sampling error		X	X	X
Administrator error	X	X		X
Random error within the test	X	X	X	X	X
Scoring error					X

3.4.4 Usability

We have discussed in detail the two chief criteria of test-validation—validity and reliability. What remains to be seen are the considerations of 'Usability'. 'Usability' mostly raises questions of feasibility with regard to test-construction, administration, evaluation, interpretation and pedagogical application.

While judging feasibility we should remember that the tests are usually administered and interpreted by teachers without the desirable amount of training in the procedures of measurement. Time available for testing and the cost of testing also deserve attention.

Besides these, attributes like the case of administration, which has little possibilities for error in giving directions, timing, etc., case and economy of scoring without sacrificing accuracy, ease of interpretation and application so as to contribute to intellectual educational decisions, are factors pertinent to the usability of tests.

3.5 LET US SUM UP

- We began this unit with a discussion of the different stages of test tool construction. There are five principles for constructing a tool of evaluation.
- We followed it with a discussion on item analysis that makes it complex to measure an educational achievement and to judge the soundness of such a measurement. Then to facilitate the judgement of an educational measurement, we identified some features of the item analysis and the important qualities to verify each one of them.
- We focused upon the two characteristics of a good tool namely, validity and reliability. We discussed the different approaches within each, the contexts in which each of these approaches becomes relevant and the procedures by which the evidence of validity and reliability of a test is to be established. Finally we referred to the aspects of a test that you should check to ascertain its usability.

3.6 ANSWERS TO CHECK YOUR PROGRESS

Check Your Progress 1

Y', question (i.e., II) is better than X's Reasons

a) Task-objective relationship

Question 1 of X demands that the learners write sentences of their own. This shifts the focus of the question from identifying the appropriate form of the vocabulary items to a more complex skill of constructing sentences. Besides, the learner also has to 'invent' contexts to suit the different forms of the given vocabulary item.

Question is of Y is better because it restricts the learner—task to the set objective of using the appropriate form—by providing the contexts and fully structured sentences.

b) Clarity, precision and adequacy of test specification

Question I is vague, because it does not specify how many different forms are to be attempted. It does not also restrict us as to which types of forms are to be used.

The infinitive forms (like to be relieved, to have relieved, to have been relieved, etc.) or the *participal forms* (like relieving, relieved, being relieved, having been relieved, etc.) the noun forms (relief and reliever).

There is no specification about the nature (simple/complex, etc.) of the sentences to be given and the kinds of context to be presented.

A sentence may be at different levels of complexity. Similarly, a context may be presented with different depths of inventive or imaginative intensity. There is, of course, no easy and accurate way of putting down the restrictions within which sentences and contexts are to be conceived and presented.

The format of question II avoids these problems of task specification, while still being functionally relevant to the objective. The elaborate instruction and the appropriate format fix the 'frame' within which the learner is to act to satisfy the set objectives of the teacher.

Check Your Progress 2

i) Facility value (Fv) is a measure of the 'difficulty' of a given question, whereas discrimination Index (D1) in a measure of the extent to which a question discriminates between the more able and the less able learners.

ii) An item can be retained in a test if the range of its Facility Value (FV) is from 25% to 75% and the range of its DI is from +0.20 to +1.00-

Check Your Progress 3

Achievement tests and diagnostic tests differ in their purpose and also in their treatment of the learning-content they are supposed to deal with.

An achievement test may enjoy a high content validity if it makes an adequately representative sampling of the learning content that it is concerned with (since its purpose is to measure the extent of an achievement, not to identity the specific lapses).

But for a diagnostic test, if meant for identifying the specific lapses in learning on the part of individual learners, to be credited with content representation of the learning-content is possible.

Check Your Progress 4

i) Content validity

ii) Predictive validity

iii) Concurrent validity

Check Your Progress 5

i) b, c, e, f and g.

ii) b

Check Your Progress 6

The three measures of internal consistency are:

i) Split-half estimates

ii) Kuder-Richardson estimates, and

iii) Cronbach alpha.

Split-half estimates are used when we want to know the extent of equivalence in 'content and difficulty' of the two halves of a test. Kuder-Richardson estimates are used when the items of a test are homogeneous and do not require splitting the test in half for scoring purposes. Cronbach alpha measure compares the variance for any single item with the variance for the entire test.

UNIT 4 EVALUATION OF DISTANCE EDUCATION SUB-SYSTEMS

Structure

4.0 OBJECTIVES

After going through this unit, you should be able to:

- describe at least three types of student assessment in the context of distance education institution;
- explain the evaluation processes involved in instructional materials;
- describe the evaluation processes concerned with support services in the distance education system ; and
- Identify the tools used for evaluating staff development activities.

4.1 INTRODUCTION

In this block you have already read in Units 1, 2 and 3 the concept, techniques and tools of evaluation. In section 1.3 of this block we have discussed the different tasks and concepts of curriculum evaluation in distance education. This would have given you an idea of the broad areas in a distance education system. Distance education has its sub-systems like services provided by the study centres, dispatch of learning materials, utilization of learning materials, personal contact programmes and evaluation. We have selected a few sub-systems, from the many, for our discussion in this unit. Here, you will find how evaluation practices are carried out in these sub-systems. While going through this unit we will realise that the boundaries between each of these sub-systems are conceptual and blurred. Each of the components/sub-systems in the distance education system is interdependent. It is essentially because of this reason that the effectiveness of one component or the lack of it is likely to have an impact upon other areas as a chain reaction. We have also touched upon various techniques and methods of evaluation (which have been discussed in Unit 2 of this block) in order to explain how we can use a tool or a technique while evaluating an area.

