CHAPTER IV
TESTING AND EVALUATION

4.0 INTRODUCTION

Testing and evaluation of language skills and competencies are very important components of language teaching. Testing becomes an integral part of teaching because it provides significant information or inputs about the growth and achievement of learner’s difficulties, styles of learning, anxiety levels. Effective teaching and effective testing are two sides of the same coin. A curriculum is what constitutes a total teaching-learning programme composed of overall aims, syllabuses, materials, methods and testing in short. It provides a framework of knowledge and capabilities, selected to be appropriate to a particular level. Test evaluates not only the progress and achievement of learners but also the effectiveness of the teaching materials and methods used.

Language test broadly classified into two types as testing skills and testing knowledge of content. Skills such as listening, speaking, reading, and writing and sub-skills such as comprehension, vocabulary, grammar, spelling, punctuation, etc. Different kinds of tests are there to test student’s knowledge in language, the tests like non-referential test, aptitude test, proficiency test, achievement test and diagnostic test.

4.1 TESTING IN INDIA

A first serious attempt to study the advantage and disadvantage of language testing in India was made during the first National Workshop cum Seminar organized by the Central Institute of Indian Languages, Mysore in 1972. Since then enough work has been done in this field by organizations like the Institute of Psychological and Educational Measurement, Allahabad, Central Institute of English and Foreign languages, Hyderabad, National Council of Educational Research and Training, New Delhi., Central Institute of Hindi at Agra and at C.I.I.L. Mysore. The non-academic organizations like Institute of Banking Personnel Selecting and Staff Selection and Staff Selection Commission which are responsible for recruitments have prepared question banks in English and Hindi. The Central Institute of Indian Languages, Mysore conducted a National Workshop on
“preparation of tests in Indian Languages” at Bhubaneswar in November, 1985 and guidelines for the preparation of proficiency tests were prepared language proficiency tests in all the major Indian Languages have been prepared by now.

However, all the concerned agencies have concentrated mainly in the preparation of language proficiency tests and their one and only aim has been to assess the proficiency of students/candidates for such purposes as admissions, job selection, training and evaluation of language courses. Not much has been done in the preparation of prognostic and diagnostic tests which form an important part of second/foreign language courses. Whatever; the materials are available need to be tested further for their effectiveness, thereby to improve the test construction.

In this regard, the Central Institute of Indian Languages holds the leading positions and already conducted more than 60 workshops across the country, and produced nearly about 50 volumes of research materials in various Indian languages, conceived and formulated the National Testing Service (NTS) in pursuance of the provisions contained in the NPE 1986, NPE RC 1990, & CABE 1992. This multi-crore project of national importance is being implemented with the approval of the Central Planning Commission and the Ministry of Human Resource Development (language division) through 3 Task Groups with 51 personnel and 60 field units across 16 States/UTs of the country, as an extension scheme of the CIIL. The NTS also helps to make inter and intra language comparability and the individual's competence in terms of aptitude, achievement, and proficiency on a common scale in the contexts of NL/SL/FL for the purposes of admission, certification, and employment.

4.2 DISCUSSION ABOUT TESTING

Within the last three decades there have been tremendous theoretical and methodological developments within the field of evaluation. Despite its progress, there are still many fundamental problems faced by this field as "unlike medicine, evaluation is not a discipline that has been developed by practicing professionals over thousands of years, so we are not yet at the stage where we have huge encyclopaedias that will walk us through any evaluation step-by-step", or provide a clear definition of what evaluation entails (Davidson, 2005). It could therefore be
argued that a key problem that evaluators face is the lack of a clear definition of evaluation, which may "underline why program evaluation is periodically called into question as an original process, whose primary function is the production of legitimate and justified judgments which serve as the bases for relevant recommendations.". However, the strict adherence to a set of methodological assumptions may make the field of evaluation more acceptable to a mainstream audience but this adherence will work towards preventing evaluators from developing new strategies for dealing with the myriad problems that programs face.

It is claimed that only a minority of evaluation reports are used by the evaluand (client) (Datta, 2006). One justification of this is that "when evaluation findings are challenged or utilization has failed, it was because stakeholders and clients found the inferences weak or the warrants unconvincing" (Fournier and Smith, 1993). Some reasons for this situation may be the failure of the evaluator to establish a set of shared aims with the evaluand, or creating overly ambitious aims, as well as failing to compromise and incorporate the cultural differences of individuals and programs within the evaluation aims and process.

