CHAPTER II
TESTING AND EVALUATION

2.1. General

As a matter of fact testing is an important phenomenon from science to arts, in order to weigh, measure and qualify the validity and the quantum of things. It is inevitable in all walks of life to measure, test and validate each and every activity to find out the nature and reliability of a person and the public as a whole. As it is important to have a testing on all the aspects in real life, language teaching and learning process also includes its role in the testing process. Language testing is as important as language teaching itself. In order to find out the nature and state of the students' proficiency, tests are to be conducted and the results are the only source, which provide valuable ideas, and suggestions that are considered for the remedial measures to be followed in the future course of action in language teaching process.

Language teaching began several centuries ago. The innovative ideas and methods adopted in the process provide valuable guidance and good models for both language teachers and learners. In order to evaluate the teaching-learning process on the whole, appropriate language test batteries become inevitable and such tests show a clear picture of the effectiveness and usefulness of the particular language teaching methods. Unless a particular teaching method is tested by reliable test batteries and empirically viewed with the test scores arrived from the learners responses, that particular teaching method would not be considered as useful and reliable for language teaching purpose. Previous
researches on the English language teaching provide valuable suggestions for language teachers, testers and test developers.

Carroll (1981: 66) says, "In designing our testing surveys, we will need to specify the communicative demands which offer variety of courses, of different levels, types and disciplines, and to device workable instruments to measure how far applications can meet those demands". The demands and the testing process do not end with merely providing the test scores, it has to do more than that by providing valuable suggestions in order to meet out the desired needs by means of the learners' language proficiency, which could be achieved through the language learning system. Apart from this, it is the duty of the testers to make the language course developers get knowledge about the reliability and usefulness of the learning tools, which they have constructed for the purpose of language learning. It is good on the part of the syllabus designer to pay attention to the ideas and suggestions of the persons conducting language test.

2.2. Definition of measurement, Test and Evaluation

2.2.1. Measurement

'Measurement' in social sciences is the process of quantifying the characteristics according to explicit procedures and rules. This definition includes three distinguishing features; quantification, characteristics, and explicit rules and procedures. (Lyle E. Bachman: 1990: 18)

Quantification involves the assigning of numbers, and this distinguishes measures from qualitative descriptions such as verbal accounts or non-verbal,
visual representations. Non-numerical categories or rankings making use of letter grades (a, b, c, etc) or labels (excellent, good, average) are used to qualify characteristics.

Testing involves, quantification of attitudes and abilities, sometimes called traits or constructs, which can only be observed indirectly. These attitudes include characteristics such as aptitude, intelligence, motivation, field dependence, independence attitude, native language, fluency in speaking and achievement in reading comprehension.

2.2.3. Rules and procedures

The distinguishing characteristic aspect of measurement is that quantification must be done according to explicit rules and procedures. That is, the 'blind' of haphazard as measurement. In order to be considered as a measure, the attitude observed must be replicable, for other observers, in other contexts.

2.2.4. Test

Carroll (1968) says test is psychological or educational. Test is a procedure designed to elicit certain behavior from which one can make inferences about certain characteristics of an individual.

A test is a measuring instrument designed to elicit a specific sample of an individual's behavior.

2.2.5. Language test

The most common use of language tests and educational tests in general is to pinpoint strengths and weakness in the learnt abilities of the students. We may discover through testing that a given student has excellent pronunciation and
fluency in the oral production in the language, but he or she may have a low level of reading comprehension. On testing further, we might find that lack of specialized vocabulary is a major factor underlying low reading comprehension for the students. Suggestions based on the suitable approaches for vocabulary development in learners are very much needed.

Evaluation can be defined as the systematic gathering of information for the purpose of making decisions (Weiss, 1972). The probability of making the correct decision in any given situation is a function not only of the ability of the decision maker, but also of the quality of the information upon which the decision is based.

2.2.6. Evaluation

Tests do not always follow evaluation procedures and in many cases the purpose of the tests is specific and they do not necessarily include the evaluation procedures. Mostly tests are conducted and made use of for pedagogical and recruitment purposes. In classroom, the tests play a major role in motivating the students to review the materials that has been taught or to know the quantum of information the students gained through teaching.

