

# **CHAPTER-I**

## **INTRODUCTION**

### **1.1 GENERAL BACKGROUND**

Language is a means of human communication through which we express our feelings, thoughts, wants, desires, and so on. Longman dictionary of language Teaching and Applied linguistics (1999) defines language as "The system of human communication which consists of the structured arrangement of sounds (or their written representation) into larger units, e.g. morphemes, words, sentences, utterances." Cambridge International Dictionary of English (1977:795) defines language as "A system of communication consisting of a set of small parts and a set of rules which decide the ways in which these parts can be combined to produce message that have meaning."

As we know the language is species specific to mankind i.e. only human beings can have the capability to speak language and their mind is genetically equipped with it. Language consists the vocal noises made by human beings. Vocal sounds such as sneezing, coughing, snoring etc can communicate some meanings but cannot be considered as language because these sounds are not deliberately used for the purpose of communication. These are involuntary sounds Thus, language as the voluntary vocal system of human communications can be considered as the more adequate definition. So, the term language refers only to the voluntary vocal sound which speakers use deliberately for the purpose of communication.

### **1.1. 1. Language Testing**

The history of language testing goes back to the history of language teaching, it was not taken as a separate discipline in the past. It was taken as the part of teaching. Testing in a broad sense has always been an inherent part of teaching. Assessment of learning is as old as education itself. From the time when teaching began, the teacher has always been keen to know the extent to which his teaching has been effective in making the learner understand what has been taught. Testing is used as a process of scrutinizing how far learners have learned what the teacher wishes them to learn. In this process, the teacher usually makes queries in the classroom during or after his teaching or administers an examination at the end of a lesson, or a unit, or a chapter, or a course of study. The intent of the teacher in making queries or administering an examination during or after his teaching on what has been taught obviously is to understand whether or not the whole or part of his teaching has been imparted and received by the learners. To ensure that the students have achieved some or whole of what has been taught is thus another purpose of testing. In order to ensure that the teaching is effective, testing is used in the classroom or after the classroom teaching.

In the case of Language Testing, pre-scientific era was described before 1960s which was also called pre-discrete point testing. It was the time for translation, long essay writing and literature. It followed whatever general principles of testing were available in humanities or social sciences.

After 1960s, there were structuralists who were interested in validity of language. Multiple choice items and short answer questions were practiced in that era. Testing focused on specific language elements such as phonological, grammatical, and lexical units. Some new tests like

dictation and cloze test emerged in 1970s due to the failure of discrete point testing. In that era, language was not broken down into pieces. It was seen as a whole complete message. But such tests did not give real value for interaction. In other words they were not communicative tests.

The approach to communicative language testing has a close relation with communicative language teaching. From the mid seventies the people involved in linguistic research and language pedagogy were influenced by the concept of communication or communicative competence. The "Communicative Competence" was first used by Hymes (1972) to refer to "The intuitive mastery what the native speaker possesses to use and interpret language appropriately in the process of interaction and in relation to social context." The influence of the approach to see language from the point of view of communication was that the context of an utterance emerged as very vital in establishing its meaning. The problem for testing is to consider how important the context is and whether or not it is possible to create a genuine context for an utterance when, by nature, testing itself provides a pseudo opportunity for genuine communication.

Teaching and testing are taken as inseparable entities. In other words, teaching and testing are so closely interrelated that it is virtually impossible to work in either field being constantly concerned with the other. Harrison (1983:1) says, it is a natural extension of the classroom work, providing teaching, teachers and students that can serve each as a basis for improvement. Davies (1968:5) says that the good test is an obedient servant since it follows and apes the teaching. Testing in a broad sense has always been an inherent part of teaching. Assessment of learning is as old as education itself. Testing is used as a process of scrutinizing how far learners have learned what the teacher wishes them

to learn. In this process, the teacher usually makes queries in the classroom during or after his teaching or administers an examination at the end of a lesson, or a unit or a chapter or a course of study. The Longman Dictionary of Applied Linguistics (1999) defines testing as "The use of test or the study of theory and practice, their use development and evaluation. Test is any procedure for measuring ability, knowledge or performance." Test may be primarily constructed as a device to reinforce learning and to motivate the students or primarily as a means of assessing students' performance in language. The importance of testing can be imagined through the statement "Teaching without testing is like painting in bad light." With the help of testing we can find the areas that need more attention and we can do the remedial works.

### **1.1.2 Theory of Foreign Language Testing**

The theory of foreign language testing is based on present linguistic understanding of language and an observation concerning the role habit in learning a foreign language .This theory is congruent with psychological knowledge and thinking but constitutes an organizing of the problem that is not found in psychology textbooks at present.

The theory of foreign language testing assumes that language is a system of habits of communication. These habits permit the communicator to give his conscious attention to the overall meaning he is conveying or perceiving. These habits involve matters of form, meaning and distribution at several levels of structure, namely those of the sentence, clause, phrase, word, morpheme and phoneme. Within these levels, there are structures of habits of articulation syllable type and collocation. Associated with them and sometimes as part of them are patterns of intonation, stress and rhythm.

When this transfer occurs, some of the units and patterns transferred will function satisfactorily in the foreign language and will not constitute a learning problem. Other units and patterns will not function satisfactorily in the foreign language. Against these the student will have to learn the new units and patterns. These constitute the real learning problems.

These learning problems turn out to be matters of form, meaning, distribution or a combination of these. They can be predicted and described in most cases by a systematic linguistic comparison of the two language structures. When several elements of the native language structures are factors, it is sometimes difficult to predict exactly what the problem is going to be. In such cases a special explanatory test or set of interviews can answer the question.

In listening to or reading the foreign language the same transfer takes place, only the sequence is reversed as the listener is exposed to the forms first, which he interprets through his native language units to the meanings of his own language and culture, except for these which he can keep in his conscious attention or which he has already mastered as a linguistic habit.

The theory assumes that testing control of the problems is testing of the language. Problems are those units and patterns that do not have a counterpart in the native language or that have counterparts with structurally different distribution or meaning.

The problems in speaking are not necessarily the same as the problems in listening. For example the question "Does he speak?" is often a problem in speaking for some foreign students because they add an "s" ending to speak and say "Does he speaks?" .In listening this is not a

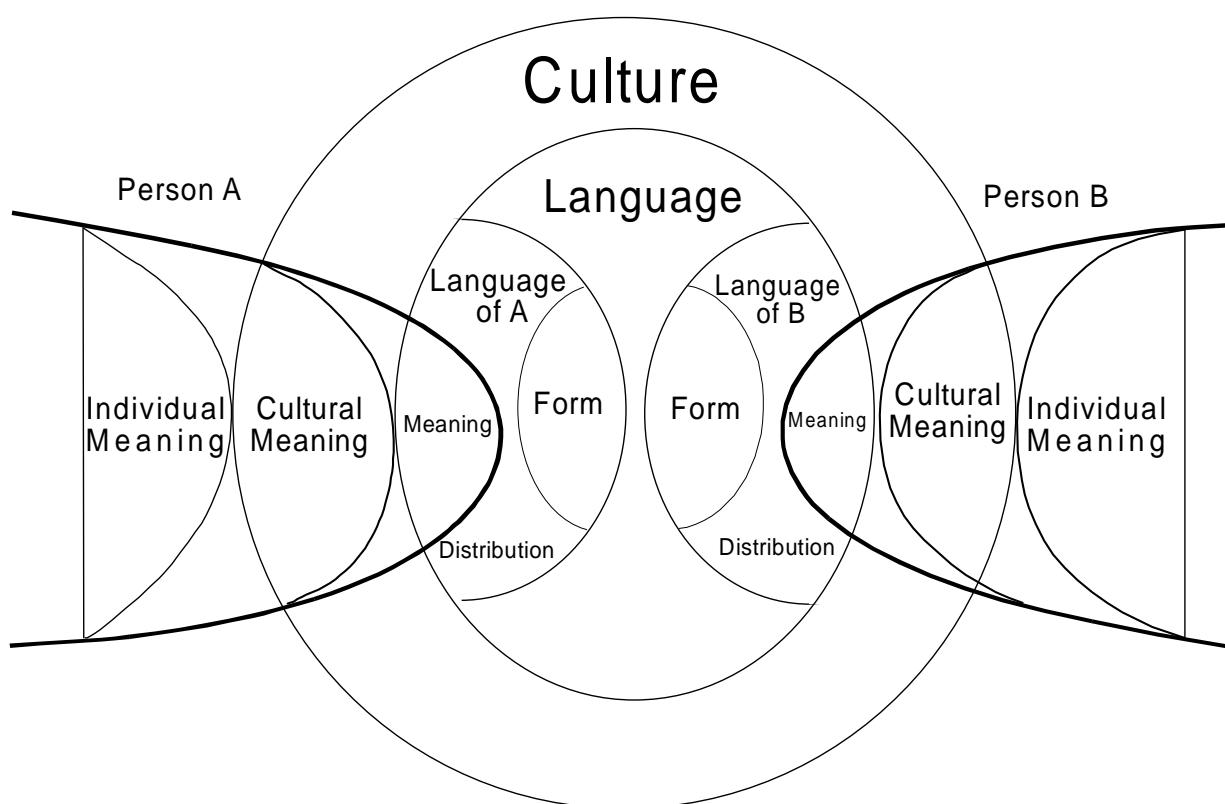
problem because the correct ending is given by the native speaker and it does not confuse the student.

The theory assumes also that the student does not know these units and patterns that are problems unless he can use them at normal conversational or reading speed in a linguistically valid situation, that is, a situation that parallels those of language in use. Definitions, for example, are not valid linguistic situations because native speakers are able to use forms and structures that they cannot define and other speakers are able to define structures that they cannot use. Stating grammatical rules is not a valid linguistic situation. Lists of words are not in themselves valid situation.

The theory assumes that linguistic and cultural meanings communicated through language are also structured and associated to the formal elements specifically stated above.

When the communicant speaks his native language, he goes from individual stimulation and meanings which are not part of the units of language to cultural and linguistic meanings that are encoded these cultural and linguistic meanings in the forms that are associated to them in the language. When these forms are uttered, the listener perceives them and through them, he grasps the same linguistic and cultural meanings that the communicant encoded. The individual meanings the listener grasps in addition are not part of the language.

**Figure: Language, culture and the individual.**



The individual is not aware of in using language is done thought a complex system of habits. When he attempts to communicate in a foreign language that he knows partially, he adopts the same linguistic posture as when using his native language. He thinks of the over all meaning and proceeds to encode consciously in addition to one or another matter of grammar or pronunciation or vocabulary, but the bulk of the encoding goes to his habit system and here it is channeled through the system of habits of his native language. This, in psychology, is known as transfer. He transfers the habit system of his native language to the foreign tongue.