None of these problems are due to a lack of a definition of evaluation but are rather due to evaluators attempting to impose predisposed notions and definitions of evaluations on clients. The central reason for the poor utilization of evaluations is arguably due to the lack of tailoring of evaluations to suit the needs of the client, due to a predefined idea (or definition) of what an evaluation is rather than what the client needs are (House, 1980).

4.3 TYPES OF LANGUAGE TESTS

Tests are basically written as well as oral type. Written tests are the most popular testing tools in our schools all over the world while the oral tests have very little scope.

➢ Oral and written test

Test may be of written or of oral type. Listening and speaking are two important objectives to teaching a language. Here oral tests have very important
role to play in this regard. However, it may not be practicable to have oral tests at the public examinations, but it should not reduce its importance in the class tests. But with the help of a little planning it is possible to have oral tests in annual examinations. And this will provide strong incentive for the pupils to learn the ‘oral skills’ and for the teacher to teach these skills. On the other hand written tests are the most popular testing tools in our schools all over the world. They can be used to test most of the objectives of teaching a language.

➢ Oral method

As our main concentration is on oral testing that in oral method of testing no book comes between the teacher and the class. Here we discover that the results obtained by oral and conversational work are almost invariably superior, rarely inferior, to those obtained by the more traditional methods of book-work and pen-work.

There are, of course, exceptions, but in most of these cases we find either that the oral course was given by an experienced teacher or that the student was unwilling or unable to adopt his mental processes to this method.

Moreover Palmer, E. H. (1965, p. 12) discusses about the relationship between oral and conversation that:

"The terms ‘oral’ and ‘conversation’ are used as synonymous terms in reality, however they are not synonymous. The term “oral” is used to denote all that is done by the medium of the spoken word and much may be performed in the way of oral work that can not be covered by the term “conversation” or “conversational”.

In oral method the pupil’s powers of unconscious assimilation is trained and developed and it brings into play the pupils capacities for memorizing and habit-forming. The oral method will re-awaken and re-educate those spontaneous capacities for language study which are inherent in the human race; those innate powers that manifested in the earliest years of childhood, but which are generally allowed by disuse to recede into a latent state. However oral method has some advantage which are as follows as pointed out by Palmer, E. H. (1965, p. 14)
(i) Advantages of oral method

1. It is able to bring about a practical mastery of speech.
2. It is in accordance with the psychological analysis of speech representation.
3. It takes into consideration of the pupils' natural disposition or gift.
4. Easy to administer and score.
5. Comprehensive: oral tests assess four different areas of language proficiency, vocabulary, comprehension, which includes articulation.
6. Valid and reliable.
7. Oral methods also source of diagnostic and program planning.

(ii) Disadvantages of oral ability

The basic problem in testing oral ability is essentially the same as for testing writing.

1. We want to set tasks that form a representative sample of the population of oral tasks that we expect candidates to be able to perform.
2. The tasks should elicit behavior which truly represents the candidate's ability.
3. The sample of behavior can and will be scored as validity and reliability.

4.4 Definition of Evaluation

Evaluation is the comparison of actual (project) impacts against the agreed strategic plans. It looks at the original objectives, at what was accomplished, and how it was accomplished. It can be formative that is taking place during the life of a project or organisation, with the intention of improving the strategy or way of functioning of the project or organisation. It can also be summative, drawing learning from a completed project or an organisation that is no longer functioning.

Evaluation is inherently a theoretically informed approach (whether explicitly or not), and consequently a definition of evaluation would have be tailored to the theory, approach, needs, purpose and methodology of the evaluation itself. Having said this, evaluation has been defined as:
A systematic, rigorous, and meticulous application of scientific methods to assess the design, implementation, improvement or outcomes of a program. It is a resource-intensive process, frequently requiring resources, such as, evaluator expertise, labour, time and a sizeable budget.

The critical assessment, in as objective a manner as possible, of the degree to which a service or its component parts fulfills stated goals (St Leger and Walsworth-Bell). The focus of this definition is on attaining objective knowledge, and scientifically or quantitatively measuring predetermined and external concepts.

'A study designed to assist some audience to assess an object’s merit and worth' (Shufflebeam). In this definition the focus is on facts as well as value laden judgements of the programs outcomes and worth.

