UNIT 7 TOOLS FOR ASSESSMENT AND EVALUATION

Structure

7.1 INTRODUCTION

In Unit-6, you have studied about the techniques of evaluation and criteria of a good tool. In Unit-5 of this Block, you have studied various techniques, such as self-report, assignment, projects, socio-metric techniques, peer assessment, debates, school and club activities, cooperative and collaborative learning and assessment, and seminars use for assessing students performance in scholastic and co-scholastic areas. Units 5 and 6 of this Block will help you to understand the discussion in this Unit.

In this Unit, we will come to know about some prevailing tests being used in our schools to comprehend, evaluate and accomplish various aspects of teaching and learning namely achievement, oral, intelligence, and aptitude of the students.

These aspects not only help you to evaluate the learners and diagnose their problems but also assess your own teaching too, so that it can be improved. The most common tools that are used in the schools such as rating scale, questionnaire, inventories, checklists, interview and observation schedules,

*(**Note** : Some sections of this Unit have been partially reproduced from Course: ES-333 'Educational Evaluation', B.Ed. Programme, IGNOU, 2010)*

anecdotal reports, portfolio and rubrics are discussed in the Unit. This will help you to do an effective and enjoyable teaching and learning.

7.2   OBJECTIVES

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

- discuss the concept of tests and tools,

- differentiate between paper – pencil and oral tests,

- explain the concept of aptitude tests, their types, uses and limitations,

- discuss and administer the aptitude, achievement and diagnostic tests,

- explain the concept of intelligence test, its types, uses and limitations,

- explain a rating scale, its types, uses and limitations,

- explain the process of developing and use of questionnaire, inventory, checklist, interview and observation schedule,

- discuss the use of anecdotal records, and

- discuss the use of portfolios and rubrics.

7.3   TESTS

Tests may be of different forms, such as psychological test used to measure mental and behavioural traits, achievement test to assess performances of students, etc. They can be used to assess both the scholastic and co-scholastic abilities of the students. Let us now understand the concept of the test. The tests are those instruments by which you, as a teacher, collect information as data through verbal and non-verbal responses of the students. A concise definition may be : a test is an instrument or systematic procedure for measuring a sample of behaviour (Gronlund, 1990). Further, a psychological test is defined as a standardized, repeatable procedures used to elicit and measure samples of human behaviour' (Kazdin, 2000). From the above two definitions, you can summarize the meaning of a test and the sample behaviour it measures as follows :

- Human abilities, including intelligence, aptitudes, skills, and achievement in various areas.

- Personality characteristics, which include traits, attitudes, interests, and values.

- Adjustment and mental health, which involves detecting signs and symptoms of psychological and neurological disorders and appraising the effectiveness of psychological functioning.

(*Sourse* : *Kazdin, 2000*)

A test can be used for two broad objectives. First, it attempts to compare the same student on two or more than two aspects of a trait, such as attitude and aptitude of the same student; and second, two or more than two students may be compared on the same trait like attitude of two students (Singh, 2002).

The tests which you are going to study in this Unit are classified based on the various objectives or purposes to be fulfilled which are not only related to the learner but also useful to the teacher and teaching-learning process. Some tests have been discussed in the subsequent sub-sections.

7.3.1 Paper Pencil Test

The paper pencil tests comprise a standard set of questions which are presented to the student in writing on paper or Optical Mark Recognition (OMR) sheets that requires completion of cognitive tasks in the form of response by the student on those papers by pencil/pen mark. These responses or answers are summarized/scored to obtain a numerical value that represents a characteristic of the student for which the test was administered. The paper pencil tests can focus on what the student knows (achievement), is able to learn (ability or aptitude), chooses or selects (interests, attitudes or values), or is able to do (skill).

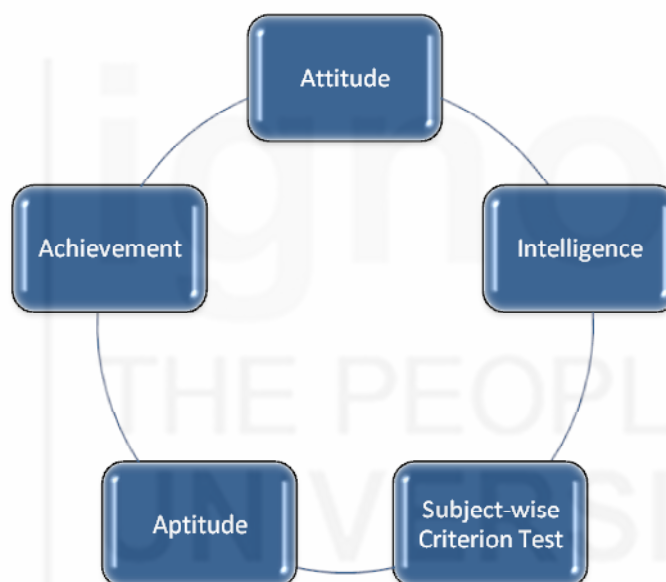Following figure shows different types of paper pencil tests:

Figure 7.1: Paper Pencil Tests

These paper pencil tests are easy to administer and score. They are economic and have standard questions which provide uniform procedure for all the students. These tests provide standard scores to the teachers.