When this transfer occurs, he produces the sounds of his native language and the sentence patterns of his native langue, in short the entire structure of his native language in the foreign one, except those few units and elements he is able to keep under conscious control and those he has

mastered to the point of habit. If his attention is brought to something he has missed and already known as the conscious level, he will correct himself but may miss something else instead. Several repetitions may produce enough immediate memory to result in satisfactory production, but when the same problem is met elsewhere it may be missed again.

### **1.1. 3 Qualities of Good Tests**

Any test, to be a good test should have some common characteristics. Although the characteristics of test differ from author to author, some common characteristics of good tests are mentioned below. These characteristics should be taken into consideration while writing the tests otherwise it becomes "... just as it is impossible to play chess without knowing how a knight moves across the board, so it is point less to write tests without a basic understanding of the principles behind them" (Harrison, 1991:10). (The qualities of good tests are validity, reliability, scorability, administrability, economy etc. If the tests lack one of these qualities, they can't be good tests.)

There are different views on what makes a test good. Some experts say that there are three constituents of exam efficiency: validity, reliability, and practicality. In this respect, validity, reliability, and practicality should be seen as relativistic concepts. The whole idea of considering the three constituents of exam efficiency is to build-up a framework for designing a good test.

Bachman and palmer (1996) argue that test usefulness involves reliability, construct validity authenticity, interactiveness, impact and practicality. Though the presentation of Bachman and Palmer may appear



to be a bit elaborative, in essence, validity, reliability and practicality constitute the quality of a test. These issues will be discussed below.

### **(i) Reliability**

Reliability is one of the essential qualities of a test which refers to the consistency of scores or performance of the same or similar test administered within a reasonable time. It is a matter of the extent to which we can believe that the performance is true, how likely it is that the performance will be repeated next time. Reliability can be dealt with at two levels: test and retest of students and marking and remarking of the examiners. It is a statistical concept. It is reported in terms of correlation coefficient. In short, it is designed as consistency of measurement. Reliability is concerned with examining consistency in the performance of the examinee. In order to establish the reliability of an examination it is necessary to answer the question: how consistent would the examinee's performance be if we asked him to take the same exam at a different time, or another examination which is supposedly similar?

### **Factors Influencing Reliability**

There are different factors that contribute to the reliability of the test. They are as follows:

- Homogeneity of items: if a test has the test items testing the more or less the same trait, the test will have high reliability.
- A test with high discriminating power items will produce high reliability.
- Variability of group-students with a wide range of ability will yield high reliability.
- Sufficient test taking time will give high reliability.

- A test with less freedom of choice will yield high reliability.
- A test with less unambiguous items will have high reliability.
- Objectivity in scoring will give high reliability.
- Length of the test also contributes to produce high reliability; longer the test, higher the reliability.

## **(ii) Practicality**

Practicality is different from other qualities of a test. Absence of this quality in a test will read the test to be of no use. It has been shown that reliability and validity are the most important aspects of exam efficiency. Another aspect of it, though non-technical, is practicality in the absence of which even a valid and reliable exam can be of no use. Heaton's (1975: 158) explanation of practicality is that the exam "must be fairly straightforward to administer." Generally, practicality involves the cost and ease of administration, and scoring. We would interpret this rather in a vague way that an exam must fit in the intended situation in all respects.

This aspect of exam efficiency is important because failing to achieve the practicalities would lead to have problems to convince the authority concerned. If any exam is likely to create some other problems because it did not fit into the given situation, it would be difficult to convince the people who would actually implement the exam.

## **(iii) Validity**

Validity is a very important quality of a text. "A measure is valid if it does what it is intended to do ....." (Davies et al 1999) Further explanation is that the validity of a text is measured on the basis of how far the information it provides is accurate, concrete and representation in

light of the purposes for which it is administered. Recent trend in determining validity of a test has included characteristics such as having meaning for teachers and students so as to demonstrate content, mentioning explicitly the standards for scoring etc. involvement of these characteristics will enhance the wash back validity- quality of the test. Heaten (1988) defines validity as "The validity of a test is the extent to which it measures what it is supposed to measure and nothing else." Validity refers to the degree to which a test actually measures what it is designed to measure. It means a test which is designed to measure the function actually measures the same, then it is known as valid test. In the same way Harrison says, "The validity of a test is the extent to which the test measures what it is intended to measure." Likewise Robert Lado (1976) says, "A test of pronunciation measures pronunciation and nothing else, it is a valid test of pronunciation."

There are five types of validity. They are:-

- Content validity
- Construct validity
- Criterion- related validity
  - (i) Concurrent validity
  - (ii) Predictive validity
- Face validity
- Wash back validity

Content and construct validity are said to be conceptual, and concurrent and predictive validity are said to be statistical validity. Face validity is said to be pseudo validity. There are other terms like consequential, systematic, discriminate, ethical, etc that come along with

discussions on validity wash back validity will also be discussed which is emerging as one of the necessary elements for a good test.

### **Criterion Related Validity**

As said above, criterion-related validity and predictive validity are statistical concepts which are also called "empirical validity." This type of validity is established employing a process of comparing the results of a test with the results of some criteria already set or the subsequent performance of the students. The validity at a test established by comparing with a set criterion measure is called criterion related validity. It is important that a test must also be empirically validation to ensure that it has elicited the information it was supposed to elicit. This can be done by checking the performance of the examinee in the exam against an external criterion. Davies (1983:141) says for the criterion-related validity in the following words. "The external criterion, however hard to find and however difficult to operationally quantify, remains the best evidence of a test's validity. All over evidence, including reliability and the internal validities is eventually circular."

Criterion-related validity can be established by giving the students an established test with similar nature which has proved to be valid. The test can be administered at the same time or in a short gap ensuring that no additional learning opportunity is given. Criterion-related validation procedures determine the efficacy of an examination in prediction the examinee's future performance in a pre-specified situation (Anastasi 1982: 137).

Criterion-related validity is discussed under two heads: Concurrent validity and predictive validity. The main difference between the two types of validation procedure is time interval if the exam scores are validated against the criterion approximately at the same time, it is

concurrent validity, and if the exam scores are validated against the criterion after a stated time interval, it is predictive validity.

### **Concurrent Validity**

By name, the concurrent validity of a test refers to the process of determining the validity against the set criterion at the same time. Test developers tend to establish the validity of the new test by comparing the performance of the students on this test against their performance on a test of similar kind already established. The correlation between two tests is said to be the concurrent validity of the new tests. The established test can take a form either a well known test or the rating of the teachers. Establishing concurrent validity of a test is very common. Most test developers would be interested to find out the extent to which the test correlates with some other standardized tests. This validity is obtained as a result of comparing the results of the test with the results of some criterion measure such as.

- An existing test, known or believed to be valid and given at the same time..
- The teacher's ratings or any other such form of independent assessment given at the same time.

### **Predictive Validity**

Predictive validity of a test is concerned about the extent to which the test can predict the future performance of the testers. This type of validity is established by comparing test results with another criterion such as success in a particular job or in higher education. Predictive validity is important in the sense that a test is supposed to predict the future performance of the candidates which can be established against the

external criteria. One of the difficulties in dealing with predictive validity is to find a satisfactory criterion with which the exam results are to be correlated. To establish the predictive validity of an exam, some standard measure of performance must be pre-specified. It is this pre-specified performance which serves as a criterion.

Despite the fact that academic achievement has been the common criterion for establishing the predictive validity of a test (Anastasi 1982:138), one of the problems in criterion selection is, it is admitted, that we still do not know the extent to which language proficiency determines academic success. Crippen and Davies (1988:65) point out that failure in academic achievement may be because of the language component. The tests-the TOEFL, the Michigan test, the IELTS – are being used as selection devices on the assumption that there is prominent role of language in academic success. This validity is obtained as result of comparing the results of the test with the results of some criterion measure such as:

- The subsequent performance of the testees on a certain task measured by some valid test;
- The teacher's rating or any other such form of independent assessment given later.

### **Construct Validity**

If a test has construct validity, it is capable of measuring certain specific characteristic in accordance with a theory of language behaviour and learning. This type of validity assumes the existence of certain learning theories or constructs underlying the acquisition of abilities and skills. For example, it can be argued that a speed reading test based on a short comprehension passage is an adequate measure of reading ability

(and thus has low construct validity) unless it is believed that the speed reading of short passages relates closely to the ability to read a book quickly and efficiently and is a proven factor in reading ability. If the assumption is held that systematic language habits are best acquired at the elementary level by means of the structural approach, and then a test which emphasizes the communicative aspects of the language will have low construct validity. Conversely, if a communicative approach to language teaching and learning has been adopted throughout a course, a test comprising chiefly multiple-choice items will lack construct validity.

Construct validity is discussed as the center of all methods of validating a test. Accepted the inclusiveness of the construct validity, it can be argued that if a test provides evidence that it has content and criterion-related validity, the exam also provides some evidence that it has construct validity as well. Brown (1976:128) writes: "construct validity is implied when one evaluates a test or other set of operations in light of the specified construct.

### **Face Validity**

Face validity is defined as "what it appears superficially to measure". This type of validity, in fact, is often referred to as face validity: if a test items looks right to other testers, teachers, moderators, and testers, it can be described as having at least face validity.

The concept of face validity is far from new in language testing but the emphasis now placed on it is relatively new. In the past, face validity was regarded by many tests written simply as a public relations exercise. Today, however, most designers of communicative test regard face validity as the most important of all types of test validity. Indeed, many argue that a test must look valid even as far as the reproduction of the material itself is concerned: thus, a test of reading comprehension using

such authentic tasks as reading and skimming newspapers must contain actual newspapers or, at least, articles printed in exactly the same way as they appeared in the newspaper from which they were taken.

### **Washback Validity**

Washback is inherent in an exam: an exam is bound to influence teaching and learning. The possible reason for why the exam is influential on education should be seen in terms of its functions. The functions of the exam can be summarized under its forward looking and backward looking purposes though they are not exclusive. The backward looking functions of the exam have to do with a sense of achievement and evaluation of the effectiveness of teaching. The forward-looking functions are related with making decisions about the examinees, for example, selection and certification. In a practical world, the procedures of selection and certification are necessary to establishing communication between educational institutes and receiving agencies.

Morrow (1986:6) argues one of the principal responsibilities of testing or examining boards and institutions, which operate in the public domain is to provide, "a package" with a powerful and positive washback effect into the classroom. The real significance of Morrow's (1986) remarks about washback validity lie not so much in their emphasis on validity as such, but on the highlighting of the purpose of validation (and hence test use). In conclusion, we can say that wash back validation proceeds from the test to the classroom.