4.5 PURPOSE OF EVALUATION

The main purpose of a program evaluation can be to "determine the quality of a program by formulating a judgment" Stake and Schwandt (2006). An alternative view is that "projects, evaluators and other stakeholders (including funders) will all have potentially different ideas about how best to evaluate a project since each may have a different definition of ‘merit’. The core of the problem is thus about defining what is of value."From this perspective, evaluation "is a contested term", as "evaluators" use the term evaluation to describe an assessment, or investigation of a program whilst others simply understand evaluation as being synonymous with applied research.

Not all evaluations serve the same purpose some evaluations serve a monitoring function rather than focusing solely on measurable program outcomes or evaluation findings and a full list of types of evaluations would be difficult to compile. This is because evaluation is not part of a unified theoretical framework, drawing on a number of disciplines, which include management and organisational theory, policy analysis, education, sociology, social anthropology, and social change.
4.6 STANDARDS IN TESTING AND EVALUATION

Depending on the topic of interest, there are professional groups which look to the quality and rigor of the evaluation process.

The Joint Committee on Standards for Educational Evaluation has developed standards for program, personnel, and student evaluation. The Joint Committee standards are broken into four sections: Utility, Feasibility, Propriety, and Accuracy. Various European institutions have also prepared their own standards, more or less related to those produced by the Joint Committee. They provide guidelines about basing value judgments on systematic inquiry, evaluator competence and integrity, respect for people, and regard for the general and public welfare. American Evaluation Association has created a set of Guiding Principles for evaluators. The order of these principles does not imply priority among them; priority will vary by situation and evaluator role. The principles run as follows:

➢ **Testing Language skills**
  ✓ Listening
  ✓ Speaking
  ✓ Reading
  ✓ Writing

➢ **Achievement test**
  ✓ associated with process of instruction
  ✓ assesses where progress has been made
  ✓ should support the teaching to which it relates
  ✓ Alternative Assessment
  ✓ need for assessment to be integrated with the goals of the curriculum
  ✓ learners are engaged in self-assessment

➢ **Proficiency test**
  ✓ aims to establish a test taker’s readiness for a particular communicative role
  ✓ general measure of “language ability”
  ✓ measures a relatively stable trait
used to make predictions about future language performance (Hamp-Lyons, 1998)

g high-stakes test

Some ways of describing tests

<table>
<thead>
<tr>
<th>Objective</th>
<th>Subjective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indirect</td>
<td>Direct</td>
</tr>
<tr>
<td>Discrete-point</td>
<td>Integrative</td>
</tr>
<tr>
<td>Aptitude</td>
<td>Achievement</td>
</tr>
<tr>
<td>Proficiency</td>
<td>Performance</td>
</tr>
<tr>
<td>External</td>
<td>Internal</td>
</tr>
<tr>
<td>Norm-Referenced</td>
<td>Criterion-Referenced</td>
</tr>
</tbody>
</table>

Evaluating language skills

Evaluating the usefulness of a language test

Usefulness = Reliability + Validity + impact authenticity + Interactiveness + practicality

Essential measurement qualities

- reliability
- construct validity

4.7 ROLE OF TESTING AND EVALUATION IN TEACHING

The ‘testing and evaluation’ is an integral part of language teaching. In other words they are very closely related and interdependent activities. The absence of any one of these will have a bad effect on the other. They are like soul and body with each other. The reason is that whatever the students is being taught must be tested.
As it has been pointed out by Lado “what the students has to learn constitutes the corpus of what we have to test”. (Sharma, 2002, p. 212)

Teaching, also known as instruction, is an important factor to be used as text content for assessing educational programmers. Narrowly, it is defined as the act of instructing in an educational institution; broadly management by an instructor of the teaching, learning situations, including direct-interaction between the teachers and the learners; the pre-active decision making process of planning, designing, and preparing the materials for the teaching-learning conditions, and post-active redirection (evaluation, re-design, and dissemination). In most cases, it is used to the formal class room teaching, and also an object for ‘process evaluation’ as well as ‘programme evaluation’.

Hence, the major differences between the strategies of language teaching and testing are that teaching may go on for years. On the other hand testing is a matter of hours. In this regard testing is a very important component in a language teaching program. Glyn Lewis (1974, p. 270) suggests that: “language testing, whether it is to assess individual group achievement, to evaluate the impact of instructional techniques, or to diagnose individual difficulties is a part of language pedagogy”. Now to achieve the expected result in any work, testing is necessary/compulsory. For example, in the end of any language teaching class there are some sessionals given by the teacher for the purpose of testing and evaluation. That’s why, in teaching a language, testing is very necessary.

