

---

## UNIT 2 APPROACHES TO EVALUATION

---

### Structure

- 2.1 Introduction
- 2.2 Objectives
- 2.3 Formative Evaluation
- 2.4 Summative Evaluation
- 2.5 Distinction between Summative and Formative Evaluation
- 2.6 External Evaluation
- 2.7 Internal Evaluation
  - 2.7.1 Advantages and Disadvantages
- 2.8 Internal vs External
- 2.9 Norm-referenced and Criterion-referenced Evaluation
  - 2.9.1 Norm-referenced Evaluation
  - 2.9.2 Criterion-referenced Evaluation
  - 2.9.3 Difference
  - 2.9.4 Relative or Absolute
  - 2.9.5 Comprehensiveness of the Test
  - 2.9.6 Application
- 2.10 Construction of Criterion-referenced Tests
- 2.11 Let Us Sum Up
- 2.12 Unit-end Exercises
- 2.13 Points for Discussion
- 2.14 Answers to Check Your Progress
- 2.15 Suggested Readings

---

### 2.1 INTRODUCTION

---

In Unit I we learnt about the concept of evaluation as well as its need and significance. As practising teachers we need to take evaluation seriously as it is an integral part of the teaching-learning process. As you have already read, there are various approaches in evaluation. We need to understand these approaches in order to choose the appropriate one for our purpose in a given situation. In this unit we will take a look at some of the most significant approaches to evaluation. These approaches are grouped in pairs, namely formative and summative; internal and external; norm-referenced and criterion-referenced. We will discuss in some detail their utility and functions.

---

### 2.2 OBJECTIVES

---

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

- explain normative and summative; external and internal; norm-referenced and criterion-referenced evaluation,
- distinguish between formative and summative; external and internal; norm-referenced and criterion-referenced,

- list the function of all these approaches,
- discuss the advantages and disadvantages of internal evaluation, and
- explain the steps involved in construction of criterion-referenced tests.

---

## 2.3 FORMATIVE EVALUATION

---

Formative evaluation is used to monitor the learning progress of students during the period of instruction. Its main objective is to provide continuous feedback to both teacher and student concerning learning successes and failures while instruction is in process. Feedback to students provides reinforcement of successful learning and identifies the specific learning errors that need correction. Feedback to teacher provides information for modifying instruction and for prescribing group and individual remedial work. Formative evaluation depends on tests, quizzes, homework, classwork, oral questions prepared for each segment of instruction. These are usually mastery tests that provide direct measures of all the intended learning outcomes of the segment. The tests used for formative evaluation are mostly teacher-made. Observational techniques are also useful in monitoring student progress and indentifying learning errors. Since formative evaluation is used for assessing student learning progress during instruction, the results are not used for assigning course grades.

### Check Your Progress 1

- i) What is the most important objective of formative evaluation?

.....  
 .....

- ii) What is formative evaluation not used for?

.....  
 .....

---

## 2.4 SUMMATIVE EVALUATION

---

Summative evaluation is designed to find out the extent to which the instructional objectives have been achieved usually at the end of a terminal period. It is used primarily for assigning course grades or for certifying student mastery of the intended learning outcomes at the end of a particular course programme. The techniques used for summative evaluation are determined by the instructional objectives. For this evaluation, there are external examinations as well as teacher-made tests, ratings etc. Although the main purpose of summative evaluation is assigning grades, it also provides information for judging the appropriateness of the course objectives and the effectiveness of instruction.

### Check Your Progress 2

Give two purposes of summative evaluation.

a) .....

b) .....

---

## 2.5 DISTINCTION BETWEEN SUMMATIVE AND FORMATIVE EVALUATION

---

The terms summative and formative evaluation were for the first time conceptualised by Michael Scriven in his classic (1967) essay on the methodology of evaluation. According to him, Summative evaluation refers to the assessment of worthwhileness of the instructional programme which has already been completed, while formative evaluation refers to the assessment or worth of the instructional programme which is still going on and can still be modified.