### **Content Validity**

Content validity is defined as "whether the items composing the test do, in fact constitute a representative sample of the content domain of concern" (Brown, 1976:122-123). In case of a final achievement test, it is



said to have content validity if its test items are considered to be a representative sample of the tasks as can be seen in the course objectives. Anastasi (1982:131) describes the nature of content validity as: "content validity involves essentially the systematic examination of the test content to determine whether it covers a representative sample of the behaviour domain to be measured."

Regarding the procedures to be followed at the construction stage to make the exam of high content validity, Anastasi (1982: 132) provides the following guidelines:

1. The behavior domain to be tested must be systematically analyzed to make certain that all major aspects are covered by the test items, and in the correct proportion.
2. The domain under consideration should be fully described in advance, rather than being defined after the test has been prepared.
3. Content validity depends on the relevance of the individual's test responses to the behaviour area under consideration rather than on the apparent relevance of item content.

What emerges on the basis of these guidelines is the concern of content validity with the *priori* stage of exam construction. What we should be doing, to make an exam educationally beneficial, as was discussed, is to make the exam a mirror of the course objectives in order to make the people concerned understand what is expected of them.

A test is said to have content validity if its content constitutes a representative sample of the language skills, structures etc with which it is meant to be concerned, Hughes (1995). The test should be constructed as to contain a representative sample of the course, the relationship between the test items and the course objectives. In the absence of a

specification of the skills or structures or course objectives, a test may lose content validity.

Content validity is, "The extent to which a test measures a representative sample of the subject matter content" Hatch and Farhady (1982:251). In the same way Richards et al (1999), "content validity is a form of validity which is based on the degree to which a test adequately and sufficiently measures the particular skills or behavior it sets out to measure. For example, a test of pronunciation would have low content validity if it tested only some of the skills which required for accurate pronunciation." Bachman (1998) has provided two aspects with the help of which we can examine the actual content validity of a test. They are content relevance and content coverage.

### **Content Relevance**

"Validity is essentially a matter of relevance. Is the test relevant to what it claims to measure? Intelligence within the limits of normalcy is an irrelevant factor. For a test to be valid we expect the content and condition to be relevant and that there will be no irrelevant problems which are more difficult than the problems being tested" Lado(1999). The investigation of content relevance requires the specification of the behavioral domain in question and the attendant specification of the task or domain. Content relevance involves the specification of the ability domain and the test method facets. The amount or the way that subject matters are closely related with the course is content relevance. The more test items are constructed, the more content validity the test paper has. If most of the contents are covered in test paper there is the establishment of content relevance.

## **Content Coverage**

It is the extent to which the tasks required in the test adequately represent the behavioral domain in the questions Hughes (1989) views that content validity depends on the how many of the functions are tested in the component and how representative they are of the complete set of functions included in the objectives. Similarly, Harrison (1991) claims that content validity is established by considering the purposes of the assessment and then drawing up a content lists.

A test can't be fully valid or fully invalid. A test is therefore judged as having content validity when the test items present the course content and the coverage objectives. Content coverage is amount or the way that the subject matter of the course covers an area.

### **1.1.4 Course Structure of the Master's in English Education**

There are altogether eleven (11) papers, out of them five (5) papers carry 100 marks each and six (6) papers carry 50 marks each. These eleven papers are divided into two groups. There are ten papers in group A and Group B has elective one paper. In elective group there are four papers but only one paper will be offered as prescribed by the concerned subject committee in the campuses for teaching learning processes.

**Table No: 1**  
**Specialization**

S.N.	Course Title No.	Subject	Marks
1	Eng. Ed. 511	Phonetics and phonology	100
2	Eng. Ed. 512	Grammar: Theory and Practice	100
3	Eng. Ed. 513	Psycholinguistic and sociolinguistics	100
4	Eng. Ed. 504	Research Methodology in language education	50
5	Eng. Ed. 505	Language testing	50
6	Eng. Ed. 551	Semantics and pragmatics	50
7	Eng. Ed. 552	Applied linguistics	100
8	Eng. Ed. 589	Thesis/Eng. Ed. 574 Discourse analysis	50
9	Eng. Ed. 590	Language testing methods and practices	100
10	Eng. Ed. 599	ELT practicum	50

**Table No: 2**

**Electives**

S.N.	Course title and No.	Subject	Marks
1	Eng. Ed. 571	English literature and its pedagogy	50
2	Eng. Ed.572	Advanced Reading and writing	50
3	Eng. Ed. 573	Translation: Theory and practice	50
4	Eng. Ed. 575	Second language acquisition	50

*Source: CDC, T.U.*

From the above data and description we know that at M.Ed. first year three specialization papers: Eng. Ed. 511 (Phonetics and Phonology); Eng. Ed. 512 (Grammar: Theory and practice) and Eng. Ed. 513 (Psycholinguistics and sociolinguistics). In second year 7 papers from the specialization group: Eng. Ed. 504 (Research methodology in language Education); Eng. Ed. 505 (language testing); Eng. Ed. 551 (semantics and pragmatics) Eng. Ed. 552 (Applied linguistics); Eng.Ed. 589 (Thesis or Eng. Ed. 574 Discourse analysis); Eng. Ed. 590 (language teaching Methods and practices); Eng. Ed. 599 (ELT practicum) are adopted. From the electives group one paper is adopted out of the four papers.

Generally in our testing system, most of the schools, colleges and universities follow the only annual examination system. Written, oral and practical examinations are administered according to the nature of courses. In M.Ed. second year in the subject "language testing" only annual written test is administered to measure the students understanding. The full marks of this subject is 50 and 20 is its pass marks. Subjective

(long and short questions) as well as objective questions are administered in the test. 8 objective questions carrying 8 marks (i.e. one mark for each item) are asked in the objective test. Similarly, in the case of subjective questions 5 short questions are asked carrying 6 marks for each question. And one long question carrying 12 marks is the weightage of the question. It is hoped that 8 objective questions, 5 short questions and 1 long question can represent the whole course as well as their weighting suits according to its contents. So the purpose of the study is to look in to whether the test papers represent the whole course or not, and whether the weighting of the given test is proportional to the weighting of the course contents or not. Do they measure what they are intended to measure or not?

### **1.1.5 The Syllabus of Language Testing at M.Ed. Second Year**

The course "English language testing" is a comprehensive introduction to language testing in general and English language testing in particular. It includes knowledge and incorporates construction and use of tests in language classroom. The course also offers strategies for analyzing the tests and their results and the ways they can be used for improving test instruments and language programmes are ten units in this course of which the first two provide theory and insights into language testing. The following units specify aspects and techniques for testing the elements and the skill of languages while the last two units deal with refining the test and include some statistical procedures.

The course is primarily designed for master's level students in education specializing in English education. However, students studying language teaching in other department and faculties or perspective language teachers can equally benefit from it.

## Course Objectives

On completion of the course the students will be acquainted with

- the prevalent theories of language testing,
- strategies and procedures of testing for various elements and skills of language,
- the techniques of the construction and use of tests of language elements and skills;
- the ways of analyzing the language tests, interpreting test results and of using those tests in language classrooms

The contents of this course can be divided into ten units as follows:

Unit I : Introduction

Unit II: Classification of language tests

Unit III: Considerations in designing tests

Unit IV: Constructing tests

Unit V: Testing language competence

Unit VI: Testing language performance

Unit VII: Testing Reading

Unit VIII: Testing Writing

Unit IX: Testing communicative Abilities

Unit X: Interpreting language test scores

**Note:** The more detail of this course contents is given in the appendix – VI

*Source: CDC. T.U.*

This paper carries 50 full marks. 20 is its pass marks. Though only external written examination and final examination the students' competence is tested.

**Table No. 3**

**Unit wise weighting and time allotment**

Unit	Marks	Time
I	10	15 hrs.
II		
III	30	45 hrs.
IV		
V		
VI		
VII		
VIII		
IX		
X	10	15 hrs.

In the case of question setting only skeleton framework of syllabus does not work while the test writer does not follow the syllabus fully and sets the questions as his/her whims. So, a good test item should have validity is one of the main qualities of a good test. A test without validity is like a ship without captain.



## 1.2 LITERATURE REVIEW

Although, there are some research works held in the field of language testing none of the researches are carried out to find out the content validity of M. Ed level language testing paper. This will be the first work to find out the content validity of the exam of the above mentioned course. Some researches carried out in the field of testing are mentioned below.

Khaniya (1990) has conducted a researches on "Examination as Instruments for Educational changes :Investigating the washback effect of Nepalese English Exams", and comes to the conclusion that the SLC exam fails to assess the language skills that the SLC English course intends to develop in students ..... because of its text book and previous exam paper oriented nature , it doesn't encourage student and teachers to focus on language skills entailed in the objectives (Khaniya 1990: 245). Finally he has concluded that

- i. Washback is an inherent quality of exam.
- ii. Ingredients of exam determine whether the washback is negative or positive and
- iii. Teaching for final exam is inevitable.

Khanal (1997) carried out a research work on "A study on the effectiveness of cloze test over conventional objective tests in testing reading comprehension in English". The purpose of the study was to compare the effectiveness of the two types of tests: cloze test Vs objective test in testing reading comprehension. The study concluded that the private school' students' performance in both objective and cloze test was far better than the objective test in testing reading comprehension.

Khaniya (2002) has written an article on "washback: Emerging Validity" in the journal of NELTA (Nepal English language teachers Association) (August-2000, vol: 5:31) and concluded that 'washback' is an inherent attribute of an examination: To be a good examination an exam shouldn't only exert a negative influence but it must also have the potential to exert a beneficial influence on teaching, where necessary. Likewise, Khaniya (2006) has also published the article on "Reform in the SLC examination: "A fiasco" in Saikshik Jyoti (year: 1 Vol: 1, Oct. – Nov. 2003), the magazine which publishes six times a year in CERID and concluded that it is urgent that persons or agencies responsible for the reform of the SLC examination take initiative to change the system, evaluation and education in a competent manner. An exam like the SLC exam should not continue rejecting about 70 percent of its participants every year.

Batala (2004) studied on 'validity of the SLC examination English question paper.' The main objective of the study was to find out the predictive and content validity of the SLC English examination. The study concluded that the predictive validity of the SLC English examination was very low because the coefficient of correlation between the two sets of scores on the SLC and grade 11 English examinations is +0.1 on the other hand in terms of representation, objectives and item-wise analysis the SLC English examination has good content validity. But in terms of weighting, the same test papers have low content validity. The finding of his research is that the content validity of the SLC examination English question paper was high.