7.3.2 Oral Test

Oral tests are those tests in which the response, solution or the task requires oral response to answer the question. Teachers' conversation with the students for the purpose of assessment, component of viva-voce examination for completion of a course, etc. can be called oral test. You might have noticed that in a written test, very little scope is left for the students to express themselves on any aspect whereas in oral test, students enjoy freedom to express themselves by citing many examples. In oral test, you can also use figures, diagrams, charts, maps, models, signs, etc. for asking students to explain the concepts covered in it. Oral tests have proved most valuable when used with students having language disabilities, the illiterates, shy feeling and the young children.

Actually, the oral tests rely on the students' oral performance to display their cognitive capabilities. This may be in the form of an oral discourse or a question and the answer session between the teacher and the students. These are effective for students with writing difficulties but may not be appropriate for students with hearing and speech difficulties or those who are prone to nervousness. For the students of speech difficulties, sign languages can be used, but in an inclusive classroom setting, sometime it is not possible to apply sign language as it requires sign language training both for the teachers and the students as well.

Oral testing is particularly important in language classes because spoken language plays an important role to express oneself. Oral tests focus on students' ability to communicate with what they know and how they articulate the things. Since spoken language is creative and flexible, students may find that oral testing gives them opportunities to self correct, ask for clarification and use language creatively. Teachers may appreciate the opportunity to hear students who choose not to speak on a daily basis in class.

The oral test is also having its own limitations as hardly it measures the skill of writing of the students. Conducting an oral test is also time consuming method, especially when it involves in-depth questioning. Oral test can be used as one of the components of overall assessment of the performance of the students.

7.3.3 Aptitude Test

An aptitude test is also an instrument used to determine and measure an individual's ability and skills to acquire, through future training. Aptitude tests may be classified into two groups : multiple aptitude test, and special aptitude tests. Multiple aptitude tests are those which intend to measure various areas of aptitude (musical, mechanical, etc.) each by independent sub-tests (Singh, 2002), whereas the special aptitude tests measure only one specific aptitude like teaching, musical, etc. The multiple aptitude tests measure abilities of students in more than one area simultaneously by using different sections of a test while the special aptitude tests measure ability of the students in one area.

Aptitude tests are used to predict the future performance of the students. These predictions are for the performance based on specific criterion (for example you all appeared before an entrance test for admission to the B.Ed. programme. The test included teaching aptitude components required to get into the B.Ed. programme) which are prior to instruction, placement or training.

The aptitude tests are also used for guidance, as well as prediction of success in some occupation. Training or academic courses are possible on the basis of scores on standardized aptitude tests. For example, for pursuing teacher education programmes, one needs to have teaching aptitude; for pursuing medical courses, one needs to have medical aptitude; for pursuing engineering courses, one needs to have engineering aptitude; for musical courses and sports related courses, one requires to have musical and the sports aptitude respectively.

7.3.4 Achievement Test

Achievement tests are administered on students to measure their learning outcomes. These tests are more prevalent in our schools. These tests show as to what has been learned by the students, rather than to predict future performance as in the case of aptitude tests. There are teacher made and standardized achievement tests. The choice of your achievement test will depend upon your purpose. The achievement tests can be used for following purposes:

- to know the learning progress of the students.

- checking any weakness in the instruction.

- in formulation of learning objectives and provide an easy means of critical examination of the content and the methods of teaching.

- adapting the instruction to the need of the individual learner.

- to know the effectiveness of any academic programme.

The achievement tests can be classified as :

- General achievement tests (batteries)

- Special achievement tests

General achivement tests attempt to measure the general educational achievement of the students at different stages which includes the common subject areas taught at the school in a particular class. Special achievement tests are meant for measuring the achievement of students in selected areas which may be grouped into two distinct groups – the diagnostic tests (discussed later in this Unit separately) and the standardized summative achievement tests. Diagnostic tests are used to know the areas of difficulties of the students and accordingly to provide suitable remedial instructions to them. Standardized summative achievement tests are used to grade the students in a particular standard and also to certify them.

An important precaution before using any achievement test is to assess the content validity with a standardized achievement test because content differs from one school to the other and even in the same subject. For using any standardized achievement test, you need to compare the content included in your curriculum for the level it has prepared and the components of the content included in the standardized test.

Though achievement tests are commonly used in schools to assess the students' achievement, there are some limitations of these tests. The following are the main limitations :

- Scores in the achievement test may not be the exact measure of achievement in the concerned field as there is subjectivity in assessment and the test selected may not completely represent the curriculum.

- Many factors such as standard of teaching, completion of curriculum, testing condition, prior information to the students, practice of the students, etc. may affect the students' achievement, but those factors may not be measured by the achievement test.