A summative evaluator gathers information and judges the merit of overall instructional sequence to retain or adapt that sequence. The audience of summative evaluation is the consumer of the instructional programme in contrast to the formative evaluator whose audience is the designer and the developer of the programme. A formative evaluator is a partisan of the instructional sequence and does everything to make teaching-learning better. A summative evaluator is an uncommitted non-partisan person who is to pass judgement on an instructional endeavour.

A very clear distinction is made between these concepts by Bloom, Hastings and Madaus. Summative evaluation, according to them is, judgemental in nature. Its purpose is to appraise the teaching-learning process and to distinguish it from formative evaluation. It is an end of the course activity concerned with assessment of the larger instructional objectives of a course or a substantial chunk of the course. Our public examinations, annual and term tests are all summative tests used for making summative evaluation. It is a measurement of pupils' achievement and not of their day-to-day improvement. Thus it is a status evaluation of students. The major function is that of grading, promoting or certification of achievement. It may take place at the end of a unit, term or a course of studies. Its emphasis is generally on measurement of cognitive behaviours, sometimes on psychomotor and occasionally on affective behaviours. Instrumentation is limited to final or summative examinations, through a weighted sample of course objectives. The average difficulty level of questions ranges from 35% to 70%. Scoring, though normally norm-referenced, can also be criterion-referenced. Reporting of scores is by objectives. Summative evaluation is thus a judgemental activity focused on certification of students' achievement.

Formative evaluation is developmental, not judgemental in nature. Its purpose is to improve students learning and instruction. Therefore, its major function is feedback to the teacher and students to locate strengths and weaknesses in the teaching-learning process in order to improve it. It operates during instruction and ideally should not be limited to assessment of cognitive behaviours. All classroom assessments which are not used for grading purpose, whether these are unit tests, informal tests, questioning during teaching, home assignments, teacher classroom observations of pupil's responses are examples of formative evaluation. For formative testing, specially designed instruments are devised. As for judgements or scoring, it is criterion-referenced, not norm-referenced as in summative evaluation. Decisions are made to relate to steps to be taken to improve the instructional programme vis-a-vis pupils' learning. Reporting of pupils progress is done in terms of an individual pattern of pass-fail scores on different tasks in the hierarchy of learning outcomes. Formative evaluation is, therefore, a means of determining what the pupils have mastered and what is still to be mastered, thereby indicating the basis for improvement of students learning.

### Check Your Progress 3

Give four important differences between summative and formative evaluation.

- a) .....
- b) .....
- c) .....
- d) .....

---

## 2.6 EXTERNAL EVALUATION

---

When examinations are organised and conducted by an agency other than the institution giving instruction to students, and all the students of a group of institutions come together under the purview of the agency, the assessment carried out is regarded as an external assessment. To this extent, the public examinations in our country as conducted by the appropriate school boards are meant for external assessment. In such Public Examinations, the teachers concerned do not get directly involved in the process of evaluation. Ordinarily such public examinations are not conducted with a specific objective in view. Thus a public examination may be conducted and the results of such an examination utilised for a large number of purposes. Actual testing is restricted to a candidate answering a question paper or a series of question papers each within a set time-limit.

Also, when the exams are conducted by the same institution which teaches but tests are prepared and marked by persons other than the ones who teach the group, the evaluation is called external. Thus when the teacher who teaches is not involved in the evaluation process, the evaluation should be called external.

---

## 2.7 INTERNAL EVALUATION

---

At present, external examinations are equated with only public examinations which become the basis of certification of students. All other examinations which come under the purview of the school are deemed as internal examinations. This notion is based on the criterion of the agency conducting the examinations. From this it follows that examinations within a school, are internal assessments whether conducted by an outside person, or a teacher who does not even teach that particular class or a section. This view does not seem to hold ground because in all these cases the examiner is not connected with the instructional process of students who are being evaluated. The criterion indeed, is the knowledge on the part of the evaluator, of what has been taught and how it has been taught. It is only the class teacher teaching the subject who is aware of this fact. If an examiner is unaware of the unit, the objectives set in advance, as also the learning experiences provided to the students, he/she is not in a position to evaluate them properly. Therefore, internal evaluation is one in which the evaluator is the one who teaches the students. All other types of evaluations, whether done in public or in schools, inside or outside the school, by a teacher of the same school or from a different school, are all external evaluations. Thus the three criteria for an internal evaluation are: direct involvement in the teaching-learning activities of the class, setting of the question paper by the class teacher teaching that class and conduct of the examination vis-a-vis evaluation of scripts by the teacher himself/herself. Where these conditions are not fulfilled, the examination should be called external.