4.2 ASSESSMENT OF STUDENT PERFORMANCE

Student assessment is at the heart of an integrated approach to students' learning. There are various reasons for which assessment can be used. They are:

- to assess/diagnose the entry level capabilities of a learner
- to measure students' progress during a course
- to give students feedback on the progress
- to consolidate students' learning and the work done so far
- to motivate students
- to give students a reward and a sense of achievement
- to give guidance to a learner as to what he/she will be tested on
- to assess the amount of effort put in
- to assess level of achievement at the end of a course
- to grade in relation to other students (norm-referenced)
- to grade in relation to criteria of excellence
- to maintain a desired standard
- to give feedback about the course
- to enable the course to be evaluated
- to provide an institution with patterns of regular entries and exits
- to provide an institution with a target and feedback information on success and failure
- to give universities/institutions public accountability
- to predict the future performance of a learner
- to assist the learner for personal development.

This clearly shows that student assessment has a significant role in an evaluation system.

4.2.1 Admission and Entrance Test

Besides the issues of distance education policies and the need for clarity about our educational goals, the calls for accountability also point to the need for greater attention to our methods of student assessment. One of the ways to improve distance teaching is through adopting various ways of testing. Just the mere collection of information about the distance learner does not mean this information is of sufficient value to use in our decisions about courses/ programmes/learners: the *quality* of information is vital. Here we will focus on one of the methods/techniques of student assessment for registering in a course/ programme.

Flexible course-entry means student admittance which is not dependent upon specific qualification. However, a few courses/programmes may require such qualifications because of their content demands, and so admission tests become necessary. Normally these entrance tests assess a prospective students' general awareness about the situation he/she lives in, his/her language competence and his/her skills at numeral and analytical abilities. These tests aim at assessing the aptitude of the learners for a particular course/programme. Success at such tests

does not qualify a candidate for any kind of certification; instead achieving a pre-determined minimum norm at such tests entitles the candidate to seek admission into the programme for which the entrance test is held.

4.2.2 Student Learning Assessment

Learning processes are not only concerned with the absorption of information but also with developing needs skills to locate and find relevant materials to assist the learning. Students often feel that they can learn by memorizing the required knowledge and practicing the required skills to get minimum marks. Indeed the students who have properly understood the principles involved and try to think problems through, may find themselves at a disadvantage in the limited time allowed in an examination. There seem to be good reasons therefore to test students' understanding, knowledge and skills with different kinds of tests.

Knowledge and skills expected of students can usually be identified quite easily against fixed performance "objectives". The assessment test should assess student performance against these objectives. Objective type questions can efficiently test certain types of knowledge and tasks related to skills, such as numeracy, literacy communication and interpersonal skills. These questions can be devised without much difficulty. Testing affective domain is however more difficult because the level of understanding and the feelings cannot be inferred from a learner's performance with certainty. Due to this reason, some use oral examinations. With distance teaching, the possibilities for oral examination are limited, though recorded cassettes can be used for this purpose. In distance education, projects are widely used for this purpose and can be designed to test understanding, though this is a very time consuming process.

4.2.3 Assessment Process

Assessment, as we have seen earlier, is a matter of judgement, not simply of computation. Marks, grades and percentages should not be treated as absolute values but as symbols to be used by evaluators to communicate their judgement of different aspects of a student's performance. The purpose of assessment is to enable distance learners to demonstrate that they have fulfilled the objectives of the programmes/courses of study and that they have achieved the standard required for the award(s) they seek.

Assessment helps to determine knowledge, ability, competence and experience in distance learning settings. These assessment activities, range from informal questioning during tutorials/counselling or working on in-text questions to formal assessment such as continuous assessment, term-end examination, project work, field visits, case studies, practical work etc.

In a distance education institution, student assessment for purposes of certification may have at least three, and in certain cases more than three components. The three components are:

i) Self-assessment;

ii) Continuous assessment; and

iii) Term-end examination

Though you have read about these components earlier in Unit 1 in the context of IGNOU, it is worthwhile to go through the following three components again, discussed from other angles.

Self-assessment

The first component of student assessment process is self-assessment. This provides the learner with tools to assess the learning processes they have been

engaged in and to consider what they mean for them. Learners do have an important role in the evaluation process not only as providers of feedback but also as evaluators of their own learning.

Evaluation can be more effectively conducted by creating a specific opportunity for the learners to reflect on their reasons for learning, what they are achieving out of learning and where they want to go next. We will discuss how self-evaluation of a learner does take place at various stages of learning.

Table 1 shows you the kinds of questions which would prompt thinking about reasons for study.

Table 4.1: Learner self-review: reasons for learning

Reasons	Strongly	Agree	Undecided	Disagree	Strongly disagree
1) It would help me in my job					
2) I would like to change my job					
3) 1 enjoy learning					

During the learning process and at its end an important part of the students' learning can be assessed by knowing their level of understanding of the facts, of their own strengths and weaknesses, inhibitions and styles of thinking and working in relation to the varieties of constraints and opportunities presented by the course.

We can use different types of tools for evaluating student performance during and after learning. Here are three types of tools which we can use for self-assessment;

i) Self-assessment questions/activities;

ii) Self-review exercise; and

iii) Keeping a journal.

Let us talk about each one of them in the given order.