- Achievement scores of students differ in similar types of achievement tests. There may be the possibility that a student scores differently in different achievement tests meant for the same class.

You will study in detail the process of developing achievement tests in Unit-9 of Block-3 of this Course.

7.3.5 Diagnostic – Remedial Test

The achievement tests which are used to identify the strengths and weaknesses of the students are called diagnostic tests. The primary purpose of the test is to identify the problems faced by the students in achieving the learning objectives and to suggest remedial measures for them. Such tests are available in various

areas of study, such as in language, mathematics, science, etc. Diagnostic remedial test is specifically required in case the teacher observes certain learning difficulties in the students. For example, if a student scores 90 marks in Mathematics in the Class IX annual examination, but if you observe that the same student scores 70 marks in Mathematics in Class X quarterly examination, thereafter 50 marks in half-yearly examination, and 30 marks in pre-board examination. In this case, there is a need for diagnosing the causes of securing lower marks in subsequent examinations in Mathematics and accordingly provide remedial teaching to the student.

Those tests which are devised to diagnose learning difficulties are considered as remedial tests. In most of the cases the remedial tests are subjective and individual oriented. If the problem is widespread among the group of students in a class or a region then the group remedial test can be used.

You will study the details of diagnostic testing and remedial instruction in the Unit-12, Block-3 of this Course.

7.3.6 Intelligence Test

The term 'intelligence' is difficult to define in a single sentence acceptable to all. But intelligence can be understood as a general set of mental traits, which is often reported as the mental abilities. Test of mental ability or intelligence tests measure convergent thinking (Singh, 2002). Convergent thinking is a process of finding out the solution of a problem. Guilford defines convergent thinking is the ability to give correct answer to standard questions that do not require creativity. For instance, most of the school based tasks can be done through different tests. Many psychologists and educationists opined that intelligence is a product of heredity and environment, but it is a matter of debate to get the real contribution of heredity and environment for intelligence of an individual. Many researches you will find in this regard. In this section, our intension is to understand the measurement of intelligence and the tests used for measuring intelligence.

Intelligence tests are used to provide a very general measure, usually reporting a global test score. As they are general, intelligence tests are useful in predicting a wide variety of tasks. Intelligence tests can be classified on various backgrounds.

On the basis of administration it can be classified into two categories :

a) Individual tests

b) Group tests

Let us talk about each category.

Individual intelligence test is one which can be administered on one person at a time. The first individual intelligence test was the Binet-Simon scale (1904). Individual test requires a highly skilled and experienced test administrator because he/she has to give exact instructions to the students in different sections and sub-sections of the test.

The group intelligence test is one which can be administered to more than one person at a time, that is, it can be administered on a group. The Army Alpha Test, 1917 (Verbal) and Army Beta Test (non-language) were the first group intelligence tests. They are generally used in educational setting for adolescents

and adults. At present you will find many such standardized intelligence tests available in different departments such as psychology and education of various institutions for use in schools, for recruiting personnel in different jobs and also for using in masters and doctoral researches.

Intelligence tests can also be classified on the basis of the nature of items included in the tests such as verbal and non-verbal tests. In verbal tests, the instructions and items are presented to the student in written language, whereas, in non-verbal tests, items are given in the form of figures, charts, diagrams, etc. Let us now discuss the measurement of intelligence, that is intelligence quotient.

Intelligence Quotient (IQ) : Intelligence of an individual can be measured by using the following intelligence quotient (IQ) formula.

$$\text{I.Q.} = \frac{\text{Mental Age}}{\text{Chronological Age}} \times 100$$

Chronological age : The exact age of the individual whose intelligence will be measured. This can be calculated counting the age from the date of birth of an individual. Chronological age of a child never varies.

Mental age : Mental age is calculated by comparing a child's obtained score from the intelligence test with the average score at his/her age level. For example, if average score in an intelligence test for six years age group children is 70, and a child of this age group scored 70, then it will be considered that the mental age of that child is 7. In case, the same child scored the average score of another intelligence test meant for the age group of eight years, then the child's mental age will be 8. Therefore, mental age of a child may vary from the chronological age. In different cases it may be same as chronological age or above or below of it.

Let us discuss it with an example :

Let the chronological age of a child is nine (CA – 9) and mental age is 10 (MA-10).

$$\text{I.Q.} = \frac{\text{MA}}{\text{CA}} \times 100$$

$$= \frac{10}{9} \times 100$$

$$= 111.11$$

To interpret the IQ score, you can refer the intelligence table as follows :

Table 7.1: IQ Reference Table

IQ score range	Category
Under 70	Feeble mindedness
70-79	Border line
80-89	Dullness
90-109	Average or normal
110-119	Superior
120-140	Very superior
Over 140	Genius or almost genius

(***Source** : https://www.iqtestforfree.net/iq-scale.html, retrieved on 10.8.2017*)

As per the above example, you can say that the child having the IQ of 111.11 is in superior category.