### Check Your Progress 4

Distinguish between internal and external evaluation.

### 2.7.1 Advantages and Disadvantages

The teacher concerned has always been regarded as the best judge of a student. Based on this principle, internal assessment of students by teachers of the institution concerned has been adopted in some institutions in our country and elsewhere. In an internal assessment, it is necessary that a variety of tests are administered and proper statistical weightage given to each and then an over-all grading arrived at, as a true index of the capabilities of a student. Further, there can be room for prejudice, there can be occasions when the student is not in the best of spirits. Taking all such factors into account, it may be necessary to resort to an over-all average of a student's performance over a period of time like two or three years. Thus a senior secondary grading may be based on the overall grading of the performance of the student during his/her last two years at school. There are also other questions relevant to internal assessment. If a class has four divisions and each division is taught by a different teacher, what will be the procedure adopted to realise uniformity in the procedure adopted and the assessment between our subject and another? Is uniformity essential? Such and many other problems exist. Even an internal assessment requires continued research to improve its reliability and validity.

If internal assessments are made for general purposes, it becomes necessary to make the results of one institution comparable with those of others. A way out for realising some uniformity is being attempted. In this attempt techniques are being developed for external moderation. It is felt that internal assessment subjected to external moderation can give results which are as satisfactory or even more satisfactory than those of public examinations. However, other views have also been expressed. It is felt that the results of internal assessment and the results at public examinations with question papers having questions of different types, should both be utilised for furnishing the results of students. In another view the results of internal assessment and of public examination should both be furnished to the candidates separately in the certificates issued to them. The latter view receives support from the fact that an internal assessment should desirably not be limited to test the scholastic characteristics only. In such a case, the teacher's over-all estimate of a student can definitely be different from the estimate obtained through a public examination. Thus internal assessment can be both continuous as well as comprehensive.

#### Check Your Progress 5

Mention one advantage and one disadvantage of internal examination.

a) Advantage

.....

b) Disadvantage

.....

---

## 2.8 INTERNAL Vs EXTERNAL

---

Should the assessment be by internal assessors or external assessors?

This is an area in which most individual teachers do not have much choice. External assessment is the norm in public examinations. These are set and marked by external agents (examination boards) and are mainly intended for external users : university selectors, employers and, in a more general sense, the public and their elected representatives in local and central government. External examinations are seen as validation courses, and their influence on the curriculum as profound.

To a limited extent public examinations also use teachers as assessors. A few teachers are involved as paid employees of examination boards in setting examinations but not necessarily do the teachers assess their own students.

Internal assessment and internal syllabuses do not necessarily go together. It is possible to assess the performance in an external syllabus by internal means. For example, project work in a technology course or computer studies, or field work in geography, may be internally assessed. Similarly, an internal syllabus may be assessed by a local consortium of teachers working on an agreed marking scheme, which, in principle, is a form of external assessment.

Internal assessment of an internally devised syllabus brings great freedom of choice and opportunities to do such things as field work projects, simulations, drama and practical work. It allows teachers to use their intimate and detailed knowledge of the work of individual pupils across a wide range of activities and objectives over a long period of time. There are dangers, not least that the teacher's judgements may be subjective or lack standardisation. The new freedom imposes new disciplines on a teacher and demands familiarity with new assessment techniques.

Internal assessment may be conducted for in-school purposes, for example, to decide who follows which course, or it may be mainly for the information of the teacher and his/her pupil. It is even possible to completely internalise assessment. A pupil's self-assessment may be a purely private affair, or may be fostered in a non-judgemental way by a teacher solely for the benefit of the pupil or for outside agencies like employees and universities. It is possible to make an impartial, disciplined subjective evaluation by a teacher for his/her students and it may be more reliable and valid in testing the instructional objectives.

