UNIT 3:TECHNIQUES OF EVALUATION.

Essay Test

The essay tests are still commonly used tools of evaluation, despite the increasingly wider applicability of the short answer and objective type questions.

There are certain outcomes of learning (e.g., organising, summarising, integrating ideas and expressing in one's own way) which cannot be satisfactorily measured through objective type tests. The importance of essay tests lies in the measurement of such instructional outcomes.

An essay test may give full freedom to the students to write any number of pages. The required response may vary in length. An essay type question requires the pupil to plan his own answer and to explain it in his own words. The pupil exercises considerable freedom to select, organise and present his ideas. Essay type tests provide a better indication of pupil's real achievement in learning. The answers provide a clue to nature and quality of the pupil's thought process.

That is, we can assess how the pupil presents his ideas (whether his manner of presentation is coherent, logical and systematic) and how he concludes. In other words, the answer of the pupil reveals the structure, dynamics and functioning of pupil's mental life.

The essay questions are generally thought to be the traditional type of questions which demand lengthy answers. They are not amenable to objective scoring as they give scope for halo-effect, inter-examiner variability and intra-examiner variability in scoring.

Types of Essay Test:

There can be many types of essay tests:

Some of these are given below with examples from different subjects:

1. Selective Recall.

e.g. What was the religious policy of Akbar?

2. Evaluative Recall.

e.g. Why did the First War of Independence in 1857 fail?

3. Comparison of two things—on a single designated basis.

e.g. Compare the contributions made by Dalton and Bohr to Atomic theory.

4. Comparison of two things—in general.

e.g. Compare Early Vedic Age with the Later Vedic Age.

5. Decision—for or against.

e.g. Which type of examination do you think is more reliable? Oral or Written. Why?

6. Causes or effects.

e.g. Discuss the effects of environmental pollution on our lives.

7. Explanation of the use or exact meaning of some phrase in a passage or a sentence.

e.g., Joint Stock Company is an artificial person. Explain 'artificial person' bringing out the concepts of Joint Stock Company.

8. Summary of some unit of the text or of some article.

9. Analysis

e.g. What was the role played by Mahatma Gandhi in India's freedom struggle?

10. Statement of relationship.

e.g. Why is knowledge of Botany helpful in studying agriculture?

11. Illustration or examples (your own) of principles in science, language, etc.

e.g. Illustrate the correct use of subject-verb position in an interrogative sentence.

12. Classification.

e.g. Classify the following into Physical change and Chemical change with explanation. Water changes to vapour; Sulphuric Acid and

SodiumHydroxide react to produce Sodium Sulphate and Water; Rusting of Iron; Melting of Ice.

13. Application of rules or principles in given situations.

e.g. If you sat halfway between the middle and one end of a sea-saw, would a person sitting on the other end have to be heavier or lighter than you in order to make the sea-saw balance in the middle. Why?

14. Discussion.

e.g. Partnership is a relationship between persons who have agreed to share the profits of a business carried on by all or any of them acting for all. Discuss the essentials of partnership on the basis of this partnership.

15. Criticism—as to the adequacy, correctness, or relevance—of a printed statement or a classmate's answer to a question on the lesson.

e.g. What is the wrong with the following statement?

The Prime Minister is the sovereign Head of State in India.

16. Outline.

e.g. Outline the steps required in computing the compound interest if the principal amount, rate of interest and time period are given as P, R and T respectively.

17. Reorganization of facts.

e.g. The student is asked to interview some persons and find out their opinion on the role of UN in world peace. In the light of data thus collected he/she can reorganise what is given in the text book.

18. Formulation of questions-problems and questions raised.

e.g. After reading a lesson the pupils are asked to raise related problems-questions.

19. New methods of procedure

e.g. Can you solve this mathematical problem by using another method?

Advantages of the Essay Tests:

1. It is relatively easier to prepare and administer a six-question extended-response essay test than to prepare and administer a comparable 60-item multiple-choice test items.

2. It is the only means that can assess an examinee's ability to organise and present his ideas in a logical and coherent fashion.

3. It can be successfully employed for practically all the school subjects.

4. Some of the objectives such as ability to organise idea effectively, ability to criticise or justify a statement, ability to interpret, etc., can be best measured by this type of test.