The first component of self-assessment comprises in-text questions/activities built in the self-instructional materials. Here the content and style of each item is designed to be appropriate for the particular learning objectives involved. As a rule, answers or suggestions leading to answers to these in-text questions are provided alongwith the study material. A distance learner is expected to work on these in-text questions on his/her own, and having worked on questions, to look for reinforcement in the answers provided in the materials themselves. Here the learner gets feedback on the progress he/she is making and this does not have to depend on the physical presence of a teacher, but can be built into the medium being used, by using a variety of in-text questions and activities. Thus, looked at differently, self-assessment in this case is a type of formative evaluation.

Self-review questions help the learner reflect back on a period of learning to review his/her own learning. He/she can pose three types of questions. They are:

i) What is the purpose of this unit/section?

ii) What have I learned?

iii) How can I apply it?

If the learner feels that he/she is not able to answer the above questions he/she can again review the unit to achieve the objectives. The value of this exercise is that the learner is given time to think about what he/she has learned and how it might be applied to what he/she does in future.

Journal keeping is especially valuable where the content of the course-work is closely related to the learner's work context or life experience. Keeping a journal is an important means through which the learner works with and learns from his/her existing experience and integrates it with the new learning gained from the course.

Self-review exercises and the keeping of a journal are important ways of recognising the significance of the learners' own involvement in the learning process, and of encouraging him/ her to spend some time thinking through the outcomes and implications of his/her study.

Continuous assessment

The second component of student assessment is operated through assignments. These assignments form a part of the total evaluation of the students' progress and provide the students feedback on how well they are progressing throughout the course. Students are expected to work on assignments provided with the self-instructional materials. These assignments demand written responses which are evaluated by the distance education institutions, these assignment-responses are commented upon and also graded with a view to helping the students improve their performance and also learn about the drawbacks or weaknesses in their responses. The comments are expected to be elaborate and of the teaching type. Thus, besides being a vehicle for two-way communication, assignments serve as tools for assessment of student performance. In other words, assignments are used for two purposes at one and the same time: they serve the purpose of formative evaluation and also the purpose of summative evaluation. The grades scored in assignments constitute a component of the overall score a student makes in a course. This component is called 'continuous assessment'. Also assignments can be written and commented upon on-line, as in the case of web-based courses.

Formative and/or assessment is possible by ensuring that each assignment contains plenty of opportunities for learners to receive detailed, positive and timely feedback, with lots of advice on how to improve. This not only informs distance learners' activities but also enables them to develop continuously and to achieve better results. Through a self-evaluation where they are asked to make review of their own study, they learn a lot about the processes of assessment and learning and this gives them insights into ways of improving their work.

Term-end examination

In open/distance education institutions the measurement of whether adequate learning has taken place is usually left to formal examination and continuous assessment systems. The overall student assessment is based on the term-end examination which in most cases is a conventional three-hour written examination. This component serves the purpose of summative evaluation. Most courses end with a formal and external examination.

Check Your Progress 1

Notes : *a) Space is given below for your answer.*
b) Compare your answer with the one given at the end of this Unit.

State the process of continuous evaluation in distance teaching.

.....

.....

.....

.....

4.3 COURSE EVALUATION

Programme evaluation comprises of evaluation of all the components of a programme, starting from the need analysis and objectives of the programme, through learning experiences, to the learning outcomes. Course evaluation is a part of this process in which a particular course and its components are evaluated. Evaluation of the course by completion rates, comparison of academic achievement between off-campus and on-campus students and qualitative feedback from students demonstrates that the distance education mode is as successful as on-campus teaching in providing high quality teaching in developing countries. Quality control in distance education frequently centres upon the review, evaluation and subsequent redesign of distance learning materials. The evaluation of distance learning materials is an umbrella term which encompasses evaluation of a single course or whole courses or a programme(s), or curriculum. The aim of evaluation is to identify whether a course is well managed, well presented and up to date. When undertaken periodically, evaluation helps statements to be made about the appropriateness, worth, coherence and relative balance of the learning situations provided in relation to their intended outcomes. Woodley (1992) states that the aim of course evaluation is to improve the quality and effectiveness of the teaching and learning that takes place. The evaluation of distance teaching materials may seek to provide information that can be used during the process of developing or preparing materials for learning experiences-formative evaluation procedures -- or information about how well the 'finished' instruction has worked in normal use summative evaluation procedures.

Reasons for course evaluation

Most evaluation concepts have as their main aim helping to develop courses under optimal conditions neglecting the review of real conditions under which the course material is actually used. These evaluation concepts involve certain risk, because in principle they suggest that it is possible to have the full range of knowledge about all future conditions and that to develop perfect course materials for use under all conditions. Researchers do not accept this point of view on the basis that the conditions will change, necessitating new didactical decisions. It is worth noting that courses are designed and developed under ideal/lab conditions. This has to be improved upon for their use in real-study conditions of a distance learner. The reason for this is that investigations into the behaviour of human beings under quasi-experimental conditions yield only limited information about real-life conditions. For this reason, it seems absolutely necessary to evaluate course materials under real-study conditions because only then will its true values become clear.

For evaluation of the effectiveness of the whole course, we need to collect information from

- students;
- course writers;
- buyers/users /employers; and
- other stakeholders.

For evaluating a course we need feedback/information from different audiences and we have to use various tools and techniques. Now let us discuss four methods in the context of course evaluation.