- **Language testing and evaluation**

  Language testing and evaluation are the two types of opinion where both are used for knowing the knowledge in a language teaching course.

- **Assessment and evaluation**

  There are many terms related to language testing. In this section only two are introduced that is assessment and evaluation as defined by Arthur Hughes (2003, p. 5).
➢ **Assessment**

In the book *Testing for Language Teachers (2nd ed.)* on page-5, Hughes clearly discusses about assessment. He suggests that testing is not the only way in which information about people’s language ability can be gathered. It is just one form of assessment. Actually, in a very general sense, Assessment is concerned with determining the amount of something, or with estimating its value. It is of mainly two types:

**i) Formative assessment**

When teachers use it to check on the progress of their students to see how far they have mastered, what they should have learned, and then this information is used to modify their future teaching plans. It could be used as a basis to get feedback about the students. Here students themselves may be encouraged to carry out self-assessment in order to monitor their progress and then can be able to modify their own learning objective.

**ii) Summative assessment**

It is used at the end of the term, semester, or year in order to measure what has been achieved both by groups and/or by individuals.

➢ **Evaluation**

In modern educational practice the term ‘evaluation’ is used in place of ‘test’ or, ‘examination’. It is an important concept as compared to either ‘test’ or ‘examination’. It is concerned not only with the measurement of the extent of learning but also determined the value of that learning.

Further, Pon Subbiah (2008) has pointed out that “suppose two students scored the same number then how can you differentiate them which one is best? Then we have to find out the other qualities what they have. And this type of judgment is called value judgment or evaluation”.

**4.8 DIFFERENCES BETWEEN TEST AND EXAMINATION**

When we are trying to make a distinction between test and examination, we can say that test and examinations are the widely used tools of evaluation. Actually a test is the activity itself and testing is a general term.
A test is regarded as an attempt to look whether the things which are taught have been learnt, while an examination is regarded as an attempt to find out that whether the students have attained certain objectives or predetermined standard. A test is directly concerned with teaching while an examination linked with an externally fixed standard of achievement. (Salim, 2001, p. 174)

Test may be classified as formal and informal. Formal test include quiz contest whereas informal test include discussion in class room. These tests may be held at any time. However, an examination is always formal and it is generally held at the end of the course. In addition, the aim of the test and examination is also different. The aim of the test is to check the progress and weakness of the students, whereas, the aim of an examination is to score marks on which their level pass or fail is ascertained.

But according to Davies, (n.d.) “the terms ‘test’ and ‘examination’ are often used synonymously because the famous ‘eleven plus’ test in the United Kingdom was sometime called an exam”. Examination is usually considered with achievement, but not always because the university scholarship examination is more a proficiency or even aptitude type of test. The type of achievement test referred to above is certainly not clearly seen as examination. R.R. Jordan (1997, p. 85) also suggested that “the ‘test’ and ‘examination’ are often used as synonymously.”

Furthermore, according to Halliday (n.d.), “by examination we mean ‘institutionalized tests’ which have an official or administrative function over and above their task of measuring performance”. A test is also school leaving certificate, or a university entrance qualification, or a matriculation equivalent, is in-fact an exam.

4.8.1 What is a test?

A Test has been defined as a “measuring device” a procedure for critical evaluation, a means of determining the presence quality or truth of something: for example, a trial: a test of one’s eyesight. We can also say “a test, in plain words, is a method of measuring a person’s ability or knowledge in a given domain”. Here the
definition captures the essential components of the test. Below, we can elaborate the definition that:

A test is the first method. As Brown (2001) points out that “a test is a set of techniques, procedures, and items that constitute an instrument of some sort”. He also suggests that it requires performance or activity on the part of the test taker (and sometimes on the part of the tester as well). The method may be intuitive and informal, as in the case of a holistic impression of someone’s authenticity of pronunciation. Or it may be quite explicit and structured, as in a multiple choice technique in which correct responses have already been specified by some “objective” means (Cf. Khan, A. Aziz. (1996).