5. Logical thinking and critical reasoning, systematic presentation, etc. can be best developed by this type of test.

6. It helps to induce good study habits such as making outlines and summaries, organising the arguments for and against, etc.

7. The students can show their initiative, the originality of their thought and the fertility of their imagination as they are permitted freedom of response.

8. The responses of the students need not be completely right or wrong. All degrees of comprehensiveness and accuracy are possible.

9. It largely eliminates guessing.

10. They are valuable in testing the functional knowledge and power of expression of the pupil.

Limitations of Essay Tests:

1. One of the serious limitations of the essay tests is that these tests do not give scope for larger sampling of the content. You cannot sample the course content so well with six lengthy essay questions as you can with 60 multiple-choice test items.

2. Such tests encourage selective reading and emphasise cramming.

3. Moreover, scoring may be affected by spelling, good handwriting, coloured ink, neatness, grammar, length of the answer, etc.

4. The long-answer type questions are less valid and less reliable, and as such they have little predictive value.

5. It requires an excessive time on the part of students to write; while assessing, reading essays is very time-consuming and laborious.

6. It can be assessed only by a teacher or competent professionals.

7. Improper and ambiguous wording handicaps both the students and valuers.

8. Mood of the examiner affects the scoring of answer scripts.

9. There is halo effect-biased judgement by previous impressions.

10. The scores may be affected by his personal bias or partiality for a particular point of view, his way of understanding the question, his weightage to different aspect of the answer, favouritism and nepotism, etc.

Thus, the potential disadvantages of essay type questions are:

(i) Poor predictive validity,

(ii) Limited content sampling,

(iii) Scores unreliability, and

(iv) Scoring constraints.

Suggestions for Improving Essay Tests:

The teacher can sometimes, through essay tests, gain improved insight into a student's abilities, difficulties and ways of thinking and thus have a basis for guiding his/her learning.

(A) While Framing Questions:

1. Give adequate time and thought to the preparation of essay questions, so that they can be re-examined, revised and edited before they are used. This would increase the validity of the test.

2. The item should be so written that it will elicit the type of behaviour the teacher wants to measure. If one is interested in measuring understanding, he should not ask a question that will elicit an opinion; e.g.,

"What do you think of Buddhism in comparison to Jainism?"

3. Use words which themselves give directions e.g. define, illustrate, outline, select, classify, summarise, etc., instead of discuss, comment, explain, etc.

4. Give specific directions to students to elicit the desired response.

5. Indicate clearly the value of the question and the time suggested for answering it.

6. Do not provide optional questions in an essay test because—

(i) It is difficult to construct questions of equal difficulty;

(ii) Students do not have the ability to select those questions which they will answer best;

(iii) A good student may be penalised because he is challenged by the more difficult and complex questions.

7. Prepare and use a relatively large number of questions requiring short answers rather than just a few questions involving long answers.

8. Do not start essay questions with such words as list, who, what, whether. If we begin the questions with such words, they are likely to be short-answer question and not essay questions, as we have defined the term.

9. Adapt the length of the response and complexity of the question and answer to the maturity level of the students.

10. The wording of the questions should be clear and unambiguous.

11. It should be a power test rather than a speed test. Allow a liberal time limit so that the essay test does not become a test of speed in writing.

12. Supply the necessary training to the students in writing essay tests.

13. Questions should be graded from simple to complex so that all the testees can answer atleast a few questions.

14. Essay questions should provide value points and marking schemes.

(B) While Scoring Questions:

1. Prepare a marking scheme, suggesting the best possible answer and the weightage given to the various points of this model answer. Decide in advance which factors will be considered in evaluating an essay response.

2. While assessing the essay response, one must:

a. Use appropriate methods to minimise bias;

b. Pay attention only to the significant and relevant aspects of the answer;

c. Be careful not to let personal idiosyncrasies affect assessment;

d. Apply a uniform standard to all the papers.

3. The examinee's identity should be concealed from the scorer. By this we can avoid the "halo effect" or "biasness" which may affect the scoring.

4. Check your marking scheme against actual responses.

5. Once the assessment has begun, the standard should not be changed, nor should it vary from paper to paper or reader to reader. Be consistent in your assessment.