When internal assessment is used in schools it can do a lot that external assessment cannot. The latter can only be used to test the learning outcomes of the cognitive domain by and large, while internal tests can be used for assessing the learning outcomes in affective and psychomotor domains, too. Since the teacher himself/herself is the assessor, continuous and comprehensive testing is possible and so an evaluation, including in all domains is not difficult.

#### Check Your Progress 6

What do you think is better for our education system internal or external evaluation? Give a reasoned answer in five lines.

.....

.....

.....

.....

.....

## 2.9 NORM-REFERENCED AND CRITERION-REFERENCED EVALUATION

Two alternative approaches to educational testing that must be thoroughly understood are norm-referenced testing and criterion-referenced testing. Although there are similarities between these two approaches to testing, there are also fundamental differences between norm and criterion-referenced testing. Part of those differences hinge on the interpretation of test scores, and part of those differences stem from the way that the two sorts of tests are constructed. Let us probe the nature of these two assessment approaches so as to be able to see the measurement purposes for which each approach is best suited.

There have been disputations about the relative virtues of norm and criterion-referenced measurements for a long time. However, a fundamental fact is recognised by most of the concerned people namely, that norm-referenced and criterion-referenced testing are complementary approaches and are needed to accomplish the full range of purposes necessary in educational measurement.

### 2.9.1 Norm-Referenced Evaluation

Norm-referenced measurement is the traditional class-based assignment of numerals to the attribute being measured. It means that the measurement act relates to some norm, group or a typical performance. It is an attempt to interpret the test results in terms of the performance of a certain group. This group is a norms group because it serves as a referent of norm for making judgements. Test scores are neither interpreted in terms of an individual (self referenced) nor in terms of a standard of performance or a pre-determined acceptable level of achievement called the criterion behaviours (criterion-referenced). The measurement is made in terms of a class or any other norm group as the function is to relate individual measurement to some norm group (Class). The purpose is to produce response variance i.e. to see the extent to which an individual varies or differs from the performance of the group to which he/she belongs or does not belong.

Almost all our classroom tests, public examinations and standardised tests are norm-referenced as they are interpreted in terms of a particular class and judgements are formed with reference to the class which is considered as a type. Who is the most intelligent boy in the class? who stood first? Who got the least marks? Is he better than 5% of the students in the class? These are the kinds of questions which involve norm-referenced judgements. Such judgements use the performance of some group as a referent of the same task. We compare an individual's performance with similar information about the performance of others. That is why selection decisions always depend on norm-referenced judgements. Prediction and many placement decisions also depend on these types of judgements. A major requirement of norm-referenced judgements is that individuals being measured and individuals forming the group or norm, are alike. The conditions under which the referent (norm) was obtained and the conditions under which the original information is obtained, is also assumed to be similar. Another criterion is that the referent used in norm-referenced judgements should have the minimum error so as to have reliable and accurate judgements. Unless the referent used is up-to-date or recent, the comparison of an individual's performance with the group (with an outdated referent) is of no use and can be misleading as it would lead to faulty interpretations. Thus norm-referenced measurement presupposes an up-to-date, reliable referent (norm group) of like individuals obtained from like conditions.

#### Check Your Progress 7

Explain Norm-referenced Evaluation.

.....

.....

.....

.....

### 2.9.2 Criterion-Referenced Evaluation

A criterion-referenced measurement has its origin in the writing of objectives by Mager who urged teachers to specify a criterion of acceptable performance while stating instructional learning outcome, and intended level of proficiency of the learner or a desired standard of performance. Thus in contrast to a norm-referenced measure we can refer to an individual

performance to a predetermined criterion which is well defined. This type of measurement is termed as criterion-referenced measurement. It determines an individual status with reference to well defined criterion behaviour. It is an attempt to interpret test results in terms of clearly-defined learning outcomes which serve as referents (criteria). Success of criterion-reference test lies in the delineation of all defined levels of achievement which are usually specified in terms of behaviourally stated instructional objectives. According to Glaser (1963) underlying the concept of measurement of achievement, lies the notion of the continuum of knowledge acquisition ranging from no-proficiency to perfect performance. It is on the continuum of knowledge that an individual's status regarding his/her achievement is to be determined. Unlike a norm-referenced measurement, the criterion level of the minimum acceptable performance for each objective is specified in advance, in criterion-referenced tests.