2.3. What is evaluation?

Evaluation is an activity through which the human behaviors, actions and happenings of the world are identified, perceived and realized. It is the only activity that controls and provides valid judgments and conclusions about each and every activity of the day-to-day events. Test is a part in the process of
evaluation but not the whole of it. An evaluation process may be complete when the tests are rightly interpreted with pros and cons of it. Evaluation related to language teaching, could be defined as a valid judgment or a fully reviewed statement about the proficiency of the students. Language skills are developed through a number of ways and methods. Assessing the proficiency of learners in a language is not an easy task. However proper evaluation procedures can provide the right judgments about the proficiency of the learners

2.4. Types of evaluation

Evaluation is nothing but the identification of language competence and performance of the learner during the course or at the end of the course. Evaluation in education includes different types like the evaluation of teaching methods, media of instruction, instructional materials etc. in addition to the learners' performance. Language tests are the measuring tools to assess the learners' achievements and therefore, they are administered to the learners and not to the materials or methods or to the teachers. They are designed to measure the learners' knowledge of the language that is being learnt or his competence both grammatical and communicative; the result of the test shows measurement and that in itself does not have much meaning. But the inference or the conclusion that can be drawn from the measurement is more crucial and important, and is called as the evaluation.

Evaluation is the qualitative and quantitative descriptions of subjects. Evaluation totally involves quantitative description (i.e. behavior described in terms of numbers) and qualitative descriptions (i.e. description expressed in words). The
terms measurement and evaluation though carry distinctly different meanings they are frequently used interchangeably. Evaluation involves the interpretation of what is measured in terms of number of words. Evaluation includes value judgments about the things described.

2.5. Process of evaluation

A number of processes are incorporated in evaluation. They are listed and discussed briefly below.

1. Identification of course objectives. (the expected or desired learning outcome)

2. Defining the objectives in terms of learners' terminal behavior.

3. Constructing appropriate tools or instrument for measuring the behavior.

4. Applying or administering the tools/instruments and analyzing the results to determine the degree of learners' achievement in the instructional program.

The above four steps are basically the same in the evaluation of instructions, curriculum or the program as a whole. Both measurement and evaluation require broad variety of tools or instruments such as, tests, rating scales, inventories, check lists, questionnaires etc.

2.6. Qualitative evaluation

A qualitative procedure might be the portfolio of evaluation. By this procedure a series of files might be centrally maintained for classes of all teachers and supervisors concerned with the implementation of the language lesson. File
folders might be organized according to lesson number, day or week of the instruction, class, sections, skill, area, etc. Teachers or teacher assistants might regularly record information such as student reactions to the lesson, appropriateness of length of materials, appropriateness effectiveness of content, adequacy of organization and sequencing, sufficiency of student opportunity for practice, problems in implementation, and suggestions for lesson improvement (Grant Henning 1977, pp: 186).

2.7. Quantitative evaluation

One of the quantitative evaluation procedures might be the unmatched group t-test. (a procedure to determine two groups by means of the test results). By this procedure students are randomly assigned to one or two different instructional groups. Each group receives a different instructional treatment. This treatment may be a method of instruction, a set of course materials, an incentive for achievement, and so on. Achievement or achievement gain is measured using the same instruments for both groups. T-test score means standard deviations are computed separately for each group. A t-value is computed and examined as an indication of the significance of the difference between the means for the two groups.

2.8. Types of evaluation

Evaluations in the context of language teaching may be divided into two main varieties. They are:

1. Ongoing evaluation (or) continuous evaluation

2. Terminal evaluation
2.8.1. Ongoing evaluation

Ongoing evaluation is meant for getting the feedback regularly after the completion of every step during its process viz. planning, preparation, production and application. This would enable the program to improve at various stages at that time of the program itself. This type of evaluation is more helpful to modify anything if necessary in the course of the didactic process.

2.8.2. Terminal evaluation

Terminal evaluation is a type of evaluation that is made after the completion of the program and it is used to know whether the program is a success or a failure. There is no other possibility in the result other than the above said two.

This type of evaluation would not be used for any improvement of the program.

In general, evaluation has been further classified into four categories: They are:

a. Formative evaluation
b. Summative evaluation
c. Brief evaluation and
d. Extensive evaluation

2.8.2.1. Formative evaluation

Formative evaluation is a process of evaluation that is made from time to time in the case of an instructional program and from one stage to the other. It does not provide a totalitarian impression of the quality either of the instructional programs, the techniques and methods, materials or media.
2.8.2.2. Summative evaluation

Summative evaluation is that kind of evaluation which takes into consideration the periodic evaluation that has been made and in addition to a total evaluation of the program: process or product made and the conclusions are arrived at keeping in view the outcome of the periodic evaluation in addition to the final evaluation.