6. Grade only one question at a time for all papers. This will help you in minimising the halo effect in becoming thoroughly familiar with just one set of scoring criteria and in concentrating completely on them.

7. The mechanics of expression (legibility, spelling, punctuation, grammar) should be judged separately from what the student writes, i.e. the subject matter content.

8. If possible, have two independent readings of the test and use the average as the final score.

Objective tests

Objective tests began to be used early in the twentieth century as a means of evaluating learning outcomes and predicting future achievement, and their high reliability and predictive validity led to the gradual replacement of the essay test.

One common misconception about the objective test item is that it is limited to testing specific, often trivial, factual details, which would sometimes lead to the use of an essay or performance test to assess students' comprehension of broader principles or their ability to apply them. However, as Robert Ebel pointed out, well written objective tests (especially multiple choice tests) can actually assess such higher-order abilities to some extent. While it is true that some types of knowledge or abilities cannot be assessed by objective tests, educators also should keep in mind that what test items can assess depends largely on the skills and effort of the test constructor, rather the test format per se.

Meaning of Objective Test

An objective test item is defined as one for which the scoring rules are so exhaustive and specific that they do not allow scorers to make subjective inferences or judgements; thereby, any scorer that marks an item following the rules will assign the same test score.

TRUE FALSE

The true/false test is the simplest form of selected response formats. True/false tests are those that ask examinee to select one of the two choices given as possible responses to a test question. The choice is between true and false, yes and no, right and wrong, and so on. A major

advantage of the true/false test is its efficiency as it yields many independent responses per unit of testing time.

Therefore, teachers car cover course material comprehensively in a single test. However, one apparent limitation of the true/false test is its susceptibility to guessing. It should be noted, however, that test givers can attenuate the effects of guessing by increasing the number of items in a test. In addition, guessing might reflect partial knowledge, which would provide a valid indication of achievement.

Matching-Type Items

Design

Matching-type items are prescribed as set of terms, events, phrases, definitions etc., called the premises, which are written on the left-hand side, say column I. Another set of name pictures, statements etc., called the responses, are placed on the right-hand side under column II. Students are asked to match each item with the corresponding response, which is considered as one test item.

Relationship may be between a term and definition, object and its functions, inventor and inventions, author and work, dates and events, problems and solutions etc. If the number of responses are equal to the number of premises, it is termed perfect matching. If the number of responses are more than the premises. it is called imperfect matching. If single criterion for matching the items is used, it is called simple matching; but if more than one criterion is used as the basis of matching, i.e. more than two lists are given, it is termed compound matching. While attempting the matching-type items, the student has more choices (2, 3,4, 5, 6 or even more) to choose from for each item (premises).

The number of choices (responses) goes on decreasing as one goes on attempting one item after another. Thus every matching-type item in a set is indeed a multiple choice-type item.

Multiple-Choice Questions (MCQ)

The most potential and usable form of objective tests is the MCQ. These questions are either used exclusively as in some selection tests, or in combination with other forms of questions as in most of the examinations conducted by the examination boards or other agencies.

If the number of candidates is very large and results are to be declared within a few days, the use of MCQ cannot be overemphasised. Exclusive use of MCQ is warranted in all such situations where 100% scoring objectivity, time constraint, computerisation of results, machine scoring, ranking, postexamination statistical moderation and record of psychometric properties of items are considered the necessary conditions or requirements.

The multiple-choice items are based on response-directed stimulus, in which responses or options may be arranged in different ways. The choice of the correct answer among the given options or responses can be made from the independent set of responses given for each item or from the same set of responses given for different items.

Sentence Completion.

Completion items require a word, short phrase, number, or symbol response. They are sometimes called *fill-in-the-blank questions.* Sometimes a word bank accompanies a set of short-answer exercises, especially for younger children or where recognition (not production) of the word is

important and spelling is not. Use completion items especially when recall of a specific word or term is important for the learning goal. Write effective completion items by following the guidelines below.

- ✓ Write each item to tap into the content and kind of thinking in important learning outcomes.
- ✓ Design items so that students must supply an important word or concept
- ✓ Insert a blank where an answer is expected.

Advantages of Objective Test

The objective tests have all the qualities of a good test, though all the abilities of the students cannot be measured by them.