Furthermore, Brown, H. Douglas (1987) pointed out about the purpose of the measuring of the test that:

“A test has the purpose of measuring. Some measurements are rather broad and inexact, while others are quantified in mathematically precise terms. The difference between formal and informal testing exists to a great degree in the nature of the quantification of data. Informal tests, the everyday intuitive judging, are difficult to quantify. Judgments are made in rather general term. For example, it is common to speak of a “good” tennis player, “fair” performance by an actor in a play, or a “poor” reader. In formal testing, in which carefully planned techniques of assessment are used, quantification is important.”

Brown also discusses that a test can measures a person’s ability also and in this regard he suggested that: “A test measures a person’s ability or knowledge --- that is, competence. A test samples performance but infers certain competence. A driving test for a driver’s license is a test requiring a sample performance, but that performance is used by the tester to infer someone’s general competence to drive a car. A language test samples language behavior and infers general ability in language. A test of reading comprehension may consist of some questions following one or two paragraphs, a tiny sample of a second language learner’s total reading behavior.
From the results of that test the examiner infers a certain level of general reading ability”. Cf. principles of language Learning and teaching by H. Douglas Brown.

4.8.2 Designing of the test

Now with the discussion of various kinds of tests some questions arise in our mind that how to ascertain if a test is a “good” test? Is it administrable within given constraints? Is it dependable? Does it accurately measure? So, if a test is a good enough then it should be able to provide answers of the above questions. A test must have certain qualities to be considered ‘good’. Now what are the qualities of a good test? According to many scholars, a good test must fulfill, at least, the following criteria. Further, these yardsticks are applicable to all tests in general.

➢ Reliability

Reliability is the accuracy of a measuring device (i.e. if a person is tested again and again or a test is marked by different persons then the result or score must always be the same or similar). Mostly reliability is considered as the measure of the stability of a test score(s) regardless of who marks it. If the same examiner marks the test at different times, the text will fetch the same score(s). A test will lack validity if it is not reliable. The reliability of a test may suffer from a number of factors i.e. examiner’s emotional or physical state, examiner’s mood, test’s inherent inadequacy such as illness, ‘a bad day’ or no sleep the night before. And also it is a clear case of unreliability that for a test of aural comprehension in which a tape recorder played items for comprehension, but because of street noise outside the testing room the students who were sitting next to the windows were prevented from hearing the tape accurately.

i) Test-retest reliability

Test-retest reliability involves administering the same test twice to the same group after a certain time interval has elapsed and then measuring the correlation between the two sets of scores. In educational research, scores collected over a two-month period are considered sufficient evidence of test-retest reliability.
ii) Equivalence reliability

Equivalence reliability occupies two different but equivalent (alternate or parallel) forms of an instrument are administered to the same group during the same time period to the same individuals and then find the co-relation between the two sets of scores. It is also possible to combine the test-retest and equivalent-forms methods by giving two different forms of testing with a time interval between the two administrations.

iii) The split-half reliability

Split-half Procedure involves by giving the tests once and then calculating the co-relation coefficient between the two halves of a test separately for each subject, one half consisting of all odd items 1, 3, 5, 7, etc., and the other half includes even items 2, 4, 6, 8 etc. This method has some advantages that split-half reliability is easy and economical and avoids many problems.

Reliability of a test can be increased by increasing the number of items and number of choices in a test. Thus reliability as Davies (1977) believes is the consistency of scores or reliability of scores.

➢ Validity

Validity of a test is perhaps most important attribute of a test. It means the test must measure the quality, ability that it claims to measure. A language test should test language and not the knowledge of the world or intelligence. A language test should measure the elements and skills of language. The simplest way to test the validity of a test is to arrange for a job sample. A test in fluency in foreign language should mean ability to converse with the native speakers.

Another way to check validity is to give the test to the native speakers of the language of the same educational level, those questions that they do not answer or doubt must be discarded. Validity of a test can be tested by correlating the test scores with some standard criterion or established tests. Validity of a test can be obtained when we clearly state the objectives of our teaching and testing. Test constructor should use contrastive analysis and linguistic analysis of L2.
i) Content validity

Content validity involves a key element is the adequacy of the sampling the content of a subject or syllabus or text-book is supposed to represent, it can claim content validity. The other aspect of content validation is the format of the instrument. Here attempts to obtain evidence that the items measure what they are supposed to measure typify the process of content-related evidence.

