
UNIT 2 TECHNIQUES AND TOOLS OF EVALUATION

Structure

- 2.0 Objectives
- 2.1 Introduction
- 2.2 Observation
 - 2.2.1 Types of Observation
 - 2.2.2 Recording Information from Observation
 - 2.2.3 Uses and Limitations of Observation
- 2.3 Interview
 - 2.3.1 Types of Interviews
 - 2.3.2 Techniques of Interviewing
 - 2.3.3 Uses and Limitations of Interviews
- 2.4 Profiles
 - 2.4.1 Types of Profiles
 - 2.4.2 Uses and Limitations of Profiles
- 2.5 Rating Scale
- 2.6 Projects
- 2.7 Tests
 - 2.7.1 Types of Tests
 - 2.7.2 Uses and Limitations of Tests
- 2.8 Let Us Sum Up
- 2.9 Answers to Check Your Progress

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 occurring 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?

<i>Very</i>	<i>Slightly</i>	<i>Average</i>	<i>Slightly</i>	<i>Very</i>
<i>Effective</i>	<i>Effective</i>		<i>ineffective</i>	<i>ineffective</i>

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 tend 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 expectation 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.

.....

<p>.....</p> <p>.....</p> <p>.....</p>
--

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 sense 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.