1. Validity : Objective tests are generally made to measure the knowledge of the students. First, the questions asked pertain to the field, the knowledge of which has to be measured. Second, the number of questions is generally large which is spread out on the entire range of knowledge.

Third, the evaluation of these questions is objective. So these tests are valid.

2. Reliability : These tests are reliable. All the questions contained in them are clear and bear only one meaning, and their answers too are definite. These are very comprehensive, and the measurement done by them is also objective. It is natural for them to be reliable.

3. Objectivity : As has been clarified above, the questions asked in these tests are clear and have a single meaning, their answers are also definite and the evaluator has not any liberty to mark them. Whoever the evaluator may be, the result of such tests is similar. It is evident that these,

tests are objective and therefore reliable.

4. Comprehensiveness : As has been stated in the beginning, small questions are asked in these tests and the answers to these questions are

given in one mark, number or word. From the viewpoint of availability of time, the time to be taken on to write one essay type answer or 4-

6 short-answer-type answers can suffice to think and write the answers of 40-50 objective questions. Thus, these tests are comprehensive in nature.

5.Discriminative : The questions in these tests are spread out on the whole course and are of **Notes** several types. Some questions have to be answered using recall, some have to be answered by knowledge, and some questions have to be answered using logic and prudence; so the classification of the students done on their basis can be reliable.

6. Practicability : Making objective test questions is a difficult task. It takes time to construct so many questions (often 100-100, 200-200). Marking them too takes time. So now in order to save effort, only multiple choice questions are constructed and their answers are marked on an answer-sheet which can be evaluated by computer. Thus, the construction, administration and evaluation of these tests have become practicable.

Limitations of Objective Test

Despite having all the merits of a good test, these tests are not devoid of some shortcomings.

1. Measurement of Cognitive Aspect Only : These tests basically help to measure the cognitive aspect of the students; these are unsuitable for the measurement of the conative and affective aspects.

2. Measurement of Memory Power Only : These tests generally examine the power of memory, the students have to hardly use the higher mental faculties such as logic and thinking.

3. No Measurement of Language skill and Expression Power : These tests do not help in the measurement of language skill and power of

expression of the students, so we cannot rely on them for language tests. It is considered to be the biggest demerit of these tests.

4. Difficulty in Construction of Questions : According to some scholars, it is very difficult to construct the objective questions for these tests. They consider it a demerit of these tests. We are, however, of the view that these questions can be constructed easily by taking a little interest and care. It depends on interest and practice.

5. Administrative Difficulty : Some scholars are of the view that these tests cannot be administered easily; according to them, these tests are expensive and present difficulties in evaluation.

However, our experience is that these tests are less expensive than the essay-type tests and their evaluation can also be done easily. We have already written that these tests can be evaluatedwith the help of computer.

6. Fluke Answers : Some people opine that these tests can be answered in a fluke. It is true, but we have already invented a statistical method to control it, which is called negative marking.

7. Use of Unfair Means : It is often heard that the students can easily copy in these tests. In our view, this demerit has occurred due to inexperienced people. The fact is that if a student copy from a book or notebook, he cannot answer all questions in the prescribed time. Of course, the students can easily copy if someone prepares the answer-sheet for him. But it is not a demerit of the test, but that of the administrators of the test.

Construction of Objective Test Items

Basically, scoring objective test items is easy : It only requires one to follow the scoring rules. However, constructing good objective test items requires much more skill and effort. The first step is to develop a set of test specifications that can serve to guide the selection of test items. A table of specifications (or test blueprint) is a useful tool for this purpose. This tool

is usually a two-way grid that describes content areas to be covered by the test as the row headings and skills and abilities to be developed (i.e., instructional objectives) as the column headings.

After specifying the content and ability covered by the test using the table of specifications, the appropriate test item format is selected for each item. At this point, not only objective test items but also other types of test items—essay test or performance assessment—should be considered, depending on the learning outcomes to be measured.

The next step is to create specific test items. Typically, it is particularly important for objective test items to be written in clear and unambiguous language to allow examinees to demonstrate their attainment of the learning objectives. If complex wording is used, the item simply reflects reading comprehension ability. It is also important for each objective test item to focus on an important aspect of the content area rather than trivial details. Asking trivial details not only makes the test items unnecessarily difficult, it also obscures what the test constructor really wants to measure. Similarly, relatively novel material should be used when creating items that measure understanding or the ability to apply principles. Items created by copying sentences verbatim from a textbook only reflect rote memory, rather than higher order cognitive skills.