Check Your Progress 1

Note : a) Write your answers in the space given below.

   b) Compare your answers with those given at the end of the unit.

1. What do you mean by a test?

   ....

   ....

   ....

2. List different types of paper-pencil tests?

   ....

   ....

   ....

3. What do you mean by oral tests?

   ....

   ....

   ....

4. For whom are the oral tests more effective?

   ....

   ....

   ....

5. What is the purpose of diagnostic tests?

   ....

   ....

   ....

6. Define the concept of mental age.

   ....

   ....

   ....

7.4   TOOLS

In the preceding section, you have studied about the various tests. In this section, you will study about various tools including the scales, schedules and inventories. To measure the scholastic abilities of the students, we usually use the tests such as achievement, intelligence, aptitude, etc., but for measuring most of the co-scholastic abilities, we use the tools such as rating scale, inventories, checklists, schedules, questionnaires, etc. In simple language, you can understand a tool as an instrument to measure something. In education, the word tool is used to measure various traits of students. Now, let us discuss each tool with details.

7.4.1 Rating Scales

Rating scale is one of the important tools widely used in psychology and education. It is used for assessing attitudes of students on any situation, idea, object, character, person or an attribute. In a rating scale, the opinions are given in various degrees such as strongly agree, agree, and disagree; highly satisfied, satisfied, and dissatisfied, etc. A rating scale is prepared always in odd number points like 3-point scale, 5-point scale, 7-point scale, or 9-point scale. It is in odd number points because, a definite middle measuring point will be possible only when the scale is odd points. You might have observed that many persons have the attitude to opine their opinions at the middle rating. Rating scales can be presented in different categories. The most commonly used categories of rating scales used in the schools are as follows :

(i)   Numerical scale

(ii) Graphic scale

Let us discuss the above mentioned scales briefly.

i)   Numerical scale : In this type of rating scale, numbers are assigned in different degrees. It is therefore called numerical rating scale. In a numerical scale a sequence of definite numbers is supplied to an observer. The assigned numbers against the item carry certain meaning/interpretation. See the example given below :

Question : How do you consider Sudhir's behaviour in your class? (in this example, Sudhir is a student)

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

Instead of nine-point scale you can have 3-point, 5-point or 7-point scale. In a 5-point scale you can have the following characteristics:

5   Most pleasant

4   Moderately pleasant

3   Indifferent

2   Moderately unpleasant

1   Most unpleasant

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 favor of other types of scales because it is believed that they suffer from various biases and errors.

ii)   Graphic Scale : The graphic scale is a popular and most widely used rating scale. In this scale, a straight line is shown vertically and horizontally

with various clues to help the rater. The line is either segmented into units or continuous. If the line is segmented then the number of units can vary from case to case. Let us try to understand from the example given below :

Question : How effective was the teacher in the class?

1	2	3	4	5
Very Effective	Moderately Effective	Average	Moderately Ineffective	Very Ineffective

Question : Semester system of education encourages learners to engage in their studies regularly.

1	2	3	4	5
Strongly agree	Agree	Undecided	Disagree	Strongly disagree

The above two examples are presented in 5-point rating scale. Scoring to this type of items are done as per the instructions given in the test. In case it is a positive item, the highest positive degrees of the response are given highest points and for the negative items lower points. For example, in a positive item, 5 points can be given for 'strongly agree', 4 to 'agree', 3 to 'undecided', 2 to 'disagree', and 1 to 'strongly disagree' in a 5-point scale. This is vice-versa in case of negative items.

Limitations of rating scales : The rating scales have certain limitations. Some of them are discussed as follows :

- **Error of leniency:** There is a constant tendency among the raters to rate those whom they know well or in whom they are closely involved, higher than they should. Such raters are called 'easy rater'. Some raters become aware of the feeling 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.

- **Error of central tendency:** Most of the raters hesitate to rate the individuals on the extremes of the scale, instead they follow the middle of the scale. Hence the result gets distorted.

- **Halo effect:** It is an error which obscures the cluster of traits within an individual. The rater develops a general opinion about the person's merit and his ratings on specific traits and is greatly influenced by the impression. It results in a false positive co-relation among the traits which are rated.

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

- **The contrast error:** It is due to a tendency of a rater to rate others in the opposite direction (contrasting) from himself in a trait.

- **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.

(**Source** : ES-333, B.Ed., IGNOU, 2010)

7.4.2 Questionnaire

A questionnaire is a device comprising a series of questions dealing with some psychological, social, and educational topic(s) given to an individual or a group of individuals, with the object of obtaining data with regard to some problems under investigation. (Koul,1997) Questionnaire is a common tool that we, the teachers usually apply to collect data from a situation, condition or prevailing practice.