Formative evaluation: This approach refers to the process of evaluating the courses to provide feedback and information that can be used during course development. This can be conducted in the following two ways:

i) critical commenting

ii) Developmental testing

You may look at the stages involved in course development process in the following diagram, and locate the activities undertaken for formative evaluation:

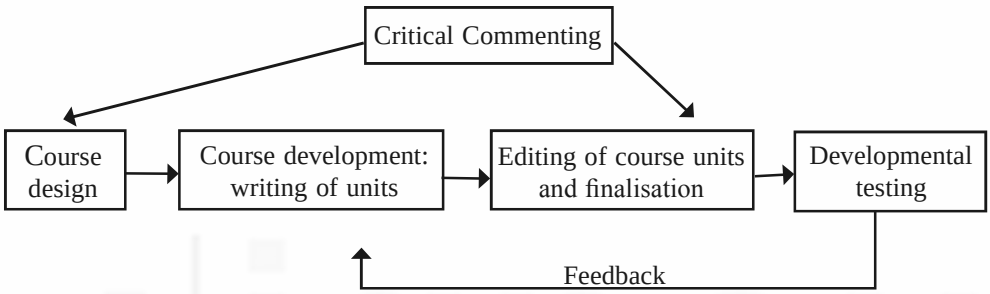


Fig. 4.1: Stages of course development

Critical commenting includes reviewing of course design, course objectives, learning experiences, instructional materials, student assessment, etc. by subject specialists, instructional designers, teaching technology experts and psychologists. This requires a clear framework against which feedback is sought further; in open/distance education contexts peer review of draft materials is a common process. At an informal level this may simply involve one or more colleagues reading or looking at draft materials and providing comments in terms of the suitability of content, the style of presentation and so on. On the other hand, arrangements may be made for systematic critical commenting, with tutors or course writers reviewing the materials prepared by others working on the same course or programme. The reactions of colleagues can also be augmented by adopting the more formal procedure of collecting data with the help of a questionnaire or inviting one or more experts in the field to act as assessors to comment on the draft materials or both.

The aim of developmental testing is to get students to work through the materials while they are in draft form in order to identify problems with such aspects as:

- clarity of aims and objectives;
- sequencing and logical arrangement of the content;
- retention of interest;
- comprehension;
- difficulty level;
- work load; and
- feasibility of student activity.

Developmental testing takes place during the preparation phase and involves trying-out draft teaching materials with students. The feedback obtained is used to guide and inform course writers' revisions of materials before they are committed to print or tape. Such testing may range from a fairly informal student trial of a single unit to an elaborate procedure for testing draft materials for a whole course of instruction.

The use of review of learning experience of the learners at the developmental stage of a course is very helpful in itself and it is a way of getting learners to identify their own problems as well as any weakness in the course. Feedback questions with tutorial support could be used to improve the learning condition as well as to make changes for the benefit of prospective learners.

In the conventional face-to-face provisions, a classroom teacher receives a whole range of informal feedback from students about the relevance and quality of a course. The attendance of the students at the tutorials, the enthusiasm and knowledge they display in their oral, practical and written work, their comments to the teacher and their success in assessment will all combine to give the teacher a great deal of information about the learning effectiveness of the course. However, with open distance learning, it is likely that there may be little or no direct contact between the course designers, the course providers and the distance learners. Course feedback may be the main, if not the only channel of communication between the students and those who carry the responsibility for their distance teaching. In effect, course evaluation at the presentation stage has to act as the students' voice. At the same time, there may also be support staff working with the students who will have their own views on the learning effectiveness of the course and on the role they are expected to undertake.

4.4 EVALUATION OF INSTRUCTIONAL MATERIALS

Evaluation of instructional materials comprises the following aspects:

- Didactic aspects : Suitability of materials for self-study, degree of difficulty, interest and clarity of texts, graphic presentation, audio-video aids, etc.
- Scientific aspects : A balanced presentation of varied scientific approaches, updating of materials, relevant bibliography, research studies, etc.

Nevertheless, these criteria are to some extent abstract. While evaluating instructional material it is essential to examine a "live course" in order to assess the effect it has upon students. Data therefore should be periodically collected with regard to all course components and made available to course teams.

For the evaluation of the instructional materials, we may have any one or more of the following approaches of evaluation:

- Piloting;
- special evaluation; and
- routine evaluation.

Piloting: In this approach, the materials are tried out with the first batch of students and revisions are brought about subsequently. This is a well known approach and very often the management takes to it almost without giving any thought to its utility in relation to its cost. There are also reasons for not adopting this approach for each and every instructional materials.

It is not advisable, as far as educational ethics is concerned, to always use the first batch of students as guinea pigs.

Financial as well as human resources may or may not be adequate for conducting piloting and then offering an improved course to the second batch of students.

Operationally, bringing in revisions immediately after the first launch affects distribution and support services adversely.

The cost of course design and rate of course production also get adversely affected by this approach.

Special evaluation: This should come in whenever needed and proposed as a project for which the resources are requested for or they are available from a source other than the institution. Obviously, such evaluation will need to have a purpose and, therefore, resources should be available for fulfilling the purpose. Such evaluation is obviously not a matter of routine and so it has to be selective. Two main considerations for opting for the special evaluation of instructional materials are:

- In case of unusual and inexplicable poor performance of students including adverse reaction by students/counsellors and academics to the materials.
- In cases when a funding external agency asks for such evaluation.