Aryal (2005) carried out a research work on "quality of English Exam. A case of content validity of Grade Twelve compulsory exam 2061." He has analyzed the question paper from different angles (by

rubrics, length, difficulty level, and content coverage). They don't cover all genres equally; the content validity of meaning into words is nearer to the coverage of units and teaching items in comparison with heritage of words. The study concluded that the rubric of the all questions were simple, scientific except few items; The length of the question matches to the allotment of time for the examinees ; the questions of English are moderate difficulty level; the questions related to heritage of words lack content validity and unscientific.

Bhattarai (2005) has carried out the research entitled, “content validity of the English textbook for grade eight.” In her findings, she found that the contents are applicable and the book has content validity in terms of content coverage and applicability. Skills are less valid, functions are less valid but language structures have the high content validity.

Neupane (2005), has carried out the research on ‘The content validity of English Textbook for Grade seven.’ In his research he found that some of the contents are valid and some are less valid. He has conducted the research on the basis of content, coverage, selection, gradation and language skills. In his research he has not mentioned the language functions and language structures.

### **1.3 OBJECTIVES OF THE STUDY**

- i. To examine the content validity of language testing exam paper at M.Ed. second year in terms of:
  - a. Content coverage, and
  - b. Content weighting
- ii. To suggest some pedagogic al implications.

#### **1.4 SIGNIFICANCE OF THE STUDY**

As a distinct research work from the rest in the department of English education, this study will be useful for the department. This study provides information on whether the administered language testing test had content validity or not.

The findings of this research will be beneficial to all those who are concerned with language learning /teaching specifically language testing. It will be equally beneficial for the policy makers and curriculum designers. No doubt, it will be helpful for the student of applied linguistics. If the teachers and test designers are familiar with constructing test papers they will be aware of the variation and mistakes hidden in the tests and try to minimize them.

## **CHAPTER - II**

### **METHODOLOGY**

The study, in this section, deals with the methodology. The researcher has followed the following methodology during the study in order to achieve the objectives specified.

#### **2.1 SOURCE OF DATA**

The researcher has used only secondary sources of data as follows:

#### **2.2 SECONDARY SOURCES OF DATA**

Test papers from the years 2058 to 2062 B. S. on course entitled 'Language Testing' at M. Ed. Second year was the main secondary source of data. Apart from this, T. U. new English syllabus, different books on testing, reports, journals and articles available related to the topic. For examples Khania (2000, 2002), Batala (2004), Aryal (2005) etc.

#### **2.3 PROCESS OF DATA COLLECTION**

The researcher collected five years (2058 to 2062 B.S.) question papers entitled 'Language Testing' at M.Ed. second year and the researcher read and analyzed both subjective and objective test items of the above-mentioned academic year.

#### **2.4 LIMITATIONS OF THE STUDY**

This study had the following limitations.

- i. The study is limited to the content validity of the test on language testing of M.Ed. Level.

- ii. The study is limited to the question paper from 2058-2062 asked in Tribhuvan University.
- iii. This study has been done only in the theoretical basis.
- iv. The statistical tools used for analyzing the data were tabulation and percentage.
- v. The study is limited to the testing of content validity of written test of English since there is no provision of oral test in the M.Ed. level second year's test on subject language testing in Nepal.

## **CHAPTER – III**

### **PRESENTATION, ANALYSIS AND INTERPRETATION**

#### **3.1 ANALYSIS OF THE CONTENT VALIDITY**

This chapter deals with the analysis and interpretation of the raw data used in the study. The main purpose of this research work was to find out the content validity on language testing test at M.Ed. level. For this research study data analysis is presented into two parts (i.e. content validity in terms of coverage and in the second part the content validity of the language testing tests in terms of weighting).

Here in the first part, the question papers of the language testing exam papers which were administered in Tribhuvan University (T.U.) examination during 5 years (2058-2062) are analyzed in terms of coverage. The researcher tried to find out whether the test paper had content validity or not.

There are altogether 10 units of the whole course in subject entitled "Language Testing" at M.Ed. second year. It is a theoretical subject and 50 is its full marks, 20 is its pass marks. Course contents and its weighting is clearly specified in the syllables, which are given in Appendix I. Table No. 4 for units 1 and 2, table no.5 for units 3-9 and table no.6 for unit 10 have in the exam from 2058 to 2062. The researcher has mainly analyzed subjective as well as objective questions of those academic years. And the question papers of those 5 years are given in appendix I, II, III, IV and V.

As we have different views on content validity, Davies et. al. (1999) focused on the basis of how far the information it provides is accurate, concrete, and representative in light of the purposes for which it is

administered. Hughes, A (1995) and Bachman 1998) have emphasized two things for content validity: 'content representation' and 'content relevance' but here, the researcher used mainly the Heaton's (1988) view on content validity. In Heaton's view there are two components through which we can judge whether or not they have content validity; they are 'representative sample of the course' and 'weighting of the course.'

Thus, using the above two principles or guidelines as 'representative sample of the course', and percentage weighting' the researcher tried to find out whether the administered tests have content validity or not. Whether the tests have representative sample of the course or not and the tests have strictly obeyed the weighting of the course or not.

### **3.1.1 Representative sample of content coverage comparison between course contents and test contents**

For the purpose of testing the test investigating the content validity of the 'Language Testing' test at M.Ed. second year during 5 years (2058-2062 B.S.), the researcher has compared the test content in relation to course contents or he examined whether the test contained a representative sample of the whole course or not. If the representative sample of the coverage of contents is above 50 percent, than it is believed that the test paper is nearer to content validity. If it is below fifty percent, then it is supposed to lack content validity. The more test items are constructed, the more chances of having content validity. If more than 60 percent course contents are covered in the test, than it is supposed to have high content validity. So, to find out the content validity of a test in terms of course representative ness, the researcher examined and analyzed the 'Language Testing' test paper from the year 2058 to 2062 of its subjective as well as objective questions. The researcher has used only descriptive method and tabulation method to analyze the data.



### 3.1.2 Examining Course Representatives in Unit One and Two

**Table No. 4**

#### **Representation of Test Contents in Terms of Course Contents in Unit One and Two**

S. No.	Course Contents	Test Contents				
	Course Items	Test Items Represented				
		2058	2059	2060	2061	2062
Unit 1	Introduction					
1.1	Language education and testing			S1		
1.2	Testing language competence					
1.3	Testing language performance					
1.4	Historical perspective				O2	
1.4.1	Traditional approach					
1.4.2	Structural approach					
1.4.3	Communicative approach					
Unit	Classification of					

2	language test					
2.1	Goal-based tests	O3	L6			
2.1.1	Proficiency test	S1			O1	
2.1.2	Achievement test				O5s1	S1
2.1.3	Diagnostic test				O6	
2.1.4	Prognostic test/aptitude test					
2.1.5	Placement test					
2.2	Medium-based tests					
2.2.1	Oral test					
2.2.2	Written test					
2.3	Mode-based tests			S5		
2.3.1	Objective test			O5	O3	
2.3.2	Subjective test					
2.4	Aspect-based tests					
2.4.1	Pronunciation test					
2.4.2	Vocabulary test					
2.4.3	Communicative function test					

2.5	Skill-based tests		O1			
2.5.1	Test of listening			O1		
2.5.2	Test of speaking					
2.5.3	Test of reading					
2.5.4	Test of writing					
2.6	Approach-based tests					O3s4
2.6.1	Discrete-point test					
2.6.2	Integrative test			O5		
2.7	Reference-based tests			S2		
2.7.1	Norm-reference test					
2.7.2	Criterion-reference test					
	Total no of asked questions in unit: 1 and 2	1 obj, 1 s.s.	1 long, 1obj, 1.s.s.	3 obj, 2 s.s.	5 obj, 1. s.s.	1 obj, 2 s.s.

**Note:** o3 = Objective question: question no. 3.

S1 = Subjective short question no. 1.

L6 = Subjective long question no. 6.

The above table indicates that in unit one there are 7 language items from 1 to 1.4.3 and in unit two there are 27 language items from 2 to 2.7.2. If we see diachronically among these 34 language items (i.e.

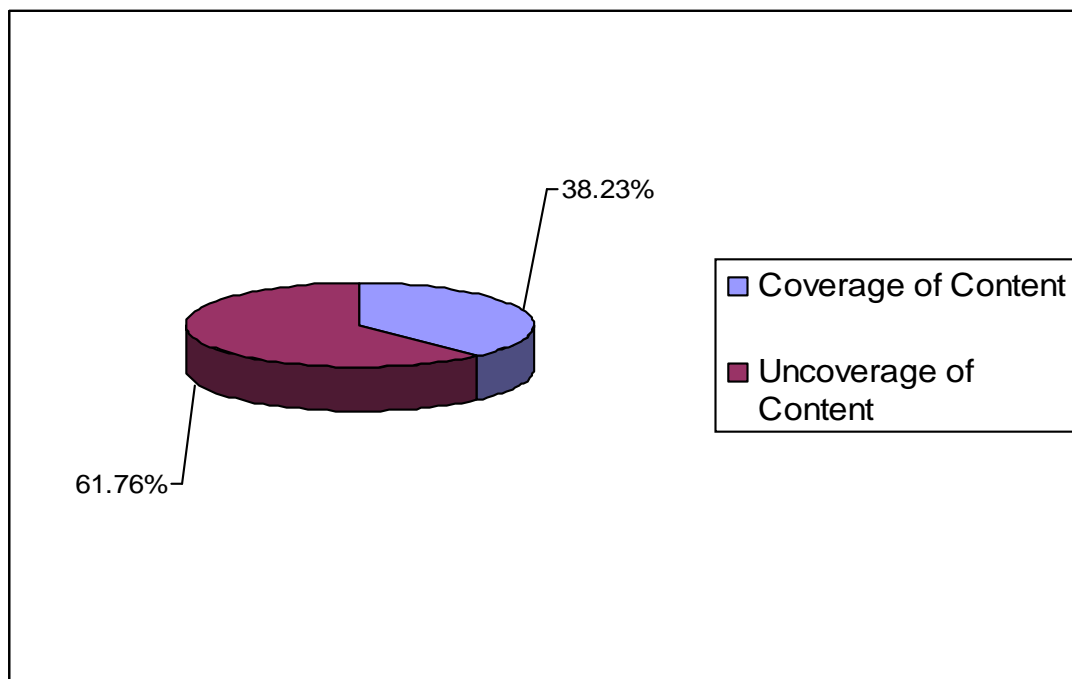
including unit one and two) there is no any most representative language item repeating each and every year from 2058 to 2062. The language items 2.1 (Goal-based tests) 2.1.1 (Proficiency test), 2.1.2 (Achievement test), 2.3.1 (Objective tests) have been represented only two years questions out of five years. The language items 1.1 (language education and testing), 1.4 (Historical perspective), 2.1.3 (Diagnostic test), 2.3 (Mode-based tests), 2.5 (Skill-based tests), 2.5.1 (Test of listening), 2.6 (Approach-based tests), 2.6.2 (Integrative test), 2.7 (Reference-based tests) have been represented only in one year out of five years. Besides them, the language items which are mentioned in the table have not been represented in any year.