ii) Criterion-related validity

Criterion-related validity is a measure of test standardization. It gathers instrument data and criterion data and also compares at the same time. If the results of an examination are used to predict future performance of students and if those who make high marks are later found to succeed in higher classes better than those who score low marks, then the examination is considered to be a valid one with reference to that particular criterion. (Cf. Pon Subbish, 2008, p.50)

iii) Construct validity

Construct validity considered the broadest of the three categories. It samples a relevant theory in accordance with its Empirical validity: to correlate test scores with a criterion or estimate which, one believes is valid and accurate constructed. There is no single piece of evidence that satisfies construct-related validity. The more evidence researchers have form different sources, the more confident they become about the interpretation of the instrument.

Ebel (1979) is one of the views that “construct validity is the accuracy with which it measures some hypothetical operationally-undefined personal characteristics such as intelligence, motivation or creativity”.

iv) Predictive validity

Predictive validity of a test indicates how accurately it predicts the future performance of an examinee. It involves when time interval elapses between administering the instrument and obtaining criterion scores. The best criterion for a test is teacher’s ratings. Recently there has been emphasis to validate in as many different ways as possible. By for the most complex criterion of a good test is its validity.
➢ **Discrimination**

A good test should always discriminate among students according to their ability and performance. Good students should be able to score better than the poor. Discriminating power of test can be determined by a method called the method of third or E1-3 (e-one-three). The scores of candidate on a test should be ranked according to total scores, and then divided into three group upper, middle, lower. Each group should consist of equal number of candidates. If a test had 30 candidates, the each group should have 10 each. Then we take each question or item of the test. As a rule, a question should be passed by more people of upper third than lower or middle third and the people of upper third should get better marks. If a question is passed merely by people of lower third or passed by equal number.

➢ **Feasibility**

By feasibility of a test we mean it should be capable of being administrated, practically designed, should not be vague which can’t be materialized. A test so framed that their administration and scoring should be easy. Tests should be economical also. This is a practical criterion. The test should measure what it aims to measure in a reasonable time, should be given under the conditions that prevail. It should not require complicated electronic equipments and trained technician. In other words, a good test ought to be practical within the means of financial limitations, time constraints, scoring and interpretation. An expensive and complicated test is not practical i.e. a test of grammar or reading that takes 6 or 8 hours to finish for students is impractical. Long hours of scoring of a test also make it impractical. The test data or results should be helping teacher to understand students better to provide ‘feedback’ to the teacher. A test that is too complex or too sophisticated may not be of practical use to the teacher.

**4.9 THE CHARACTERISTICS OF TESTS**

Testing is an essential part of almost every educational system. It occupies an important position in language teaching program. The basic purpose is to calculate the knowledge or ability of a person, the evaluation or assessment is made for various purposes. Therefore, we can say that language tests are designed for
various purposes, which helps in making decisions about possible course of action, keeping in view the purposes, the tests may be characterized into:

➤ **Proficiency tests**

   It is constructed with a view to judge the knowledge of a person about a particular language, subject or profession. The proficiency tests give a general picture of a student’s knowledge and ability (rather than measure progress). These tests are generally conducted before giving admission to a particular language course and may not be based on any syllabus and their content depends largely on the mind of the tester. They are frequently used as stages people have to reach if they want to be admitted to a foreign university, get a job, or want to obtain some kind of certificate.

   Examples of a proficiency tests are competitive tests or admission tests that are conducted by almost all schools, colleges and universities in India. A very good example is PMT, MBBS, MBA and engineering tests.

➤ **Progress or achievement test**

   These tests are conducted to find out how much of a course a person has actually mastered and measures learner's progress of language and skills in relation to the syllabus they have been following.

   Achievement tests only work if they contain item types which the students are familiar with. This does not mean providing them with similar texts and familiar task types. If students are faced with completely new material, the test will not measure the learning that has been taking place, even though it can still measure general language proficiency.

   Achievement tests should reflect progress, not failure and it includes only what has been taught in a class and they are generally conducted after the end of a course.

   These tests are strictly based on a prescribed syllabus and they include only the section which has been taught in a class will form the part of achievement test. The amount of language mastered by students in a class can be measured by giving the same test or before and after the teaching or by giving two equal tests one at the
beginning and another at the end of the teaching. Example of achievement tests are the half-yearly and final exams of B.A, M.A, or high school.