Use of criterion-referenced measurement at the elementary stage where learning of basic skills and fundamental concepts is essential, is a must to lay a proper foundation for learning at the secondary stage. It does away with the unfair comparison of an individual with other children. The major difficulty is the establishment of an achievement continuum in accordance with the complexity of the skills or the concepts involved. This notion also leads us to the concept of Minimum Levels of Learning (MLL) and Mastery Learning. MLL for primary classes has already been prepared by NCERT and Ministry of Human Resource Development, Government of India. The work in the field for secondary level is in progress. MLL can serve as criteria for evaluating students at any particular stage. In fact, both norm-referenced and criterion-referenced measurements have a place in the teaching-learning process.

### Check Your Progress 8

Explain Criterion-referenced Evaluation.

.....

.....

.....

.....

.....

.....

### 2.9.3 Difference

We now come to a consideration of the fundamental difference between norm-referenced and criterion-referenced tests. Basically, this distinction depends on the manner in which we interpret the results of an examinee's test performance. If you were to inspect a particular test, for example, a test of mathematical computation skills, it might be impossible for you to discern, merely from analysing the test. Items used on norm- and criterion-referenced tests may be very much the same. You would need to consult the technical and descriptive materials accompanying the test to see what constitutes the basis by which the test's scores are to be interpreted. Just as you can't tell a book by its cover, you typically cannot distinguish between criterion-referenced tests or norm-referenced tests merely by inspecting a test's items. How we can do it, is discussed in detail in the following sub-sections.

### 2.9.4 Relative or Absolute

In the case of norm-referenced test, we interpret someone's test performance according to the performances of others; in the case of criterion-referenced test, we interpret someone's test performance in relationship to clear description of what is being measured, for example, a domain of skills, knowledge, or attitudes. In a very real sense, then, interpretations are made relatively for norm-referenced tests and absolutely for criterion-referenced tests.

A reasonable working definition for a norm-referenced measure is:

**A norm-referenced test is used to ascertain an individual's status with respect to the performance of other individuals on that test.**

Although a norm-referenced test focuses on the relative status of an examinee's test performances, a criterion-referenced test endeavours to tie down the nature of an examinee's test performance more tightly or, if you prefer, absolutely.

Thus a simple working definition for a criterion-referenced measure will be:

**A criterion-referenced test is used to ascertain an individual's status with respect to a defined achievement domain.**

Whereas a norm-referenced test refers an examinee's performances to that of a norm group, a criterion-referenced test refers an examinee's performance to a defined set of criterion behaviours, that is, an assessment domain. An assessment domain might be a specific type of reading skill such as the ability to infer the main idea of a reading passage or, perhaps, a particular mathematics skill such as the ability to solve work problems based on two or more arithmetic operations. An assessment domain might also be affective in nature, such as a student's attitude towards learning.

A criterion-referenced test can measure one or more assessment domains. Most criterion-referenced tests that educators encounter are based on assessment domains dealing with skill or knowledge. Here are some other examples of the sorts of assessment domains measured by criterion-referenced tests.

Typical Assessment Domains:

- The ability to spell a set of 250 hard-to-spell words
- Knowledge of specified events in Modern Indian History
- Skill in solving algebraic equations

### 2.9.5 Comprehensiveness of the Test

A norm-referenced test typically measures a more general category of competencies (for example, reading comprehension), knowledge (for example, familiarity with the federal government system) or aptitude (for example, problem solving potential).

A criterion-referenced test, on the other hand, typically focuses on a more specific domain of examinee behaviours. Several criterion-referenced tests, each measuring its own specific assessment domain, would ordinarily, be required to tap the skills and knowledge being assessed by a typical norm-referenced test.