If we see synchronically, 2 questions (1 objective and 1 subjective) were asked in 2058. In the case of objective question that is asked from 2.1 (Goal-based tests) and one subjective question was asked from 2.1.1 (proficiency test). In 2059, 3 questions (1 long question, one short question and 1 objective) were asked. Long question was asked from 2.1 (Goal-based tests), short question was asked from 2.3.1 (Objective test) and objective test was asked from 2.5 (skill-based tests). In 2060, 5 questions (three objectives and two short questions) were asked. One objective question was asked from 2.3.3 (subjective test), another question was asked from 2.5.1 (testing of listening) and another question was asked from 2.6.2 (integrative test), short questions were asked from 2.3 (mode-based tests) and 2.7 (reference-base tests). In 2061, 5 questions (4 objectives and 1 subjective short question) were asked. The objective questions were asked from 2.1.1 (proficiency test), 2.1.2 (achievement test), 2.1.3 (diagnostic test) and 2.3.1 (objective test) short question was asked from 2.1.2 (achievement test). In 2062 3 questions (1 objective and 2 subjective short questions) were asked. The objection question was

asked from 2.6 (approach-based tests) and subjective short questions were asked from 2.1.2 (achievement test) and 2.6 (approach based tests).

In conclusion, the above presented table and description can be shown in the pie chart as follows:

**Figure No. 1**



From the above table and description the researcher found that there are 34 language items (including unit one and two because unit one and unit two cover 10 marks) according to course content but the representation of the test items is only 13 language items. Twenty one (21) language items were neglected while constructing tests items. It means the course in test contents in unit one and unit two is 38.23 percent. 61.76 percent content were not covered in the question papers. According to unit one and two have low content validity because it (test papers) covers less than 50 percent course contents.

### 3.1.3 Examining Course Representativeness in Unit Three, Four, Five, Six, Seven, Eight, and Nine

Table no. 5

#### Representation of Test Contents in Terms of Course Contents in Unit Three, Four, Five, Six, Seven, Eight and Nine.

S. N	Course Contents	Test Contents				
	Course Items	Test Items Represented				
		2058	2059	2060	2061	2062
Unit 3	Considerations in designing tests			L6		
3.1	Validity		O7		S2	O7
3.1.1	Content validity	O5			O8	O8
3.1.2	Criterion related validity					
3.1.3	Construct validity					
3.1.4	Face validity	O6				
3.1.5	Empirical validity					
	- concurrent validity					
	- predictive validity					
3.2	Reliability	O7		O6		
	- test-retest method					

	- alternative method					
	- split-half method					
	- internal consistency method					
3.2.2	Factors influencing reliability	S2				
	- length					
	- restriction of group performance					
	- methods of estimating reliability					
	- objectivity of scoring					
3.3	Administrability	O2				
3.4	Scorability					
3.5	Economy					
3.6	Wash back effect	O8	S2		L6	O4
Unit: 4	Constructing Tests					
4.1	Stages of test construction		S1			
4.1.1	Test specifications				S4	S2

4.1.2	Contents					
4.1.3	Weighting					
4.1.4	Time					
4.2	Criterion level of performance					
4.3	Piloting tests		O2			
4.4	Improving test items			O7		O5
4.5	Developing test for final administration					
Unit: 5	Testing Language Competence				O3	
5.1	Testing sound system		O4			O2
5.1.1	Testing segmental sounds					
5.1.2	Testing suprasegmental Features		S3			
5.2	Testing grammatical system					O1
5.2.1	Testing grammatical units					
5.2.2	Testing grammatical structure					



5.2.3	Testing grammatical function					
5.2.4	Testing grammatical categories					
5.2.5	Testing grammatical transformation					
5.3	Testing semantic system					
5.3.1	Testing vocabulary					
5.3.2	Testing other semantic features					
5.4	Testing Communicative functions		O3	O2S3		
Unit: 6	Testing language performance				O7	
6.1	Testing listening	S5				L6
6.1.1	Testing discrimination					
6.1.2	Testing comprehension		O6			
6.2	Testing speaking					S3
6.2.1	Testing discrimination					
6.2.2	Testing connected speech					

Unit: 7	Testing reading	O1				
7.1	Testing reading aloud					
7.2	Testing comprehension					
Unit: 8	Testing writing	S3				
8.1	Testing sentence level					
8.2	Testing supra-sentence level					
8.3	Testing creative writing		S4			
Unit: 9	Testing communicative abilities	L6			O4S3	S5
9.1	Testing communicative functions					
9.2	Testing of the pragmatic sensitivity					
	Total no of asked question in units: 3 to 9	6-obj 3-s.s. 1-s.L.	5-obj 4-s.s.	3-obj 1-s.s. 1-s.L.	4-obj 3-s.s. 1-s.L.	6-obj 3-s.s. 1-s.L.

The above table shows that in unit there are 14 language items from 3 to 3.6; in unit four there are 10 language items from 4 to 4.5; in unit five there are fourteen language items from 5-5.4. In unit six there are 7 language items fro 6 to 6.2.2; in unit seven there are 3 language items from 7 to 7.2; in unit eight there are 4 language items from 8 to 8.3

and in unit nine there are 3 language items from 9 to 9.2. Altogether, there are 55 language items from unit 3-9.

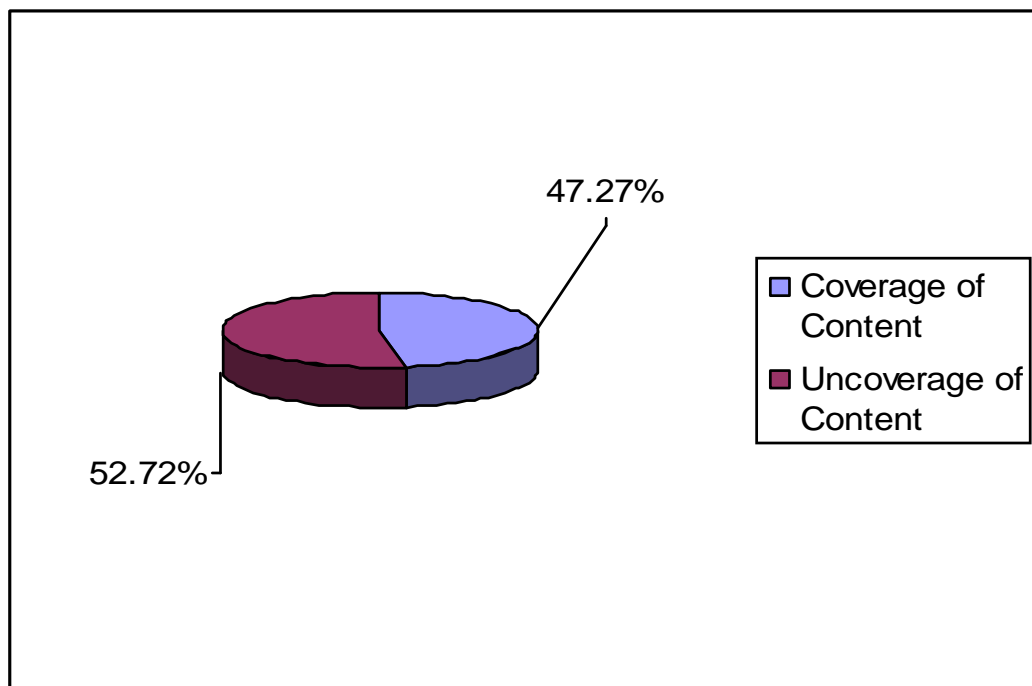
If we see diachronically among these 55 language items (i.e. including unit three to unit nine), there is no any most representative language item repeating each and every year from 2058 to 2062. The language item 3.6 (wash back effect) has been represented four years questions out of five years. The language items 3.1 (validity); 3.1.1 (content validity); 9 (testing communicative abilities) have been represented only three years questions out of five years. The language items 3.2(reliability); 4.1.1.(Test specifications); 4.4.(improving test items); 5.1 (Testing sound system); 5.4 (testing communicative functions); 6.1 (Testing listening); have been represented only two years questions out of five years. The language items-3 (considerations in designing tests); 3.1.4 (face validity); 3.2.2 (factor inflecting reliability and objectivity of scoring) 3.3 (Administrability); 4.1 (stages of test construction); 4.3 (Piloting tests); 5 (testing language competence); 5.1.2 (testing suprasegmental features); 5.2 (testing grammatical system); have been represented only in one year out of five years. Besides them, the language items which are mentioned in the table have not been represented in any year.

If we see synchronically, 10 questions (6 objectives, 3 subjective short questions and one subjective long question) were asked in 2058. In the case of objective questions that are asked from 3.1.1 (content validity); 3.1.4 (face validity); 3.2 (Reliability); 3.3 (Administrability); 3.6 (washback effect) and unit 7 (testing reading). In the case of subjective short questions that are asked from 3.2.2 (factors influencing reliability); 6.1 (testing listening) and unit 8 (testing writing). In the case of subjective long question that is asked from unit 9 (Testing communicative abilities).

In the year 2059, 9 questions (5 objectives and 4 subjective short questions) were asked. In the case of objective questions that are asked from 3.1 (validity); 4.3 (piloting tests); 5.1 (testing sound system); 5.4 (testing communicative functions) and 6.1.2 (testing comprehension). In the case of subjective short questions that are asked from 3.6 (wash back effect); 4.1 (stages of test construction); 5.1.2 (testing suprasegmental features) and 8.3 (testing creative writing). In the year 2060, 5 questions (3 objectives, one subjective short question and one subjective long question) were asked. In the case of objective questions that are asked from 3.6 (Reliability) 4.4 (Improving test items) and from 5.4 (testing communicative functions). In the case of subjective short question that is asked from 5.4. (Testing communicative functions). In the case of subjective long question that is asked from unit 3 (considerations in designing tests). In the year 2061, 8 questions (4 objectives, 3 subjective short questions and 1 subjective long question) were asked. In the case of objective questions that are asked from 3.1.1 (content validity); unit 5 (testing language competence); unit 6 (testing language performance) and unit 9 (testing communicative abilities). In the case of subjective short questions that are asked from 3.1 (validity); 4.1.1 (test specifications); and unit 9 (testing communicative abilities). In the case of subjective long question that is asked from 3.6 (washback effect). In the year 2062, 10 questions (6 objectives; 3 subjective short questions and 1 subjective long question) were asked. In the case of objective questions that are asked from 3.1 (validity); 3.1.1 (content validity); 3.6 (wash back effect); 4.4 (Improving test items); 5.1 (testing sound system) and from 5.2 (testing grammatical system). In the case of subjective short questions that are asked from 4.1.1 (test specifications); 6.2 (testing speaking) and fro 9 (testing communicative abilities). In the case of subjective long questions that is asked from 6.1 (testing listening)

In conclusion, the above presented table and description can be shown in the pie chart as follows:

**Figure No. 2**



From the above table and description the researcher found that there are 55 language items (including unit three to unit nine because unit 3-9 cover 30 marks) according to course content but the representation of the test items is only 26 language items. Twenty nine (29) Language items were neglected while constructing test items. It means the coverage of course in test contents in three to unit nine is 47.27 percent. 52.72% content were not covered in the question papers. According to the unit from 3-9 have low content validity because it (test papers) convey less than 50 percent course contents.