- **Prognostic/ aptitude test**

  These tests are conducted for the purpose of finding out about future performance of the students. These tests deal with future learning or progress of a particular language or profession. These tests try to conduct whether a person will be a poor performer or he does better in future. In other words tells us about the ability of a person that he will show in future. There is no fixed time for this test and there is no particular syllabus.

- **Placement test**

  Placing new students in the right class in a school is facilitated with the use of placement tests. Usually based on syllabuses and materials the students will follow and use once their level has been described on, these test grammar and vocabulary knowledge and assess student’s productive and receptive skills. Some schools ask students to assess themselves as part of the placement process, adding this self analysis into the final placing decision.

- **Diagnostic tests**

  It is conducted to find out the weakness and short-comings of the learners. This is given with the purpose to find out the weakness and problems of the students. These tests can be conducted from time to time. It may be taken at the middle of the course. After finding out the problems of the students, the teacher will come to know what things have to repeat, and what remain to be taught. And it can be seen by the results of the test.

  A particular learner of a language may be weak in all or one of the skills. It is the duty of a teacher to find out where their problems lie. Diagnostic test is just like a medical diagnosis, the teacher is like a physician and student may be taken as patient. The student will be given various tests and find out the disease/weakness. Then the teacher will try to remove these in various ways.
4.10 TESTING AND EVALUATION IN ENGLISH LANGUAGE TEACHING

English Language Teaching becomes sound and meaningful when they take into account the prevailing theories of communication. The understanding of communication implicitly and explicitly influences our teaching learning programme. It is therefore worthwhile to examine the various aspects of English Language Teaching in the light of new insights into communication theories to assess how far the students are theoretically sound and to make necessary changes in the existing programmes to make them more effective. English Language teaching came into its own as a profession in the twentieth century. English undoubtedly taught as life skill but in institutions where it is taught as subject without for seeing the consequence of it in the future has to be replaced to the present day needs. English even if taught as subject life skills should be incorporated as part of the curriculum changes in language teaching methods throughout history have reflected recognition of changes in the kind of proficiency learners, such as move towards oral proficiency rather than reading.

Language teaching in the twentieth century was characterized by frequent changes and innovations and by the development of some competing language teaching ideologies because of the integration of globalization across and because of the remote economics of the world, the quest for better method was a pre-occupation of many teachers and applied linguists throughout the twentieth century. ELT is gaining momentum day by day in India. The growing importance of English puts the pressure on English teachers. The traditional method of teaching English has become obsolete. The demand for English is not only increasing but keeps changing. Taking the language classroom as its point of departure and return, the aim focuses on concerns about language learning and language teaching.

The importance of decision-making is introduced in a number of discussion that seek to relate language testing to evaluation, whether prominently as in Bachman (1990) or incidentally as in Henning (1987). Bachman (1990: 22-24) discusses the relationship between evaluation, measurement and testing along the following lines. All tests involve measurement, but not all measurement involves testing. Some
evaluation also involves in measurement, both in tests and in other forms of measurement, while some does not conversely, of course, not all tests and other measurements are used for evaluation. Bachman assets that "It is only when the results of tests are used as a basis for making a decision that evaluation is involved".

Language testing has long been an important area in applied linguistics, partly because constructs such as language proficiency have to be made explicit if they are to serve as models for test design and validation purposes. Validation involves ascertaining whether a test is effectively measuring what it was designed to measure. On the other hand, some of the specialized concerns of language testers have tended to keep them apart from other developments or discussions in applied linguistics and language education.

Evaluation has its own long tradition in education. Beretta’s discussion goes on to trace a current development of serious evaluation studies in applied linguistics. Murphy (1985) gives a picture of predominant neglect at the time, but more optimistically points out that a lot of teacher activity involves evaluation under other guises. Nunan (1992), quoting Brown (1989), draws attention to the idea that concerns over “needs analysis” in teaching English for specific purposes are not so far removed from questions of curriculum evaluation.

**4.11 WRITING SKILLS TAKEN FOR EVALUATION**

Writing skills has taken into the consideration for the evaluation. Writing skills can be evaluated with more accuracy compared with other skills. As the testing of oral abilities is often complex and time consuming, attempts are sometimes made to produce highly indirect tests, such as paper and pencil responses to “what would you say in the following situations?” to estimate oral proficiency. Suppose that such attempts extend to an entire examination, as opposed to just one class activity among many; it then becomes essential to validate such an approach to testing. Some kinds of validation studies, besides establishing the internal consistency (reliability) of the paper and pencil tests in question, would seek to compare performance on the tests with performance during oral interviews. Even if the out come of such comparisons were judged satisfactory, however,
fundamental questions would still remain about the content and construct validity of paper and pencil tests as measures of oral abilities.