Routine evaluation: Routine evaluation is best done as a part of management process. This approach keeps the faculty/schools in constant touch with the students and the materials, resulting in an appropriate and timely feedback to the schools, course writers and academic counsellors. It will eventually feed into the process of course maintenance which can be carried out through supplementary materials without wasting resources. In addition to instructional materials, assignments and the term-end examination question papers also are evaluated.

You may note that no single approach can provide reliable and valid data for evaluating instructional materials. A particular approach adopted by any particular team of course designers must depend on the *circumstances*, the *constraints* and the *contexts* within which they are working.

4.4.1 Print Materials

Print material is utilised well in distance education courses as this form of technology is the cheapest. It is also the only form of technology that can reach a wider range of learners and especially those in the very remote areas. Print medium remains by far the most widely used medium for tertiary level distance education. They are made up of many textual components and design features with psychological inputs. Various research studies have been conducted on how the presentation of texts for distance education be improved by seeking opinion from theoreticians, practitioners, psychologists, distance learners and other users so that it should be effective from a pedagogical point of view.

You may note that by their very nature, distance education print materials are subject to re-drafting and updating with the aim of improving their usability. We use different types of tools for evaluating this specific medium. Many tools for evaluating this medium include attitude scales, rating scale, and checklists, questionnaires, interview schedules and tests. We can also conduct the following quasi-experimental design for evaluating the materials.

i) pre-test/post-test evaluation; and

ii) follow-up evaluation.

For the procedure for using these, you can refer to Units 1 and 2 of this block, in which we have already discussed how to use them for evaluation of course materials.

4.4.2 Electronic Materials

Some instructional objectives, especially those concerned with acquiring *skills and/or attitudes* cannot be realised by written materials alone. A few research studies have focused upon the use of every medium within the total media mix of a course. Here we will focus upon audio broadcasts and computer based electronic materials, while keeping matters such as functionality, costs, feasibility and compatibility with the principles of distance education constantly in mind. Audio tapes are a simple, unsophisticated medium. In terms of accessibility the audio tape runs almost equal to print-based materials. Despite its inherent simplicity and the ubiquitous availability of cassette players, audio tapes are a much underutilized medium for the delivery of educational materials.

While evaluating the audio tapes the following **criteria** should be considered.

i) Simplicity—easy to use, can be played, paused and stopped for a period.

ii) Flexibility—can complement a variety of other media.

iii) User control—the learner can use whenever she/he needs.

iv) Humanising tutor—student relationship.

v) Stimulating and motivating learners by using sense organs (ear), variations in phase etc.

vi) Variable concentration.

vii) Time.

Depending on the above criteria, audio cassettes can be pilot tested with the help of a questionnaire administered to the tutors/students (present and past).

Television-based materials

The various forms of television-based materials used for distance teaching are broadcast, cable, satellite, video tape, video disc, etc. The use of video for distance teaching is well established and its advantages are well documented (Kirkwood, 1990; Bates, 1993). This medium is a potentially powerful component because it has the capacity to integrate the two most important senses in the transfer of information: visual and aural. Broadcast television has the potential of reaching mass audiences. The video tape medium allows storage of information for individual, group or mass audiences. These can be used or repeated as often as required. Combinations of video tape, satellite, micro wave, digital and cable technologies allow targeting of specialised groups and audiences, and the possibility of two-way communication and even interactive networks.

While evaluating television based materials the following criteria should be considered:

i) delivery system;

ii) picture quality;

iii) efficiency of operation;

iv) access;

v) correctness;

vi) Reliability;

vii) validity;

viii) maintenance;

ix) life-span;

x) documentation; and

xi) user-friendliness.

Three main areas considered for evaluation of computer-assisted learning (CAL) are:

a) Case of use (user evaluation);

b) Programme performance (functional evaluation); and

c) Accuracy, clarity and depth of evaluation (content evaluation).

The major purpose of *user evaluation* is to determine whether or not the interfaces are consistent and easy to use. *Functional evaluation* consists of determining whether or not the programmes perform as specified. The *evaluation* of content will ensure accuracy and determine whether or not students are able to meet the objectives.

Check Your Progress 2

Notes: *a) Space is given below for you to write your answer.*
b) Compare your answer with the one given at the end of this Unit.

List at least five criteria for evaluating an audio cassette.

.....

.....

.....

.....

.....

4.5 STUDENT SUPPORT SERVICES

The student support services provided to a distance learner are an aggregate of the organisation of the teaching-learning process to enable him/her to acquire learning skills. It also includes individual tutoring and group tutoring, counselling, contact programmes, two-way communication between the learners and tutors. The evaluation of tutoring and counselling will help improve the support services and the course design and thus will help to meet the needs of distance learners.

To evaluate student support services we seek answers to the following questions:

- Do these services cater to information needs promptly, adequately and convincingly?
- Do these services provide for advice at pre-course, on-course and post-course stages adequately and effectively?
- Do these services provide for academic-counselling adequately and conveniently? This includes the effectiveness of tutors with regard to their work on assignment-responses and face-to-face situations.
- Are these services easily accessible?
- Are the staff involved in these services attitudinally tuned to the kind of work assigned to them'?

- How do the students rate the support services?
- Do the support services make a pedagogically rational use of various media in operation?
- Do the support services help the learners get a good score in the examination?

These questions pertaining to evaluation of student support services involve a number of variables.

These variables are:

i) human beings in the roles of coordinators, academic counselors, etc.;

ii) geographical situations such as towns, remote areas, rural areas, etc.

iii) pedagogical factors such as classroom techniques, counselling, advising, etc.

This should give us an idea that evaluation of support services is a complex area and so should be dealt with carefully.