### 3.1.4 Examining course Representativeness in Unit Ten

**Table No: 6**

#### **Representation of test contents in terms of course contents in unit ten**

S.N.	Course Contents	Test contents				
	Course items	Test items represented				
		2058	2059	2060	2061	2062
Unit: 10	Interpretive language test scores					
10.1	Frequency distribution	S4				
10.2	Measurement of central tendency		S5	S4O4O 8		O6
10.3	Item analysis	O4	O8		S5	
10.4	Moderating language test					
10.5	Scoring techniques					
10.6	Reading test scores					
	Total no of asked question in unit ten	10bj 1 S.S	10bj 1 S.S	20 bj 1 S.S	1.S.S	10bj

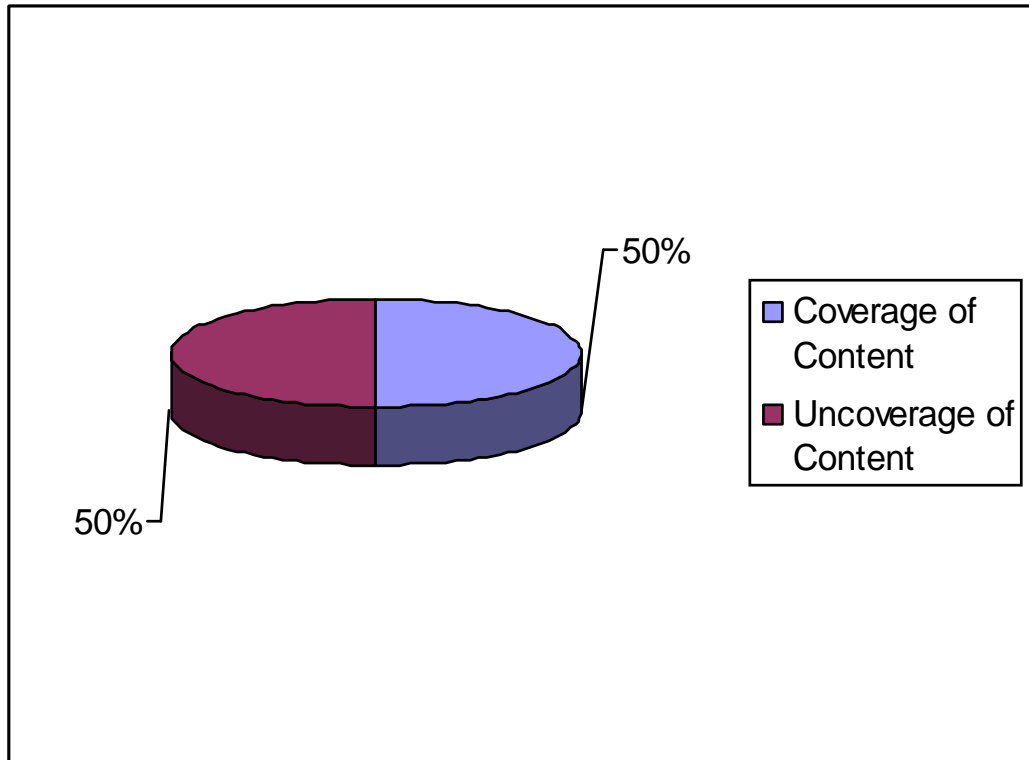
The above presented table shows that in unit ten there are 6 language items from 10.1 to 10.6. If we see diachronically among these six language items, no any language item was represented in every five

years' examination from 2058 to 2062. The language items 10.2 (Measurement of central tendency) and 10.3 (item analysis) have been represented in three years' examination out of five years. The language item 10.1 (frequency distribution) has been represented in only one year out of five years. The language items which were not represented at all are 10.4 (Moderating language test); 10.5 (scoring techniques and 10.6 (Reading test scores).

It we see synchronically, 2 questions (1 objective and 1 subjective short question) were asked from this unit in 2058. In the case of objective question that is asked from 10.3 (item analysis and in the case of subjective short question is asked from 10.1 (frequency distribution). In 2059, 2 questions (1 objective and 1 subjective short question) were asked. In the case of objective question that is asked from 10.3 (item analysis) and in the case of subjective short question that is asked from 10.2 (Measure of central tendency). In 2060, 3 questions (2 objective questions and 1 subjective short question) were asked. In the case of objective questions that are asked from the same language item 10.2 (Measurement of central tendency) and in the case of subjective short question that is also asked from the language item 10.2 (Measurement of central tendency). In 2061, only one subjective short question is asked from the language item 10.3 (item analysis). In 2062, only one objective question is asked from the language item 10.2 (Measurement of central tendency).

In conclusion, the above presented table and description can be shown in the pie chart as follows:

**Figure No. 3**



For the above table and description there are 6 language items in unit ten from 10.1 to 10.6 according to course contents but the representation of the test items is only 3 language items. Three (3) language items were neglected while constructing test items. It means the coverage of course contents in test contents in unit ten is 50 percent fifty (50) percent content were not covered in the question papers. Anyway, language testing papers have content validity according to unit ten because it (test papers) covers 50 percent course contents.



**Table No: 7**

**Examining Content Validity of the Test Papers On The Whole In  
Terms Of Coverage**

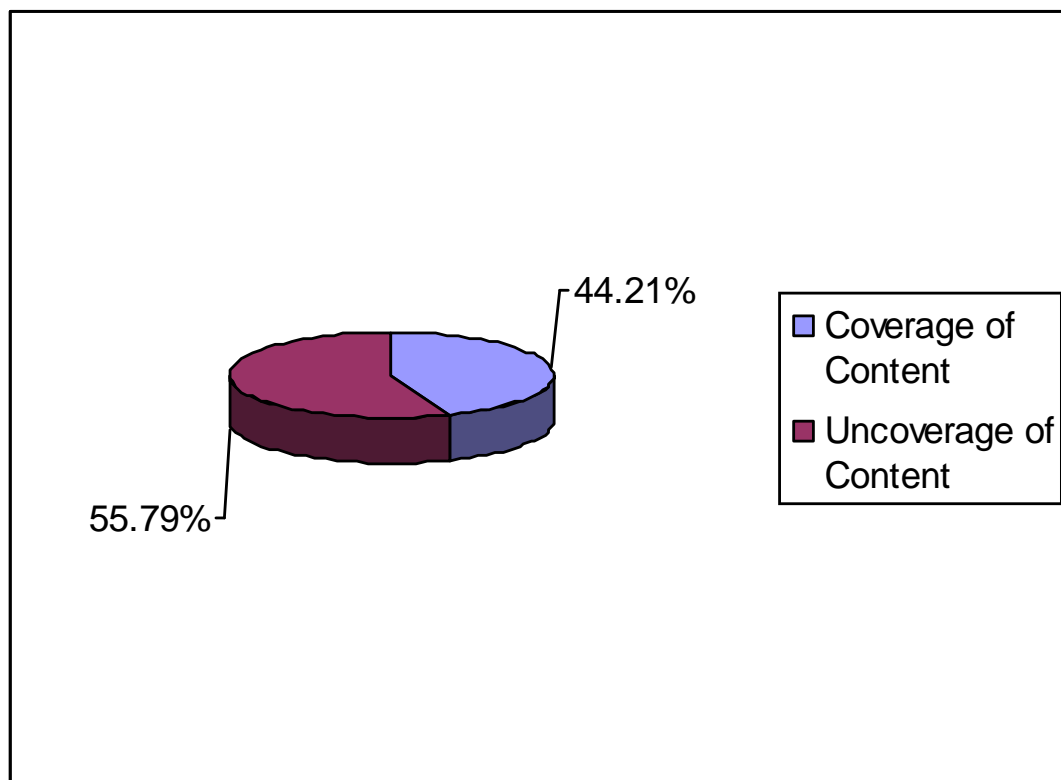
S.No.	Units	Course contents language items	Test contents language items	Test converge in percentage
1	1	34	13	38.23
	2			
	3			
	4			
	5	55	26	47.27
2	6			
	7			
	8			
	9			
3	10	6	3	50
Total		95	42	44.21

The above table indicates that on the whole in the ten units, there are 95 language items in the whole language testing course at M.Ed. second year. Out of 95 course contents language items the representation of the test contents language items are 42. Fifty three (53) language items were neglected while constructing the test items. It means the coverage of

contents in whole units is 44.21 percent. Fifty five (55.79) percent contents were not covered in the question papers.

In conclusion, the above presented table and description can be shown in the pie-chart as follows:

**Figure No. 4**



As the description mention that the content validity of the test papers as a whole is not satisfactory because it covers less than 50 percent course contents. The unit wise, distribution of marks is not scientific because in unit one and two covers 10 marks. It does not indicate that which one is more important or both units covers equal status. And in case of unit 3-9 (as a whole) covers 30 marks. The syllabus designers do not focus which one is important and which one is less important. And last unit 10 marks. This distribution of marks creates problems to test paper designers. On the other hand, the model of test paper is not also

scientific. It (test paper) should contain 8 objective questions, 5 subjective short questions and 1 subjective long question. Altogether there are 14 questions (including subjective and objective). And there are 95 language items. This is the reason that the test designer should not cover the course content properly or the model of test paper should be changed.

## **3.2 WEIGHTING PERCENTAGE**

### **3.2.1 Comparison between Weighting of the Course Contents and Weighting of the Test Contents**

This is the second part of this chapter which deals with an analysis of the proportionality of weighting of the language testing test papers for the purpose of examining content validity of the language testing test, the researcher compared / examined whether or not the marks weighting in the course contents according to syllabus is proportional with the weighting of the test contents for this purpose, the researcher examined language testing 5 years (2058-2062) test papers as a whole.