Such tests are more likely to measure other abilities, such as aspects of knowledge about language varieties, than to measure oral abilities themselves, even if the two sets of abilities happen to be highly correlated. For such reasons, test validation becomes a matter of informed judgment that must draw on multiple perspectives, and not merely a question of establishing “the facts”. Construct validity are motivate by explicit descriptions of abilities, such as oral communicative abilities, that the test fails to embody, they are clearly acceptable in test validation. In contrast without such explicitness, otherwise comparable reactions on the part of non-specialists can easily be dismissed by specialists as mere appeals to face validity. This incurs a danger that potentially important information will be overlooked. When calls for explicitness deserve respect, people concerned with classroom language testing in particular cannot afford to ignore the impressions, however in articulate they may appear, of the people most concerned with the tests being set and taken. Indeed, eliciting impressions as explicitly and reflectively as possible from learners and others is a vitally important aspect of programme evaluation in general, and of classroom test validation as one aspect of this.

It is worth emphasizing that classroom language tests form part of curriculum, and can themselves be studied from the perspectives of language programme evaluation. Links between testing and teaching are of obvious important in classroom contexts. In making use of tests, language teachers will be concerned about messages given to learners and other non-specialists, and the importance or otherwise that learners will subsequently attach to different kinds of activities in class. The crucial notion of wash back effect (called backwash effect by some writers) is a name given to the impact of testing practices on teaching and learning the wash back effect of a complete paper and pencil examination of oral abilities would typically be highly detrimental to the direct encouragement of classroom oral interaction wash back effect is of central importance to the relationship between classroom language testing and programme evaluation.
Writing Skills Tested: I

Morphological Usage

Difficult Area

Easy Area

Frequent Error in Usage

Assessment & Evaluation

Easy Area

Difficult Area

Frequent Error in Usage

(Fig 4.1)

Noun
Pronoun
Verb
Adverb
Adjective
Preposition
Conjunction
Interjection
Articles
Prefix
Suffix
Pharas

12 Tense
Pattern
Simple
Complexé
Compound
Direct
Indirect
Positive
Negative
Interrogative
Superlative

Please purchase PDF Split-Merge on www.verypdf.com to remove this watermark.
Writing Skills Tested - II

Word ménage
Word usage
Sentence ménage
Sentence usage

Semantic Usage

Easy Area

Difficult Area

Frequent Error in Usage

Assessment & Evaluation

(Fig 4.2)

Essay writing
Paragraph
Note making
Note taking
Letter writing

Easy Area

Difficult Area

Frequent Error in Usage
Explicit accounts of language ability and its development, in relation to contexts of learning and use, will evidently be helpful to evaluate the effectiveness of programmes that set out to enhance the ability. The views of test takers and test users constitute essential data for the validation of forms of testing the aspects of educational practice. Test take and user views, though certainly an important one, of the attitudes and perceptions that classrooms language learners and others bring to a language programme as a whole. There are both ideological and educational reasons to view, learners as participants in language curricula, rather than as recipients of knowledge, so the views and values of participants are important in their own right as well as for their effects on leaning practices.

4.12 CONCLUSION

This chapter clearly described about language testing and evaluation and specifically explained about testing in writing. In writing morphological, syntactical and semantic levels are discussed more. The chapter evaluated all kinds of errors in writing of under graduate students. Testing and evaluation takes major role in language teaching and learning. Teachers can conduct different kinds of tests to know about the students skills in language. After test teachers will do the standard evaluation. The result shows the performance level of students. Evaluation is a systematic gathering of information for purposes of making decisions. It is the collection, analysis and interpretation about any aspects of a programme of education and training as part of a recognized process of judging its effectiveness. Evaluation is defined as an attempt to understand what is going on to judge its worth and make decisions. One can add many more reasons for doing evaluation. Evaluation is not just for measuring students learning at the end of a course. Tests are useful in diagnosis, prediction, selection, grading, guidance, self-correction, etc. There is internal as well as external evaluation in most educational institutions.