2.8.2.3. Brief evaluation

Evaluating a program can also be made taking into account only some aspects and the evaluator can also give a judgment based on the few aspects chosen for evaluation. But it will be subjective and impressionistic and not a realistic one. This can be useful to roughly compare two (or) more programs.

2.8.2.4. Extensive evaluation

Extensive evaluation involves the analysis of a program in its entire main and sub aspects. The evaluator has to rate and weigh each of them individually and consolidate the total rating based on which he makes his value judgment. This is more objective and valid.

For the task of evaluating the procedures or methods, materials and media etc., we need a monitoring device. That is to say that we have a continuous or constant feedback about the effectiveness of the methods, materials and media.

2.9. Language Testing through Skills

Language testing means the testing of the four language skills namely listening, speaking, reading and writing. Language testing will not be fulfilled unless it includes the tests of all the four skills, since all these skills have one-to-one
relationship and though the modes of reception, production and quality differ. So, in order to test a learner's proficiency, the test batteries related to all the four skills of language become important. Test batteries can be developed to test one skill through another and it is evident that all the four skills are interrelated both in active and passive manner. During the phase of production, the active skills are supported by the passive skills and the passive skills are supported by productive skills.

Munby (1978: 126) says language is devisable into four skills of reading, writing, listening and speaking. The skills are, in turn, devisable into finer language skills (or functions) such as 'understanding conceptual meaning' with its related micro-skills such as 'quantity and amount', 'comparison and degree'. Language skills are measuring tools which help to understand a particular thing or a concept of things and facts viewed to be qualified and quantified by means of the skills a person possesses and language skills that help him to measure its value, quality, quantity, nature etc.

2.10. Language Tests in Curriculum

The present curriculum method, which is followed in India, does not follow the testing methods, which are meant for testing of four language skills. Either any one of the International Language Systems of ELTS (English Language Testing System), (Caroline Claphan, 1996; p.1) or any one of the standard language testing system has been followed in India. In India, English occupies the position of a second language. In India the learners, learning English as the SL differ by their mother tongue since they belong to different states. Even in one state
uniform type of testing system is not followed. No testing tools or testing devices are adopted and followed uniformly. Language teachers are the test developers and they develop the batteries according to the needs. It is found from the question papers of various universities and colleges of Tamil Nadu, that several types of tests are conducted to test the language ability of the second language learners. The present research is related to the Graduate level second language learners. So the testing systems pertaining to the graduate level language learning were viewed and adopted by the researcher.

The following components have been included in the syllabus.

1. Comprehension (Reading and Writing)
2. Grammar
3. Precise writing
4. Sentence patterns
5. Letter writing
6. Prose
7. Poetry
8. Hints developing
9. Phonetics
10. Stress / Intonation patterns
11. Conversation
12. Antonyms and Synonyms
13. Reorganizing the texts
14. General Essay writing
15. Grammar related exercises (word, sentences and passage levels)
All such above-mentioned types of teaching items are included in the syllabus. These items are taught in classroom and the students were tested through the writing skills.

2.11. Linguistic view on the curriculum design

Language is an activity pertinent to the human beings and it has developed a sophisticated need of the contemporary society. It is learned and practiced through the four skills. Can the present curriculum provide a base for the development of all the four skills in graduate level? The answer to this question is highly negative because the syllabus is designed with a view to develop the language skills of the students particularly to develop reading and writing in second language. The chances for the development of the oral communicative skills (listening and speaking) are almost nil. No separate teaching-learning syllabus has been included in the graduate level second language teaching system.

With regard to the teaching of grammatical aspects, only the grammatical categories are taught. The syntactic and semantic studies are not included. This may lead the second language learners to total confusion while using the patterns in second language like:

1. Structures of the sentences.
2. Usage of adjective and adverb.
3. Diversified use of same lexical items.
4. Inappropriate use of vocabularies and
5. Incomplete sentences.
The present researcher has found out errors in the above types of language use. Most of the students committed errors due to irrelevant substitution of verb, adverb, adjective and nouns. The probable reason for the error is due to the insufficient knowledge in the use of such items.

The present day syllabus does not advocate any scientific strategy to test the language elements taught to the students. Generally, a scientific syllabus, and teaching materials should have an inbuilt testing/evaluation procedure, so as to obtain the feedback from the learners. Since no testing parameter has been given in the syllabus or material, each one involved in the task of imparting English resorts to go for devising indigenous test device for evaluation, which in turn does not extract the real feedback from the learners.

2.12. Evaluation of the language tests

Importance must be given to the selection of the test and its appropriateness to the purpose for which it is administered. It should be based on reliable source materials for testing.