The questionnaire can be classified in terms of the nature of the questions constructed to gather the data from the stakeholders. It may be closed or open ended. *Closed questionnaires* are those where the respondents answer in limited way, like responding in 'yes' or 'no'; underlining the replied among the pre-defined responses, putting the sign 'correct' or 'incorrect'. Whereas in *open questionnaires* the respondents are free to share, clarify and put their view. Let us discuss certain examples of closed and open ended questions.

Examples of open-ended questions :

● Describe any one of the wars fought for independence in India.

● Write an essay on Quit India Movement.

Examples of closed-ended questions :

● In which year Quit India Movement took place?

   a) 1941   b) 1942   c) 1943   d) 1944

According to other classification, questionnaire may be Structured or Non-Structured. A structured questionnaire contains definite, concrete and directed questions, whereas non-structured questionnaires are often used as the interview guide which is non-directive. It is an important instrument being used to gather information from widely scattered sources and also used when factual information is desired. (IGNOU, 2007)

Advantages of questionnaires : Some of the advantages of questionnaires are enumerated as follows :

● The responses are gathered in a standardized way.

● Questionnaires are more objective, certainly more so than interviews.

● Relatively quick to collect information.

● Potential information can be collected from a large portion of a group.

Limitations of questionnaires : Although the questionnaire is one of the widely used tools but it has also certain limitations which can be enumerated as follows :

● Not easy to be used with children or illiterates.

● Respondents may not agree to respond in writing.

● Sometimes it is difficult to construct questions on complex and crucial topics.

● Respondent interprets the questions from his/her angle of understanding.

● External factors may affect the response.

7.4.3 Inventories

An inventory is constructed in the form of a questionnaire. But the inventories are more exhaustive than questionnaire. Inventories have been mostly used for measuring personality traits, interests, values and adjustments i.e. for assessing self-reporting affective domain of behaviour. It consists of a series of questions or statements to which the subject responds to by answering 'Yes' or 'No',

or 'Agree' or 'Disagree'. This can also be answered in some similar ways to indicate preferences or to make those items that describes the subject's typical behaviour.

In the inventory, the statements are put in first person, For example, "I think I am comparatively more tense than others". In the questionnaire, there is a question in a second person, for example "Do you think you are more tense than other persons around you?"

The term inventory is also used in another sense. That is, an inventory of various items is available in an office room or inventory of clothes to be given to the washer man, etc, but in the educational evaluation, in a school and classroom situation, inventories are used for the aforementioned concepts.

Guidelines for the administration of the inventory are as follows

● The teacher should explain printed instructions very clearly to the learners.

● The teachers should make it clear to the learners' that the data will be kept confidential.

● The teacher should remove doubts, if any, regarding the manner of filling the inventory or the questionnaire.

● The teacher should take all other timely precautions for preparing a state of mind conducive to response.

7.4.4   Checklist

According to Koul (1997), a checklist is a simple device consisting of a prepared list of items which are prepared by the teacher to be relevant to the problem being studied. After each item a space is provided for the observer to indicate the presence or absence of the item by checking 'yes' or 'no', or a type of number of items may be indicated by inserting the appropriate word or number.

The checklist is a systematic and quick way to gather data of the relevant factors and take actions accordingly. A very simple example is being given below :

For example :

Is the school building fire proof?   Yes/No

Does the school follow rain harvesting policy?   Yes/No

Is the school building earthquake proof?   Yes/No

It is noteworthy that the responses collected on the checklist are as a matter of fact not any judgment. It is a good tool in gathering facts for educational survey, checking your school library, laboratory, game facilities, school buildings, textbooks, etc. It may also be used to check the availability of other facilities in your school.

Checklists are also applied to classroom instructional activities such as studying the working habits of students, supervising classroom instructions, teacher –pupil relationships, etc.

7.4.5   Interview Schedule

Interview is a communication or conversation by which a person asks interviewer and interviewee responds verbally in the face-to-face situation. An interview can also be conducted through skype electronic media. It can be conducted through telephonic conversation or through by using internet. It is a common technique for collecting required information about an individual. You have learnt about interview technique in Unit-5 of this Block. The interview schedule is a tool

with the help of which the interview is conducted.

Interview schedule can be classified according to the purpose for which it is structured and used. If it is to resolve the research hypothesis, it is a research interview schedule and if it is for clinical purpose, it a clinical interview schedule.

On the bases of the structure, the interview schedules are categorized as structured or unstructured.

● **A structured** interview schedule is one in which the procedure to be followed is standardized and it is determined in advance of the interview. The same types of questions are presented to the interviewee and the wordings of the instructions to the interviewee are specified.

● In the **unstructured interview** schedule**,** the series of questions are decided in advance but the interviewer is largely free to reorganize the questions and timing to attain the objective of the interview.