**Table No. 8**

**Examining the content validity of the test papers on the whole  
in terms of weighting**

Weighting of the course contents			Weighting of the test papers				
S.N	Units	Full weight age	2058	2059	2060	2061	2062
1	1 2	10	1+6	2+6+12	3+18	4+6	1+12
2	3 4 5 6 7 8 9	30	6+18+12	5+24	3+6+12	4+18+12	6+18+12
3	10	10	1+6	1+6	2+6	6	1
	Total marks	50	8+30+12	8+30+12	8+30+12	8+30+12	8+30+12
	Total		50	50	50	50	50

**Note:** In 2058 (from unit 1 and 2) 1+6 =1 objective question which contains 1 mark and 1 subjective short question which contains 6 mark.

**Here :-** 12 means subjective long question which contains 12 marks.

The above table indicates that out of 50 marks of the whole language testing course at M.Ed. level, unit one and two have carried 10 marks according to syllabus. In comparison between weighting of the course contents to weighting of the test paper, the researcher found that though unit one and two are said to be carrying 10 marks, it is not found to be followed in test papers because in 2058, there asked 7 marks, (1 marks for 1 objective question and 6 marks for 1 subjective short question). In 2059, there asked 20 marks (2 marks for 2 objective questions, 6 marks for 1 subjective short question and 12 marks for 1 subjective long question). In 2060, there asked 21 marks (3 marks for three objective questions and 18 marks for 3 subjective short question and 18 marks for 3 subjective short question). In 2061, there asked 10 marks (4 marks for 4 subjective question and 6 marks for 1 subjective short question). In 2062, there asked 13 marks (1 marks for 1 objective question and 12 marks for 2 subjective short questions). In unit 3 to 9 carry 30 marks weightage according to syllabus but the test papers have been carried different weightage as follows: in 2058, 36 marks (6 marks for 6 objective questions, 18 marks for 3 subjective short questions and 12 marks for 1 subjective long question). In 2059, 29 mark (5 marks for objectives questions and 24 marks for 4 subjective short questions). In 2060, 21 marks ( 3 marks for 3 objective questions, 6 marks for 1 subjective short question and 12 marks for 1 subjective long question). In 2061, 34 marks (4 marks for 4 objectives question, 18 marks for 3 subjective short questions and 12 marks for 1 subjective long question). In 2062, 36 marks (6 marks for 6 objective questions, 18 marks for 3 subjective short question and 12 marks for 1 subjective long question). It seems this kind of distribution of marks for test papers rather followed the syllabus weightage.

At last unit ten carries 10 marks weightage according to the syllabus but the test papers have been carried different weightage as follows, in 2058, there was asked 7 marks (1 marks for one objectives question and 6 marks for 1 subjective short question). In 2059, there was asked 7 marks (1 mark for 1 objective question and 6 marks for 1 subjective short question). In 2060, there was asked 8 marks (2marks for 2 objective question and 6 marks for 1 subjective short question). In the case of 2061, there was asked 6 marks (6 marks for 1 subjective short question). And in the case of 2061, there was asked 1 mark (1 mark for one objective question). This shows that these weighting marks only represented nearly half marks in the years 2058, 2059, 2060 and 2061 but, in the year 2062, only 1 mark was asked. It shows the negligence in the part of the question setters which may lead the lack of content validity and negative washback as well.

As a whole during 5 years (2058 to 2062) in language testing papers at M.Ed. second years, the researcher found that in totality unit one and two which carries 10 marks but the question setters did not care according to schedule in syllabus. In 2058 the distribution of mark is under weight but in 2059, 2060, 2061 and 2062 have over weighting which secure more marks according to Scheduled in syllabus. In the case of unit 3 to 9 carries 30 marks. In 2059 and 2060 have under weighting (it means below according. In the case of unit ten which carries 10 marks according to Schedule in syllabus, all the five years (i.e. 2058, 2059, 2060, 2061and 2062) have in under weighting. It shows the negligence in the part of content validity and negative wash back as well. Though it does not mention the fixed course contents marking (weighting), the test setters also do not care according to Schedule in syllabus.

So far we have seen the course contents weighting and tests weighting at language items which is one aspect of content validity, there is no satisfactory finding. Some course language items are found to be over weightaged and some course language items are found to be under weightage. It seems that there was no specific norm for asking question. Thus, from the above data and description the researcher found that Language Testing papers lack content validity in terms of weighting of the course contents.

## CHAPTER - IV

### FINDINGS, PEDAGOGICAL IMPLICATIONS AND RECOMMENDATIONS

What are the findings of the study is the main concern of this chapter. After the analysis of the question papers from different angles, the researcher has found the following findings.

#### 4.1 FINDINGS

The major finding of this research work is as follows:-

According to the coverage/representativeness principle, the language testing tests have low content validity. It is because out of 95 language items in totality of the course, the test items have represented 42 language items, i.e. 44.21 percent during 5 years (2058 to 2062).

On the other hand, according to the weighting principle, the language testing tests have low content validity. It is because there was not following any norm or tendency in the distribution of marks in the test papers what it was given in the weighting schedule of the syllabus. For example, in 2062 from the unit 10, it has carried 1 mark in the test papers. Where as its weightage is 10 marks according to the syllabus.

The unit specific findings of this research work are as follows:

1. In terms of coverage / representativeness comparison between course contents and test contents.
  - i. In unit one and two, out of 34 (unit one-7 and unit two-27) language items of course contents, test items represented 13 language items during 5 years (2058 to 2062). Twenty



one (21) language items were untouched. It means the coverage of course contents is 38.23 percent. In conclusion, language testing tests have low content validity in unit one and two because test contents have been represented less than 50% course contents.

ii. In unit three to nine, out of 55 language items of course contents, test items represented 26 language items during 5 years. Twenty nine (29) language items were untouched. It means the coverage of course contents is 47.27 percent. In conclusion, language testing tests have low content validity in unit three to unit nine because test contents have been represented less than 50% course contents.

iii. In unit ten out of 6 language items of the course contents, test items represented 3 language items. Three language items were untouched, it means the coverage of course contents is 50 percent. In conclusion, language testing tests have nearer to content validity in unit ten because test contents have been represented 50% course contents.

2. In terms of weightage / weighting comparison between course contents weighting and the test contents weighting.

i. In unit one and two, out of 10 marks weightage according to syllabus, the test papers have been carried out different weightage as follows:

In 2058: 7 marks from 1 objective and one subjective short question. In 2059: 20 marks from 2 objectives 1 subjective short question and 1 subjective long question. In 2060: 21 marks from 3 objective question and 3

subjective short questions. In 2061, 10 marks from 4 objective questions and 1 subjective short question. In 2062: 13 marks from 1 objective question and 1 subjective long question.

Thus, as a whole in unit one and two it is found that in 2058 the marks weighting in the test papers were under weightage, in 2059,2060 and 2062, the marks weight age in the test papers were over weighting and in 2061 carries 10 marks as according to the syllabus. It is found that in all the years the marks weighting in the test papers were not proportional to the weighting of its course contents. This shows that the language testing test papers lack content validity according to unit one and two.

- ii. In unit three to nine out of 30 marks weightage of course contents, the weightage of the contents were as follows.

In 2058:36 marks from 6 objective questions 3 subjective short questions, and 1 subjective long question. In 2059, 29 marks from 5 objective questions and 4 subjective short questions. In 2060: 21marks from 3 objective questions, 1 subjective short question and 1 subjective long question. In 2061:34 marks from 4 objective questions, 3 subjective short questions and 1 subjective long question. In 2062:36 marks from 6 objective questions, 3 subjective short questions and 1 subjective long question.

Thus, as a whole in unit three to nine it is found that in all the years the marks weighting in the test papers were not proportional to the weighting of its course contents. It is found that in 2058, 2061 and 2062 marks weighting in the

test papers were over weightage and in the years 2059 and 2060, the marks weighting in the test papers were under weight age. Thus, this shows that the Language Testing test papers lack content validity according to unit three to nine.

- iii. In unit ten, out of 10 marks weightage of course contents the weightage of the test contents were as follows.

In 2058:7 marks form 1 objective question and 1 subjective short question. In 2059:7 marks form, 1 objective question and 1 subjective short question. In 2060:8 marks form 2 objectives and one subjective short question. In 2062:1 marks form 1 objective question.

Thus, as a whole in unit ten it is found that in all the years the marks weighting in the test papers were not proportional to the weighting of its course contents and the language testing test papers lack content validity according to unit ten.

3. It has been found that the language testing test have low content validity in terms of coverage or representation of the course contents and low content validity in terms of weighting of the course contents syllabus.
4. According to the weighting of the course contents, it is argued that the language testing tests have not tested what it has been supposed to test in the testees because there is not found to be obeyed any weighting system of the course contents in the test contents. And according to the representation of the course contents, it is argued that the language testing tests have not tested what it has been supposed to test in the testees because the test items represented less than 50 percent course items.

5. More variation is found in asking questions in all the types of questions (objective, subjective short, and subjective long questions) in the different years in same unit as well. For example, in unit 10 there was asked only objective questions and subjective short question, but there was not asked any long questions during 5 years.

## **4.2 RECOMMENDATIONS**

On the basis of the findings of the research work and the researcher's own intuition some recommendations are made which are listed below.

1. It has been found that Language Testing tests have represented only 44.21 percent course contents as a whole. It is low content validity in terms of representative principle. Thus language testing tests should be covered/represented more than 60 percent course items to be good content validity.
2. To get a test high content validity the question setters have to give the right of question format in his own but don't confine him to follow the ready made format.
3. A test would be high content validity if it follows both representative principle and weighting principle. Thus Language Testing test should follow both principles: representativeness and weighting to get a test high content validity.
4. There should be prepared any fixed criteria, for example 'specification table' for asking the different question types (objective, subjective short, subjective long questions) in terms of unit wise. Asking different types of questions differently in

different years create harmful effects on the validity of the test papers as well as to the testees. Thus, Language Testing tests should follow fixed criteria about how many and which type of questions will be asked from each unit in each year.

5. Instead of repeating the same test items in each and every year examination, it should be touched the different and more language items in the examination. Thus in language testing test, the repetition of the same language items should be excluded as possible. It increases the representation of the course contents and activates the testees as well.
6. While constructing questions all the units should be given emphasized equally.
7. To get a test high content validity, the question setters have to have minute study on the course objectives, course contents and weighting of the course contents before developing the question papers. The question papers have to be piloted if possible.
8. The office of the controller of examinations should have positive spirit and should play a vital role in improving the Language Testing tests as a whole. It should be cared about not only content validity of the test papers but also other types of validity as well. For this the concerned authority should conduct some seminars, workshops and conferences.