4.5.1 Tutoring and Counselling

We have talked about these concepts at length in MDE-411. Tutors/academic counsellors are essentially engaged in a wide range of tasks as shown here.

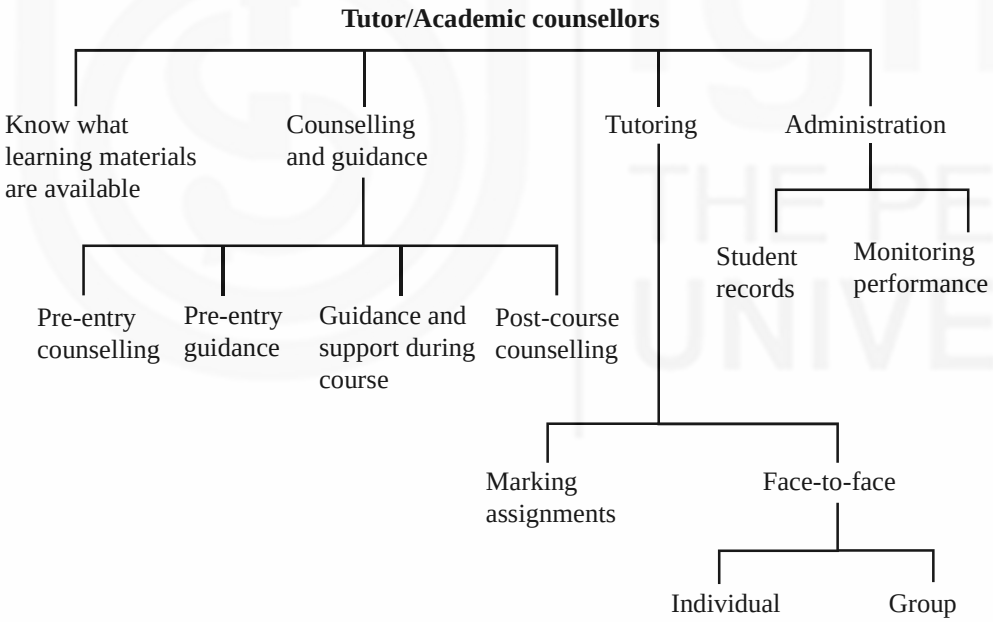


Fig. 4.2: Tasks of academic counsellors

In other words, one can say that the role/task of a tutor is to:

- inform, advise and counsel students;
- meet students in face-to-face situations for a number of sessions during an academic year;
- assess assignment-responses; and
- help distance learners learn to learn.
- Correspondingly, evaluating the tutoring/counselling sessions involves:
- understanding the nature of communication between academic counsellors and the students — through letters/telephone/or in face-to-face situations;

- assessing the quality of their comments on assignment-responses;
- evaluating the reliability and validity of the assessment of assignment-responses; and
- analysing the turn around rate of assessed assignment-responses.

As you know in most forms of distance learning, face-to-face tuition has a voluntary aspect. It is provided with the assumption that learners should use them. But do they in fact use it, and for *how long and how often?* Were the facilities adequate? Is it for better learning or do the arrangements match learner convenience? These are important questions particularly for providers where tutorial attendance is not made compulsory in most of the programmes.

Evaluation of face-to-face tuition is essential for two main purposes: the first to ascertain the nature and distribution across the year of tutorial and counselling tasks; second to provide indicators and its effectiveness as a medium for teaching and learner support. In this case, the value of attendance serves as indicators. This requires the cooperation of learners, clerical staff and tutors depending on the system. Monitoring of attendance can help us know how often students have used the session. This is vital for organising tutorial sessions.

As far as face-to-face sessions are concerned, we evaluate academic counsellors' competencies, clarity of presentation and handling of discussions, punctuality, regularity, nature of rapport with students, quality of the conduct of sessions, motivation or enthusiasm and students' satisfaction.

For example, we would like to explore the value of tuition and quality of the role of the tutor in general. For this purpose, we can design a questionnaire and students can be asked to respond to a very wide range of statements both about the course and about their experience in the tutoring sessions.

For this, we can develop a 5-point attitude scale as given in Table 4.2.

The evaluation of the support system also consists in evaluating the counselling. In open distance education, counselling involves:

i) developing study skills in people who left or did not have formal education;

ii) helping the students with some constructive responses when they come out with personal problems over the phone/or to the study center; and

iii) helping students with study methods at (a) start of course (b) during the course, and (c) near examinations.

Table 4.2: Student evaluation of the role of the tutor and tuition

Attitude statements	Strongly agree	Agree	Undecided	Decided	Strongly disagree
1) A good tutor can make a course more interesting					
2) Some aspects of most courses can only be taught effectively on a face-to-face basis.					

Attitude statements	Strongly agree	Agree	Undecided	Decided	Strongly disagree
3) If the amount of tutor support is reduced, it would adversely affect my ability to cope with open university studies.					
4) Without the help of a tutor I would probably have dropped out from at least one course that I stuck with.					

Issues in evaluation of counselling session

Distance education practitioners turn to evaluation of counselling to provide clues about the need for counselling, perhaps more realistically, evidence of the demand for it, which can inform decisions about how best to provide it. Thus, the major issues in counselling evaluation are:

- How much is counselling used and by how many learners?
- Should tutors also offer counselling or should there be separate counsellor?
- How well do part-time counsellors as well as full-time staff perform these functions effectively? What training do they need?
- Is it possible to counsel at a distance with the learner using the telephone or written communication or both?