7.4.6   Observation Schedule

Observation has been the first practice of assessment that we do in our classroom. Each observed incident, expression and reaction has useful data for teachers, hence the observation is an effective tool for us. Observation is the process in which one or more persons observe what is happening in a real-life situation, and he/she classifies and records pertinent happening according to pre-planned scheme. It is used to evaluate the overt behaviour use uniformly either American or British English i.e., spellings of words, etc. of individuals in both the controlled and uncontrolled situations. (Koul, 1997, p.170)

The observation schedules are the enumerations, listing of the facts or other data that are observed under observation process. Like questionnaires, the observation schedules are also classified as structured or unstructured. This can also be classified as participant and non-participant observation. You have already studied various types of observation techniques in Unit-5 of this Block. In case the things to be observed are properly defined, the style of recording the observed information, standardized conditions of observation and the selection of pertinent data of observation then it is a structured observations (Kothari & Garg, 2014, p.91) and the schedule is a structured observation schedule. On the other hand when these are not pre-determined aspects of observation, such observation schedules are unstructured schedules.

Merits of observation : The following are some of the merits of observation

● Through observation the data are gathered from a natural setup.

● The data observed are direct or first hand.

● As the data are first hand or direct, while doing observation we can correlate what is being said and what is being shown.

● Does not rely upon people's willingness or ability to provide information.

Demerits of observation : The observation has also the following demerits :

● The observation if not done with planning then it will not bring out authentic information.

● Observer's biases may affect the result.

● There may come '*Hawthorne effect*', that is when the person gets to know that he/she is being observed then the real problems may not be shared or shown.

● Does not develop the understanding why people behave in a particular manner.

Table 7.2: Strengths and weaknesses of various tools

Tool	Strengths	Weaknesses
*Tests**	- Economical - Standard Questions - Commercial tests Strong in Technical Qualities - Objective Tests Easy to Score - Standardized tests provide Uniform Procedure for all subjects and standards	- Norms may be inappropriate - Standardized tests may be too broad and general - Standard Score may distort differences - Standardized tests may give false sense of validity - Locally developed tasks often technically weak test anxiety - Restricted to subjects who can read and write
Scales	- Consumes less time to administer - Interesting to rater - Can be used by little training - Wide range of Application	- Consumes time to construct - Items if not made valid data can be misleading - More than three point scale are difficult for some raters to answer
*Question-naires**	- Economical - Can be anonymous - Standard Questions & Uniform Procedure - Usually easy to score - Provides time for subject to think about responses	- Response rate of mailed questionnaire - Inability to probe and clarify - Scoring open ended items - Faking and social desirability - Restricted to subjects who can read or write - Biased or ambiguous terms - Response set
Inventories	- Exhaustive - Appropriate for assessing self reporting affective behaviour. - Used when quick overview is required	- Can be only used when required to enlist things. - Lacks explanations. - Ill-preparation may lack essential ingredients of inventories.
Checklists	- Simple Device - Easy to Use - Quick collection of data - Systematic	- Limited in response - Restricted to one aspect - Factual - Useful when used in support of other tools
*Interviews**	- Flexible - Adaptable - Ability to probe and clarify - Ability to include non verbal behaviour - High response rate - Used with non readers	- Costly - Time consuming - Interviewer bias - Not anonymous - Subject effects - Effect of interviewer characteristics - Requires training - Leading questions
*Observation Schedules**	- Captures natural behaviour - Mitigate social desirability, response set and subject effects - Relatively unobtrusive - Reliable for low interface observations	- Costly - Time consuming - Effect of observer on subject - Observer bias - Requires training - Reliability difficult for complex behavior and high inference observations - Inability to probe and clarify - Usually not anonymous - Interpretations of high inference observations

(***Source** : McMillan & Schumacher, 2006)**

Table 7.2 summarizes the strengths and weaknesses of tests and tools. All tools have their strengths and also weaknesses. A single tool can not be suitable to use for all types of measurement. The data gathered through a checklist can not be possible by using a scale and accordingly, data received through interview is also somehow different from observation.

Check Your Progress 2

Note : a) Write your answers in the space given below.

b) Compare your answers with those given at the end of the unit.

7. What is a numerical rating scale?

....

....

....

8. How do the questionnaires differ from inventories?

....

....

....

9. Give two dimensions of a school where checklist can be used to collect information.

....

....

....

10. How is an unstructured interview schedule better than structured one?

....

....

....

11. Write any two demerits of an observation schedule.

....

....

....

7.4.7 Anecdotal Records

Every teacher (also parents, friends, relatives, etc.) observes his/her students on the day-to-day basis and those observations can be formally recorded to his own understanding or to that of others who will later deal with the students. Such reports of a student recorded by informal observations in the form of anecdotes (not only of present, but past too, that is childhood, nursery and pre-school days, etc.) by parents, siblings, friends and peer group are called anecdotal records. (IGNOU, 2010)

The anecdotal records provide a means for the observation of one's behaviour. A record of events is one dependable data that will remain unchanged from

the time it is made until the time we want to refer to it. A set of such records provides stable evidence on which later appraisals can be raised. Making a record of an observation of a child's behaviour, a prompt record, while the behaviour is still fresh in the mind, is dependable as it eliminates the limitations and distortions of memory. Such a record can be used, with practice, to provide a relatively direct and objective report of actions, with reactions of some observers as well.