A 100-item norm-referenced test would attempt to cover the entire range of a learner's reading comprehension skills. In contrast, five separate twenty-item criterion-referenced tests would focus only on five well defined skills within the overall realm of reading comprehension.

### 2.9.6 Application

If the marking system used is impression marking, rating scale or checklist, the method of result determination may be classed as criterion-referencing; the student passes if he/she fulfils certain predetermined criteria. This is an apparently logical method of determining results, but is only reliable if the criteria are well defined, so that the standard of the assessment remains the same from year to year. For some practical skills, it is relatively easy to define the criteria reliably (e.g. the student should tabulate the given scores in a continuous series in class intervals of 5 each and calculate the mean, median and mode) and hence to maintain

the standard from year to year. If, however, the criteria are mainly in the minds of the markers, it becomes very much more difficult to ensure that they are applying the same standards as one another and as they did in previous examinations; this is particularly true of essay questions and of oral assessment.

For written examination and tests, criterion-referencing cannot usually be applied; A question paper tests only a sample of the objectives of the course and in any one year, may test an easier or more difficult selection of application questions. Testing the objectives may be easier or more difficult and even comprehension and recall questions may be made more or less difficult. Unless the questions are all pretested and banked, thus keeping the difficulty level same, and the marking is of guaranteed reliability, it is impossible to be sure that the test is of the same standard as that of a previous year. One solution to this problem is to determine results by norm-referencing, which assumes that the students are of the same standard from year to year and that, any differences in the level of marks awarded are due to the question paper or its marking. The same percentage of students is passed each year. This approach can be valid in a large national examination, but its use is questionable in college or small group examinations (i.e., where the group is less than a few hundred, unless there are special reasons for believing the standard to be constant). It is potentially unjust, since the same percentage of students will pass, whether their overall standard is high or low.

### Relationships between the three pairs

Now that we have discussed the three pairs of evaluation approaches - Formative - Summative, Internal - External and Norm-referenced - Criterion-referenced, can you perceive a relationship between them?

It has been hinted in the foregoing sections that it is generally found that formative evaluation uses internal and criterion-referenced tests, while summative evaluation is external and uses norm-referenced tests.

Since formative evaluation takes place during the course of the teaching-learning process it has to be internal. The teacher himself/herself has to assume the responsibility of formative evaluation and by the very fact that the teacher himself/herself is evaluating, it becomes internal assessment. Moreover, formative internal assessment is not aimed at grading or certification. So the tests which fit in place with them will be criterion-referenced. The teacher too have started with some specific goals in mind, and these goals or objectives form the criteria to be evaluated, Whatever has been taught/learned in the class will become the basis on which these internal formative tests will be constructed. Evidently, they will have to be criterion-referenced.

On the other hand, summative evaluation, which will be most of the time external and directed at grading and certifying students, will have to use norm-referenced tests, which enable the evaluator to make comparisons between the individuals and the individual with the whole group. The results of these norm-referenced, external summative tests will be used for various purposes that go with this kind of assessment.

### Check Your Progress 9

Explain the difference between Norm-referenced and Criterion-referenced Evaluation.

.....

.....

.....

.....

---

## 2.10 CONSTRUCTION OF CRITERION-REFERENCED TESTS

---

There are varieties of criterion-referenced tests and due to confusion in the nature and scope of such tests, it is quite difficult to provide clear cut guidelines for their construction. However, keeping in view the basic tenets of criterion-referenced tests, the following steps can be suggested. Some of these steps may be combined while others can be split up into two or more steps.