4.5.2 Two-Way Communication

Tutoring and counseling, and communication through assignment-responses facilitate personal interaction between a tutor and a distance learner. Tutor comments (written on assignment-responses) provide feedback to students about their own progress. Although grades awarded on assignment-responses are very important to the learner, the purpose of the assignments includes the assessment of the student's progress in a course. Tutor comments are thus a vehicle for two-way communication between a distance learner and tutor. In a distance teaching context, monitoring of tutor comments requires that assignment-responses already evaluated by a tutor may be assessed by a different person. A particular percentage of evaluated assignment-responses may be used for this purpose.

The reliability and validity of the assessment-responses: The speed with which tutors mark and return assignments is not the only important issue. There is also the issue of reliability and validity in awarding grades. In the present case, we are concerned with how truthfully an assignment question measures what it purports to measure. In Block 4 of this course we have already studied that reliability is a prerequisite of validity, that is, to be valid a test must be reliable. Reliability controls validity. Some of the most effective measures for improving the reliability of a test however revolve around opportunities for tutors to discuss their expectations together—the marking process.

The turn around rate of assessed assignment-responses: In this process, we evaluate how quickly the learner gets back an assignment, after posting or handing in, and how much time is taken by the tutor in evaluating it. Many distance education institutions can monitor tutor turn-around very easily because distance learners mail/submit their assignment-responses directly to the institution. The institution therefore knows and records when the

tutor should have received what, and can 'chase' late returners, if there are any. However, when the student number is larger, this does pose operational problems.

Check Your Progress 3

Notes: *a) Space is given below for you to write your answer.*
b) Compare your answer with the one given at the end of this unit.

State how you can evaluate the tutoring/counselling sessions.

.....

.....

.....

.....

4.6 STAFF DEVELOPMENT

In a staff development programme we generally evaluate two things: learning and performance. Although these two components are related, each needs to be considered separately. If you are not clear at the start of the programme about what you intend to find out from evaluating each area, it will be impossible for you to get this information when the training programme ends. Since evaluation deals with change, you will first need to determine the distance education practitioners' current level of performance. In this way you will be provided with information about what has changed and what changes you can expect when the employee returns to his/her work place.

A training programme delivered through a contact mode or distance mode should have specific objectives. These are stated as training objectives, but a programme coordinator is more concerned with anticipated changes in performance than with whether the training objectives have been met. This being the case, the practice in staff development programme is that the objectives are stated in terms of *change in performance.* You should not, however, ignore the learning aspects, as without this the performance cannot be expected to change much. Various methods for identifying training needs are:

- task group;
- consultation of experts, clients and target audiences;
- quantitative analysis of knowledge/skill area;
- monitoring and analysis of demand for existing courses; and
- studies of employees to identify current anticipated training needs.

As a coordinator of a staff development programme you may be required to collect base-line data. This data will help you plan the training programme as well as show how performance has improved after the training programme. If you do not have the information it is impossible to compose? performance after the training experience. Qualitative data usually takes the form of narrative reports, journals or other verbal assessments. It is possible for the institution where the staff is employed to give you more information than you need or to give it in a form that is readily comprehensive or usable. The institution in this way can prepare a qualitative report before the training and after the training.

There are a few criteria which are useful for evaluating a staff development programme.

1) What information is essential to know about performance evaluation?

   This deals with transferring the new knowledge from the learning situation to work situations. You will want to know through a systematic evaluation, whether the trainees' performance has changed. You may need to have some assessment: how much did performance change? Too often a course coordinator feels that once the employee has returned to work from the training programme, it is difficult to assess his/her performance in the actual work situation. So the trainer/organizer of the training programme needs to know enough about the change in performance to determine the relative success of the programme with the help of *follow-up* activities.

2) Was the cost of the programme worth the results?

   To answer this question you and the institution must first know what the total cost of the programme is. You can count the rejected materials and wasted time or the need for additional services of experts or resource persons and other associated staff calls, etc. Assessing costs becomes difficult when we are dealing with performances and training output.

   At the end of a staff development programme you will have to decide whether providing training will cost more than the benefits you can derive from it. If you did not make that decision before providing the training programme, it should be done afterwards and used as a basis for future training programmes and decision-making.

3) What kind of data should be collected?

   The two categories of data are classified as quantitative data, things that can be counted or measured, and qualitative data, where counting is not possible but observation can be effective. That is, after a training programme an individual can he observed to determine if his/her performance has improved or is up to the set standard.

4) How much change can we expect in a training programme?

   For example, if a job requires delivering of course materials, no amount of training pertaining to writing course materials will be able to bring about the desired change. Your expectations have to be realistic.

Evaluation of staff development involves:

- training needs analysis;
- prioritisation of training needs;
- continuous assessment of training programmes; and
- follow-up training programme to see the effectiveness/quality of a programme.

We should note here that evaluation of staff development programmes ranges from the use of financial measures to subjective measures of change seen in the behaviour of participants after training. For this, several measures can be used:

- financial management;
- information provided to participants (on content, formal material provided, etc.);
- observation of the training activities;
- using questionnaires (open and closed type) to get feedback; and
- follow up.

Many research studies have focused on the need for an explicit policy by the institution about staff development for distance education (Parer, Croker and Shaw 1988; Benson 1990). Such a policy emphasises effective teaching and learning at a distance, the preparation for the production of study materials, support services and promotion rewards for those who are successful in distance education. Few institutions all over the world have such an explicit policy. (We have discussed at length evaluation of staff development for distance education in MDE-419.)