These records serve two purposes. The first purpose may be to give you practice in studying students with a view to develop their understanding and increasing their sympathetic insight. Second purpose is that it provides an informal and usually qualitative picture of certain aspects of an individual's behaviour. It is a primary aspect of social functioning or adjustment to personal problems that one hopes to illuminate by records of incidents of school behaviour. The interaction of a child with the other children in a room, aggression or withdrawal, events that throw light on the child's role in the group and his reaction to it are good materials for records. Indications of personnel tensions and adaptation to them, habitual mood and temper or special crises and adjustment are worth recording.

Anecdotal records should be factual report of an event in a child's life reported with adequate details so that it is a meaningful evidence of his/her behaviour.

A good Anecdotal record has the following features:

- It provides an accurate description of any event,

- It describes a setting sufficiently to give the events meaning.

- If it includes interpretations or evaluation by the recorder, this interpretation is separated from the description and its different status is clearly identified.

- The event relates to the child's personal development or social interactions.

(***Source :*** *Anecdotal Record, Unit 7, ES-333, IGNOU, 2010*)

7.4.8 Learners Portfolios and Rubrics

You have already studied about portfolio in Unit-5 of this Block. We will discuss in this section some more details about portfolio as an evaluation tool. A portfolio is a purposeful, systematic collection and evaluation of student work that documents progress towards meeting learning objectives. Portfolios have been used for years in the various fields such as architecture, art and journalism as the primary method of evaluating performance. In the field of education, portfolios are being used with increasing frequency, especially with assessment of reading and writing skills. (McMillan &Schumacher, 2006, pp.193) A portfolio (a collection of the student's work in an area, showing growth, self-reflection and achievement) is a systematic collection of work, often including work in progress, revisions, student self analysis and reflections on what the student has learned. (Popham, 2008)

Portfolio comprises artistic pieces or written work of the student. The student might also include letters in the portfolios describing each entry and its importance. They may include graphs, diagrams, pictures or digital slideshows, power point presentations, etc. They may also include recordings of the student reading, their work, unedited and final drafts of essays or poem, list of books

and references to a particular theme, peer comments, videotapes, laboratory reports and computer programmes. Anything that demonstrates learning in the area being taught and assessed can also be included. (Popham, 2008)

Few common portfolios and their purposes are as follows :

(i) Growth portfolios

- To show growth or change over time

- To help develop process skills such as self-evaluation and goal-setting

- To identify strengths and weaknesses

- To track the development

(iii) Showcase portfolios

- To showcase end-of-year/semester/trimester accomplishments of the students

- To prepare a sample of best work by the students of the school

- To showcase student perceptions of favorite, best or most important works

- To communicate a student's current aptitudes to future endeavours

(iv) Evaluation portfolios

- To document achievement of students' for grading purposes

- To document progress towards standards set by the school

- To place students appropriately in a particular section or community of learning group

Creating portfolios : Students should be involved in selecting the pieces that will make up the portfolios. The criteria for creating a portfolio is as follows :

- Ask students to include a rationale of their selection.

- Have each student write a 'guide' to his or her portfolio, explaining how strengths and weaknesses are reflected in their work.

- Include self and peer critique, indicating specifically what is good and what needs improvement.

- Model self criticism of your own production.

To evaluate these portfolios, scoring rubrics are used that actually determine the quality of students' performance.

*(**Source** : Woolfolk et.al. 2012)*

Rubric

A rubric is an assessment tool for the teachers to assess performances of the students in different areas with fixing certain criteria of assessment. To measure students' performance based on a pre-determined set of criteria, a rubric or scoring scale is typically created which contains the essential criteria for the task and appropriate levels of performance for each criterion.

A rubric comprises two components: criteria (vertical column) and levels of performance (horizontal column). Any rubric has at least two criteria and at least two levels of performance (which may be added as per the requirement of the teacher). The criteria for the performance on a task, are listed in the left-hand column in the rubric below (concepts, facts and figures, organization/presentation and references) and the performance is marked in three levels-poor, average and good. Weightage to each criteria is also given in the second column (Refer Table 7.3).

Table 7.3: Students' performance rubrics

Criteria	Weightages to the criteria	Performance		
		Poor 1	Average 2	Good 3
Concepts	X2			
Facts and figures	X3			
Organization and Presentation	X3			
References	X1			

Finally, the rubric above needs a mechanism of scoring. The weigthage to the above criteria are : two times to concept (X2); three times to facts and figures (X3); three times to organization and presentation (X3); and one time to references are (X1). Accordingly, points for the criteria 'concept' will be: poor–2; average – 4; and good – 6. Scoring to other criteria can be done accordingly.