Nevertheless the steps given below seem quite appropriate:

- a) **Identification of Subject Area:** The first step in the development of criterion-referenced tests is the decision about the subject matter area to be worked out. For example one may take up Mathematics, English, Environmental Studies etc. Depending upon the need of the area and the resources one can utilize for one or more subject areas.
- b) **Selection of Unit/Topic:** After the decision taken about the area, the next step is to select the unit on which the test is to be developed. This unit may have more than one modules/sections/domains which comprise the total unit. Depending upon the need, one or more than one sections or chunks of content may be taken up for test construction.
- c) **Delineation and Description of Domain of Testing:** Since domain refers to a particular segment of the content, one may examine the topic and delineate it into various segments which can be developed into well defined separate sub-domains. Each sub-domain can then be analysed in terms of facts, concepts, principles, processes etc. That may be arranged in order of their increasing complexity. Description of sub-domain is very important as it provides the basis for item writing; it should reflect clearly the nature and scope of the content specification in sequential, hierarchical or development order.
- d) **Specification of Domain Objective:** Having decided about content elements of a domain selected, the next task is to formulate the instructional objectives on expected learning outcomes which may be categorised in terms of knowledge, understanding, application, skills, attitudes etc. These objectives should be stated so precisely that the performance of students is clearly interpretable in terms of adequacies or inadequacies of intended learning outcomes. For more clarification sample prototype items can be framed that may accompany each specific objective.
- e) **External Review of Steps (c) & (d) :** The tasks identified in steps (c) & (d) should be reviewed by those who are not involved in identifying the domain and its descriptions in terms of content elements and the specific objectives. However, a teacher who teaches that particular subject may also be involved in this task so that he/she may be able to clarify doubts, if any, raised by the external reviewers. The main purpose of this review is to sharpen domain description and the specific objectives in order to make them more realistic and functional. For this, sample items accompanying the specific objectives are checked for their congruence with each objective and content elements.
- f) **Internal Review:** After step (e) the internal reviewer i.e. the constructor himself/herself will examine all specific objectives alongwith one or more sample items which accompany each specific objective. The focus of this review is on sharpening the specific objectives further, if necessary.
- g) **Construction of Test Forms A & B:** It is desirable to develop two forms of the test, A & B so that one of the two could be used after post instructional remedial measures. Moreover, it would be easier to compute reliability on the two forms. A copy of the key or the correct answers should accompany the test and may be retained by the developer. As far as the construction of items is concerned, these items are to be

developed in accordance with the domain description. This is, however, not always possible or practical, especially in teacher made tests. Therefore, in teacher made tests only one test form will do.

- h) **Internal Review of Step(g):** An internal review by the teacher is essential after the test/s are ready i.e. after step (g). The purpose of this review is to see whether all questions in the test/s are congruent with the specific objectives, besides, having a cursory check for any glaring deficiencies in the test e.g. placement of key, arrangement of items, sufficiency of instructions etc. It is at this stage that items can be further improved and the congruence of items with the specific objective, is ensured.
- i) **External Review of the Test/s:** Prior to the field trial the test may be re-examined by the practising teachers. The main purpose is to detect content flaws, if any, and check for congruence of items with the domain description. This should be done preferably by those very people who do external review for steps (c) and (d) as mentioned under step (e).
- j) **Field Trial of the Test:** At this stage other test/s may be tried out on a limited number of students, say 5 to 10, to get a fix on the element in the instruction which might be proved drastically wrong. It is better that the subject teachers other than those involved in the development of the test, administer these tests. However, one team member of the developers may accompany them in order to meet queries, if any, relating to the content of procedure that might arise during try-out of the test. This is possible only if the member who is conversant with the development of the test is associated in the field try-out.
- k) **Internal Review:** After step (j), internal review would provide a last look at the test which would depend on the changes, if any, that have been made as a result of the review or the field try-out. The purpose of this review is to certify the final format of the test and pass it for print or use.
- l) **Final Form of the Test/s:** Now the final form of the test/s is ready for use and may be administered after having it published or cyclostyled, depending upon the size of the group to which this test is to be given. A sign-off sheet may be used to accompany the various tests as they move from step (a) to (l). Each test may be kept in a folder to which the signed sheet may be attached. This enables the developer to keep track of the given test. To monitor the progress of the test as it goes through the various stages of development, a master progress sheet can be used.
- m) **Using the Test in a Classroom:** Test copies can now be used in the classroom. The test can be administered to test the domains which are covered by the test. The domains being tested can be arranged according to the needs of the teacher and administered accordingly. Students' responses may be recorded and tabulated in accordance with the scheme of analysis which has to be mostly in terms of specified domain objectives.
- n) **Finding Validity and Reliability of the Test:** Since the data are now available on the test we may find out the reliability and validity of the test using various techniques.