Check Your Progress 4

Notes : *a) Space is given below for your answer.*
b) Compare your answer with the one given at the end of this unit.

List the four criteria useful for evaluating a staff development programme.

.....

.....

.....

.....

.....

.....

.....

4.7 LET US SUM UP

In this unit, we have discussed how to evaluate a few key areas in distance education systems which are interdependent. We have presented to you, the way to assess student performance which occupies a significant role in an evaluation system. As you know, quality control in distance education frequently centres upon the review, evaluation and redesign of learning materials. We have discussed course evaluation with reference to developmental testing, formative and summative evaluation. We have focused on how to evaluate the electronic materials with the help of test tools. Besides this, we have touched upon the evaluation of student support services and staff development programmes.

4.8 ANSWERS TO CHECK YOUR PROGRESS

Check Your Progress 1

Assessment of student performance is operated through assignments. The students work on assignments and prepare assignment-responses. These assignment-responses are commented upon and also graded with a view to helping the students improve their performance. The grades scored in assignment-responses, constitute a component of the overall score a student makes in a course. This component is called continuous assessment.

Check Your Progress 2

i) Flexibility
ii) Simplicity
iii) User control
iv) Stimulating and motivating.
v) Time

Check Your Progress 3

Evaluating tutoring and counselling sessions involves understanding the nature of communication between academic counsellors and students through letters, over the telephone or in face to face situations, (the quality of their comments on assignment-responses.)

Check Your Progress 4

i) Information essential to know about performance of the participants in a staff development programme

ii) Cost of the programme

iii) Data for need assessment

iv) Change we expect in a programme.

REFERENCES

Alvarado B.J., Agostino, S.G.D. and Balajios, M.G. (1991) : 'Orientation and conceptualisation of the academic quality control centre', in R.Schuemer (Ed.) *Evaluation Concepts and Practice in Selected Distance Education Institutins*, (pp 37-51), Hagen: Fern Universitat –Gesamthoebschule Commonwealth of Learning, Vancouver, British, Columbia, Canada

Beattie, A. (1993) : *Figuring out Quality of Open University Services* (Report No.8), Milton Keynes: IET, Open University.

Calder, J. (1994) : *Programme Evaluation and Quality: A Comprehensive Guide to Setting up an Evaluation System*, London: Kogan Page Limited

Cronbach, L.J. (1964) : *Essentials of Psychological testing*, London.

Cronback, L.J. (1983) : 'Course improvement through evaluation', in C.f. Madaus, M.Scriven, and D.L. Stufflebeani (Ed.) *Evaluation Models: View Points on Educational and Human Services Evaluation*, (pp 101-115), Bosten/ The Hague: Kluwer Nijhoff.

Deshpande, P.M. (1994) and Mugridge, I. : *Perspectives on Distance Education Quality Assurance in Higher Education*, Vancouver: The Commonwealth of Learning.

Freeman H.E. (1989) : Evaluation: Concepts and Practices, Newbury Park: Sage

Ganor, M. (1991) : 'The open university of Israel' in R. Seliuemer (Ed.) *Evaluation concepts and Practice in Selected Distance Education Institutions* (pp 83-85), Hagen: Fern Universitat _ Gesamthochsehule.

Gooler, D.D. (1979) : 'Issues Concerning Evaluation of distance Education in Asia', paper presented at the meeting of syracuse University, USA.

Gronlund, N.E. (1981) : *Measurement and Evaluation in Education: Guide to Best Practice*, London. Harper and Row.

Holmberg, B. (1989) : *Theory and Practice of distance Education*, London, Routledge.

John, B. (1995) : *Outcomes, Learning and the curriculum*, London: The Farmer Press.

Kelly, A.V. (1989) : *The Curriculum theory and Practice,* 3rd Eq.) London: Routledge and Kegan Paul.

Koul, B.N. (1991) : Evaluation at Indira Gandhi National Open University, New Delhi, India, in S.Redolf, (Ed) *Evaluation Concepts and Practices in Selected Distance Education Institutions*,Hagen: ZIFF.

Meintijes, L. (1996) : Distance Education at the Pacific Advertise college: An Evaluation, *Papua New guinea Journal of Education*, 1, 165-175.

Morris, A. (1990) : *Performance Indicators,* Report of a Committee of Inquiry, Polytechnics and colleges Funding Council, London.

Regal, O. (1992) : *The Academic Credit System in Higher Education: Effectiveness and Relevance in Developing Countries.* (The World Bank Report), Washington, DC: The International Bank for Reconstruction and Development.

Rossi, P.H. and Freeman H.E. (1989) : *Evaluation: A Systematic Approach.*

Rowntree, D. (1992) *Exploring Open and Distance Learning,* London: Kogan Page.

Rudolf, S. (Ed.) (1991) Third Ed. : *Evaluation Concepts and Practice in selected Distance Education Institutions,* Hagen: ZIFF.

Thorpe, M. (1993) : *Evaluating Open and Distance Lelarning* (Rev. Ed.), London:Longman.

Woodley, A. (1991) : 'Evaluation at the British Open University', in R.Schuemer 9Ed.) *Evaluation concepts and Practice in Selected Distance Education Institutions,* (pp 208-222),Hagen: Fern Universitat – Gesamthochschule.

Young, K.E.: Chambers, C.M. Kelts, H.R. and Associates : *Understanding Accereditation*, Jossey – Bass.