Grant Henning (1987,p: 9) says that in order to develop an appropriate test, the following information is to be taken into consideration.

i. Purpose of the test

ii. Characteristics of the examinees,

iii. Accuracy of measurement

iv. Suitability of the format and features of a test

v. Developmental sample,
vi. Availability of equivalent or equated forms
vii. Nature of the scoring and reporting of scores
viii. Procurement
ix. Political compatibility of the test.

2.13. Testing the learners outcomes

Testing the learners is as important as teaching learners. When the learners are measured correctly, their level of understanding can be clearly identified. The testers should consider many things before they conduct test to the learners. The important points to be remembered are:

i. When to test?
ii. What type of test to be used?
iii. How is the test going to be maintained?

2.14. Testing procedure adopted to the present study

The present study has to do with the learning of the second language learners belonging to the undergraduate course. The present curriculum does not provide language curriculum or paper after second year of the graduation studies except for those who study language subjects as their major subjects. This study concentrates on the second year under graduate students who are in the final stage of their language learning process in the academic domain. They have been selected as informants for the test. This test was conducted during the months of January and February 2001. No test or examination was conducted during these selected months by the college authorities. Hence the researcher selected these two months to conduct test for the present study.
2.15. Methodology adopted in the language tests

In order to find out the language proficiency and fluency of the second language learners especially the graduate students, and to know how they use the language for both academic and non academic application purpose, a specially designed testing system related to the testing of all the four language skills was developed and the same was issued to the learners from various colleges for testing. The methodology adopted for the preparation of questionnaires had not strictly followed any of the standard test items. The researcher made models and materials, which are suitable and reliable in testing the informants. The testing material is in accordance with the following models.

2.15.1. Writing Test

The questionnaire framed for testing writing skill was mainly based on the testing of the items and aspects of writing such:

1. Vocabulary
2. Spelling
3. Grammar
4. Expression
5. Fluency and
6. Style.

In order to test the above, seven questions based on general topics in which the graduate students are familiar were given, and the informants were asked to write on any of the five questions. The questionnaire is given in the appendix.
A writing comprehension has also been included in the writing test. In the writing comprehension questions were framed in the following levels namely:

1. Word
2. Sentence and
3. Passage

The comprehension skills namely, recognition, inference, recreation, and reorganization have been tested through the comprehension test on writing.

Finally one question related to the translation of text from one form to another form was included. The students were given a conversation to convert into a passage form. The conversion skill has been tested through this exercise.

Responses obtained for all these above types of questions were analyzed both quantitatively and qualitatively and a detailed comment has been given for each and every exercise.

2.15.2. Reading test

In order to test the reading skill questions were prepared giving importance to reading comprehension. 20 questions have been framed as objective, or yes/no decision types. Questions were framed in three types of sub skills namely recognition, inference and reorganization. The aim of the reading test was to find out the reading ability of the students. So only objective type questions were selected and no writing based answers were included in the reading test questionnaire.

On the whole 20 test items were included in the questionnaire. For reading comprehension three long and, small passages and one advertisement column
were also selected for the questionnaire. All the test items are commented with the probable reasons for the errors and the correct answers along with the remedial measures to develop the reading skills are given with corresponding exercises.

2.15.3. Listening tests

Listening test was conducted with the view to test the students' listening capacity in perception and identification of speech sounds in English language. A questionnaire was designed for the same, which included the tests of

1. Identification of words
2. Inference related to the informants given in a passage
3. Recognition of meaning to the words given in a passage.
4. Retention and recall of the sentences given in spoken form.

This test was conducted with the help of the audio tape recorder.

Test materials

1. One conversation. (Previously recorded and given for listening comprehension)
2. Five individual sentences (previously recorded and given for testing the recall capacity) and
3. One passage having four sentences (given for testing retention and recall capacity)

All the listening materials were previously recorded by the students of same standard and the same was played for listening. Questions were printed in papers and given to the students for writing answers.
2.15.4. Speaking test

The aim of speaking test was to identify the level and quality of production of speech sounds of students in the speech of second language. In order to test the same the students were asked to speak on a topic of their interest and no restrictions were laid on it. Each student was given five minutes to speak and their speeches were recorded in an audiocassette for analysis.

For testing pronunciation, it is necessary to test all the sounds in English language. So a number of selected words, having all the sounds in English were selected and given for loud reading. The reading voice was recorded and the same was analyzed with the Daniel Jones pronunciation model.