---

## 2.11 LET US SUM UP

---

In this unit we discussed various approaches to evaluation. We took, these approaches in three opposing pairs; formative and summative; internal and external; norm-referenced and criterion-referenced. As a practising teacher we do at times have to think and take a decision as to which approach will be the most appropriate in a given situation or with a particular group of students or even in the interest of an individual pupil. The difference in subject or our objective can also make an impact on the approach we choose. Even otherwise, we need

to be aware of the various approaches as enlightened educators. Most of the time we may not be in a position to take a decision but if we are aware and have an insight into the theoretical aspects of education, we may be able to contribute in our own way to a change in the system, after we are able to explain the advantage of the change.

---

## **2.12 UNIT-END EXERCISES**

---

In Section 2.10 the steps involved in the construction of criterion-referenced tests are explained. Take up a unit in your subject and construct a criterion-referenced test following all these steps.

---

## **2.13 POINTS FOR DISCUSSION**

---

As we have seen in Section 2.3, formative evaluation is internal most of the time and uses criterion-referenced tests, while summative evaluation is generally external and uses norm-referenced tests. This gives us two distinct approaches to evaluation — criterion-referenced, internal, formative and norm-referenced, external, summative. Do you agree with this classification? Can there be other combinations? Further, which one of these two approaches is more important? Can we choose only one of them for our education system? In the move towards continuous and comprehensive evaluation we will have to gradually do away with summative, external examinations or at least minimise their importance. On the other hand, internal formative evaluation will gain significance. This is also closely connected with teacher empowerment, status as well as accountability and responsibility.

---

## **2.14 ANSWERS TO CHECK YOUR PROGRESS**

---

1. i) To provide continuous feedback to teachers and students on their teaching-learning process while it is in progress.  
ii) Grading
2. a) Find out the extent of achievement; (b) Grading
3. a) Judgemental - Developmental; b) Assign grades - improve Learning and Instruction;  
c) generally norm-referenced tests used-generally criterion-referenced tests used;  
d) certification-extent of masters of learning is going on.
4. Internal — Teacher teaching the class/subject is the evaluator.  
External — Someone else, an outsider is the evaluator.
5. a) The teacher is the best judge.  
b) Subjectivity or partiality may creep in.
6. The Justification may be based on the objectives for which evaluation is to be carried out. For example the external evaluation is intended for university selectors, employees and is seen as validation courses. Internal examination/evaluation can be used for assessing the learning outcome in all the domains i.e. cognitive, affective and psychomotor and provide continuous and comprehensive assessment.
7. In Norm-referenced evaluation, the focus is not on what one has learnt but is on where one stands in relation to the others in a group. The group is a norm group because it serves as a referent of norm for making Judgement. Hence in norm-referenced evaluation, the main criteria is the individuals relative position in the group.

8. In criterion-referenced evaluation, the ability of one is assessed against the standard criterion performance what has been set as an acceptable level of ability - demonstration.
9. Relative - Absolute; covers most of the course - covers a chunk or particular abilities/ outcomes used for comparison-used for determining the extent of learning according to fixed criterion; assumes average achievement - aims at mastery level achievement.

---

## 2.15 SUGGESTED READINGS

---

Ebel, Robert, L. (1996) : *Measuring Educational Achievement*, Prentice-Hall of India, New Delhi.

Ebel, Robert, L. and Fristic, David A. (1991) : *Essentials of Educational Achievement*, Prentice - Hall of India, New Delhi.

Harper, A. Edwin, and Harper, Erika S. (1992) : *Preparing Objective Examinations. A Handwork for Teachers, Students and Examiners*. Prentice-Hall of India, New Delhi.

Popham, W. James (1990) : *Modern Educational Measurement. A practioners perspective*, Prentice-Hall, USA.

Rammers, H.H. Gaje, N.L. Rummel, J. Francis (1967): *A Practical Introduction to Measurement and Evaluation*. Universal Bookstall, Delhi.