Setting of good question papers is not everybody's cup of tea. It requires a long, consistent practice of writing, criticising, editing and improving the items. It demands thorough knowledge of the criteria of a good question paper (test), i.e. the technical qualities like validity, reliability and usability, which make an evaluation instrument dependable for ranking, selecting and certification of students' achievement. Thorough understanding of technology of paper setting involves preparation of a functional design of question paper that provides framework for setting the question paper. Such a framework includes identification of content units or areas of syllabus, assessment objectives to be included for testing, form of questions to be used, scheme of sections if any, and difficulty level of question type of options to be provided, besides taking cognizance of some predisposing factors like total time, total marks, number of papers, assumption to be made etc. Giving proportional weightage to various content areas, objectives and forms of questions is also necessary in designing. The next step is development of blueprint conforming to the prescribed design by the board or the examining agency. This blueprint provides distribution of questions over various content units in relation to assessment objective, along with indication of marks. Then it is necessary to frame the questions in conformity with the blueprint requirement, along with key, outline answers and marking scheme. Consolidation of questions to give question paper a particular format and organisational pattern is followed.

Writing of general instructions and directions for various parts comes next. Editing and review of questions is undertaken by means of question-wise analysis, on whose basis the needed modifications and improvements in questions are made and paper is finalised to make it press-worthy. However, for a final look an expert review is needed to ensure further improvements through moderation, which is undertaken by different

subject experts and trained evaluators. Sometimes a panel of paper setters and moderators is used to make question papers more valid, reliable and usable through various modifications and improvements made by the moderators, who do a final review to make the question paper perfect. Thus paper setting is a long-drawn process that demands technical know-how, hard labour and critical acumen.

General Tests

This type of test includes those tests which are constructed by the teachers, so these are also called **Teacher-Made Tests.** These tests are mostly used in the field of education, in all weekly, fortnightly, monthly, quarterly, half-yearly and annual examinations. Though attempts are being made to make these tests valid, reliable and objective, yet nothing much can be said with certainty about their validity, reliability and objectivity. Since fully valid, reliable and objective tests cannot be constructed for general tests, so much emphasis is laid on making these tests more and more valid, reliable and objective.

Standardized Tests : This type of tests includes those tests which are constructed by specialists in a specific way. They prepare a draft according to the measurable objectives, then they administer it on different groups of students, and then they delete unnecessary material from these tests and add necessary material to them. Thus, they make them valid, reliable and objective.

Efforts are on to make the achievement tests more and more valid, reliable and objective, but success has not been tasted so far.

But it should not be taken to mean that we should not standardize the achievement tests. We should keep endeavouring in this direction so that they are made more and more valid, reliable and objective.

A standardised test is expertly and carefully planned, constructed and administered; it provides tables of norms for interpretation of scores. It provides for comprehensive coverage of achievements for comparable scores in different areas. Profiles of test scores indicate pupils' general level of achievement and their specific strengths and weaknesses. Subset scores may be quite unreliable for effective use. Standardised achievement test scores are useful in facilitating inspection and in evaluating its results. Standard test scores provide information that contribute to evaluation of teacher competence, but should not be used as the sole basis for this purpose. In general, locally designed testing programmes are more relevant to local needs, although these are less expertly designed than external testing programme; which in turn are criticised for emphasising only verbal and quantitative skills, penalising creative thinkers. Essential functions served by tests of technical or professional competence would be better served if those who develop such tests were better trained end had paid more attention to the problems of test validity.

Most of the, standardised tests are broad survey measure of areas of achievement. Such survey tests are connected primarily with selection, placement, guidance, curricular decision and public policy decisions. Since diagnostic testing is a time-consuming enterprise, their development and use is reserved for the students who face serious problems or weaknesses in some specific area of achievement. At secondary level, centrally produced achievement tests provide evidence of outstanding competence for advanced placement and of minimal competence as a basic outcome of schooling. State-wise programme of testing vacillates between attempt to make objective referenced interpretation of achievements in local schools and global appraisal of effectiveness of different schools and school systems.