Criteria	Weightages to the criteria	Performance		
		Poor 1	Average 2	Good 3
Concepts	2-times	2	4	6
Facts and figures	3-times	3	6	9
Organization and Presentation	3-times	3	6	9
References	1-time	1	2	3
Total		**9**	**18**	**27**

Accordingly, score of each student may be calculated. You can also use many other methods to score the rubrics.

Activity 1

Cite an example of an activity that you would like to conduct among the students to whom you teach and prepare a rubric for its assessment. Also specify the scoring procedure.

....

....

....

7.5 LET US SUM UP

In this Unit, you have studied about the tests and tools which can be used in schools easily. We discussed the tests, their types such as paper-pencil tests, oral tests, etc. We also discussed about commonly used tests in the school system such as aptitude tests, achievement tests, diagnostic-remedial tests and intelligence tests. You have also studied when to use the aptitude test and how the achievement tests have become the most used tests in our schools.

You also studied about the different tools like rating scales, questionnaires, inventories, checklists, interview and observation schedules, anecdotal records, portfolios and rubrics and their uses and limitations.

The contents and discussions covered in this Unit will help you to conceptualise and understand Unit-9, 10 and 11 presented in Block-3 of the same Course. The details about developing achievement test, diagnostic remedial test, and commonly used tests in schools have been elaborately discussed in Block-3 of this Course.

7.6 REFERENCES AND SUGGESTED READINGS

Anastasi, A. (1968). Psychological Testing. London: MacMillan Company.

Binet, A. (1916). The Development of Intelligence in Children (translated by E. S. Kite). Baltimore: Warmick and York.

Brody, N. (1976). Intellegence , New York: Academic Press.INC

English, H.B. & A.C. English (1958). A comprehensive Dictionary of psychological and Psychoanalytical Terms. New York: David McKay Company, Inc.

Freeman, S.F (1962). Theory and Practice of Psychological Testing (3rd ed.), New Delhi, Oxford & IBH Publishing Co. Pvt. Ltd.

Gronlund, Norman E. and Robert L. Linn. (1990). Measurement and Evaluation in Teaching. 6th ed. New York, Macmillan, p. 5.

IGNOU (2010). B.Ed. Course ES-333, 'Educational Evaluation' Unit 7, Types of Tools, IGNOU: New Delhi.

IGNOU (2007). *Tools and Techniques of Data Collection* in MES-016-Educational Research : Research Design. New Delhi: School of Education, IGNOU.

Kazdin, Alan E. (2000). Encyclopedia of psychology. Washington, D.C., American Psychological Association, vol. 8, p. 47-48.

Kothari, C.R. & Garg, G. (2014). Research Methodology-Methods and Techniques, New Delhi: New Age International Publishers.

Koul, L. (1997). Methodology of Educational Research, Delhi: Vikas Publications.

Popham,W.J. (2008). Transformative Assessment. USA:ASCD

Singh, A.K. (2002). *Test Construction* in Tests, Measurements and Research Methods in Behavioural Sciences. Patna: Bharti Bhawan Publishers.

Sternberg, R.J. (1985). Beyond IQ- A triarchic Thoery of intelligence , Cambridge: Cambridge University Press

Woolfolk, A., Mishra, G. & Jha, A. K. (2012). *Classroom Assessment, Grading and Standardized Testing* in Fundamentals of Educational Psychology, Delhi: Pearson.

Website Referred

http://busyteacher.org/14344-6-assessment-testing-strategies-esl-classroom.html dated 14/-5/16

www.icbl.hw.ac.uk/ltdi/cookbook/info_**questionnaires**/printable.**pdf**

http://www.elon.edu/students/portfolios/dated 10/07/16

http://jfmueller.faculty.noctrl.edu/toolbox/rubrics.htm dated 14/07/2016

https://www.iqtestforfree.net/iq-scale.html, retrieved on 10.8.2017

7.7 ANSWERS TO CHECK YOUR PROGRESS

1. A test is a standardized instrument designed to measure objectively one or more aspects of a total personality of a person through the sample of verbal and/or non-verbal responses.

2. Achievement, aptitude, subject-wise criterion test, intelligence test, etc.

3. The oral tests are those tests in which the response, solution or the task requires oral response to answer the question.

4. Oral tests are effective for students with language disabilities, illiterates, shy feeling and young children.

5. The main purpose of diagnostic test is to identify the problems faced by the students in achieving the learning objectives and to suggest remedial programmes for them.

6. Mental age is calculated by comparing child's obtained score from the intelligence test with the average score at his/her age group.

7. In numerical rating scale, numbers are assigned to different degrees for responding to an item in the rating scale.

8. In questionnaires, the questions are written in second person whereas the questions in inventories are written in first person.

9. School library, laboratory, game facilities, school buildings, textbooks, etc.

10. Unstructured interviews are focused, in depth and non directive hence more data can be gathered on the problem under investigation, whereas in the case of structured interview options of answers are given and it is also restricted.

11. Observer biases and Hawthorne effect.