## REFERENCES

- Aryal, Kamal Prasad (2005), A Study on Content Validity of Grade Twelve Compulsory English Examination 2061. An unpublished M.Ed. thesis, Kathmandu: Tribhuvan University.
- Batala, K. B (2004), Validation of the SLC English Examination, An unpublished M .Ed Thesis, Tribhuvan University.
- Bachman, L. F. (1998), Fundamental Considerations in Language Testing, Oxford University Press.
- Bhattarai, G.R. (2001), A Thematic Analysis of Research in Reports. Ratna Pustak Bhandar, Kathmandu.
- Bhattarai, Jharna (2005), The Content Validity of Compulsory English Textbook for Grade Eight, An unpublished M.Ed. Thesis Kathmandu: T.U.
- Cambridge International Dictionary of English (1977:795).
- Davies, A. (1968), Language Testing Symposium: A Psycholinguistic Approach, London. Oxford University Press.
- Harrison, A. (1991), Language Testing Handbook London, ELTS, Longman.
- Hatch, E. and Farthady, H. (1982), Research Design and Statistics for Applied Linguistics. Rowley, Mass, Newbury House Publishers, Inc.

Hughes, Arthur (1989), *Testing for Language Teachers*. Cambridge University Press.

Heaton, J. B. (1998) *Writing English Language Tests* Longman.

Khaniya, T. R. (2000), Washback: Emerging Validity, In *Journal of NELTA*. Vol. 5, Kathmandu: NELTA.

Khaniya, T. R. (2005), *Examination for Enhanced Learners*, Lalitpur: Millennium Publication.

*Longman Dictionary of Language Teaching and Applied Linguistics* (1999).

Kumar, R. (1996), *Research Methodology*. New Delhi: SAGE Publication.

Lado, Robert (1961), *Language Testing: The Construction and Use of Foreign Language Tests* London, Longman.

Lado, R. (1999), *Language Testing*: Longman.

Neupane, Sudeep (2005), *The Context Validity of English Textbook for Grade Seven*, an Unpublished M.Ed. Thesis Kathmandu T.U.

[www.google.com](http://www.google.com).

## APPENDICES

### APPENDIX-I

# TRIBHUVAN UNIVERSITY

2058

Roll No.:.....

Master Level /II year/ EDUCATION

Language Testing (Eng.Ed.505)

Group "A"

8

A. Tick (✓) the best answers.

4

1. Which one of the following cannot be a test of reading?
  - a. Picture and sentence matching
  - b. Reading a short text
  - c. Composition test
  - d. Completion test
2. Which one of the following does not come under "test rubrics"?
  - a. Organization of items
  - b. Specification of the test
  - c. Time allocation
  - d. Instructions
3. Which of the following is not a goal-based test?
  - a. Proficiency test
  - b. Achievement
  - c. Placement test
  - d. Written test
4. Discrimination index (DI) of an item indicates
  - a. How difficult an item is for the students
  - b. How easy an item is for the students
  - c. How effective an item is to group students in terms of their ability
  - d. How effective an item is to predict the future performance of the students



**B. Fill in the gaps with appropriate words.**

**4**

5. If the content of a test contains a representative sample of the objectives of the course, the test is said to have .....

6. If the test looks as if it measures what it is supposed to measure, the test is said to have .....

7. If two markers award similar scores to a piece of written work, the test is said to be .....

8. If a test appears to have positively influenced what goes on in the classroom teaching, the test is said to have .....



**TRIBHUVAN UNIVERSITY**

**2058**

**Master Level /II year/ EDUCATION**

**Full Marks: 50**

**Language Testing (Eng.Ed.505)**

**Times: 3 hrs**

*Candidates are required to give their answers in their own words as far as practicable.*

*The figure in the margin indicate full marks.*

**Attempt all the questions.**

**Group “B”**

**5 × 6 = 30**

1. Write a short note on proficiency test.
2. What is reliability of a test? What factors are responsible for influencing the reliability of a test?
3. What are the methods of testing the writing skill? Explain.
4. What is descriptive statistics? Why and how do you present the frequency distribution of the scores?
5. How do you test listening skill? Give examples.

**Group “C”**

6. Discuss what components of language are tested by a communicative language test, and what features of language that the tasks to be included in the communicative test must reflect.



# TRIBHUVAN UNIVERSITY

2059

Master Level /II year/ EDUCATION

Full Marks: 50

Language Testing (Eng.Ed.505)

Times: 3 hrs

*Candidates are required to give their answers in their own words as far as practicable.*

*The figure in the margin indicate full marks.*

**Attempt all the questions.**

## Group "B"

5 1/6 = 30

1. Describe in brief the stages of test construction.
2. Explain the concept of 'wash back effects' and point out the positive and negative aspect of such effects on language teaching.
3. Design an objective test that the knowledge of the suprasegmental features of the English sound system. The full mark of the test is 6 and the testees are B.Ed. students with English major.
4. What is 'creative writing'? What type of test would you use to test creative writing? Give reason along with examples.
5. An objective test consisting of 20 items was given to a group of 10 students. The result of the test is as follows:

Student	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Score out of 20	10	15	5	11	16	11	8	10	3	11

Now answer the following question base on the test result:

- a) What is the main score?
- b) What is the mode score?
- c) What is the median score?
- d) What is the range of the scores?

## **Group "C"**

**12**

6. Name the various types of 'goal-base' tests and explain how they differ from one another.

APPENDIX-III

<b>TRIBHUVAN UNIVERSITY</b>	
<b>2060</b>	<b>Roll No.:.....</b>
<b>Master Level /II year/ EDUCATION</b>	
<b>Language Testing (Eng.Ed.505)</b>	

**Group "A"**

**8**

**A. Tick ( $\checkmark$ ) the best answers.**

**4**

1. A dictation test is basically a test of a. listening skills c. reading skills	b. speaking skills d. writing skills	
2. "Underline the subject in the sentence 'John married Mary in Kathmandu'. a. structure c. category	b. function d. transformation	
3. In an objective test an extremely difficult item separates a. the very good students from the good students b. the very good students from the average student c. the very good students from the poor students d. the very good students from the very poor students		
4. Which of the following does not refer to a central tendency of the distribution of test scores? a. mean c. median		b. mode d. range

**B. Write True or False against each statement**

**4**

5. A cloze test is an integrative test .....
6. A test is said to be valid if it gives the same result no matter who scores it. ....
7. A pilot test is carried out after the actual test. ....
8. The standard deviation is a technique of showing the spread of scores. ....

# TRIBHUVAN UNIVERSITY

2060

Master Level /II year/ EDUCATION

Full Marks: 50

Language Testing (Eng.Ed.505)

Times: 3 hrs

*Candidates are required to give their answers in their own words as far as practicable.*

*The figure in the margin indicate full marks.*

**Attempt all the questions.**

## Group “B”

5 × 6 = 30

1. Discuss the role of testing in language education.
2. Distinguish between 'norm-referenced test' and 'criterion-referenced test'.
3. Construct a cohesive and coherent dialogue which includes expressions that illustrate the following communicative functions:
  - a. Introducing
  - b. Apologizing
  - c. Asking to repeat
  - d. Expressing inability
  - e. Seeking information
  - f. Extending an invitation

The sequencing of the communicative functions needn't be the same as the one given above. Identify the expressions corresponding to the communicative functions they serve.

4. Explain, with a hypothetical data, how (he standard deviation of a test result is calculated.
5. Design **two** different types of *subjective* questions and **four** different types of *objective* questions of your choice to test the comprehension of the following passage:

*Nepal is a land of unparalleled variety. Imagine a rectangle, 800 km by 240 km, divided lengthwise into three strips. The northernmost strip is the Himalaya, meaning "abode of snow". It includes eight of the ten highest mountains in the world. The Himalayan region is sparsely settled by people who speak languages of the Tibeto-Burman family and practice Tibetan Buddhism. The southernmost region, which is the narrowest of the three strips, is called the Tarai. It is an extension of the Gangetic plain of northern India, a jungle with elephants, rhinoceroses and tigers. These inhabitants contrast markedly with the yaks and snow leopards less than 160 km to the north. This area is populated by people who speak Indo-European languages and practice Hinduism. Between the two outer strips lies an interface region of hills and valleys. The inhabitants speak languages of both the Tibeto- Burman and Indo-European families and generally practice Hinduism with many Buddhist and shamanistic influences. This region is the unexpected treasure of Nepal.*

**Group "C"**

12

6. Discuss the qualities of a good test as comprehensively as possible.

APPENDIX-IV

**TRIBHUVAN UNIVERSITY**

**2061**

**Roll No.:.....**

**Master Level /II year/ EDUCATION**

**Language Testing (Eng.Ed.505)**

**Group "A"**

**8**

**A. Tick ( $\checkmark$ ) the best answers.**

**4**

1. TOFEL is an example of  
a. an achievement test                      b. a proficiency test  
c. a placement test                          d. a prognostic test
2. The idea that language ability is essentially unitary or holistic can be phrased as "A single language competence". This idea was predominant in language learning / teaching in  
a. 1960s                      b. 1970s                      c. 1980s                      d. 1990s
3. Which of the following tests does not test-overall competence?  
a. Cloze test                                      b. C - test  
c. Multiple choice test                          d. Dictation test
4. Which one of the following was not included as a separate component of communicative competence in Canale and Swain (1980)?  
a. grammatical competence    b. discourse competence  
c. strategic competence              d. socio-linguistic competence

**B. Fill in the gaps with appropriate works.**

**4**

5. A test designed to assess students' level of language ability so that they can be placed ift the appropriate course is a/an ..... test
6. A test used to find out the strengths and weaknesses of the students is called a ..... test
7. A distinctive feature of performance test is that it assesses ..... in addition to linguistic competence.
8. If a test covers a representative sample of the skills and abilities of the course given, the test is said to have .....

# TRIBHUVAN UNIVERSITY

2061

Master Level /II year/ EDUCATION

Full Marks: 50

Language Testing (Eng.Ed.505)

Times: 3 hrs

*Candidates are required to give their answers in their own words as far as practicable.*

*The figure in the margin indicate full marks.*

**Attempt all the questions.**

**Group “B”**

**5 × 6 = 30**

1. Highlight the features and functions of 'Achievement Test'.
2. What is validity? Discuss the content validity of a test.
3. What are the major features that a communicative task for testing language ability should reflect? Discuss.
4. What is a test specification? Discuss in detail the major elements of the test specification.
5. What is facility Value of test item? Discuss its significance.

**Group “C”**

**12**

6. Wash back effect of an examination is said to be its inherent part Justify this statement Also, discuss how you would construct a test ensuring that it has positive wash back effect.



## APPENDIX-V

### TRIBHUVAN UNIVERSITY

2062

Roll No.:.....

Master Level /II year/ EDUCATION

Language Testing (Eng.Ed.505)

#### Group "A"

8

**A. Tick (✓) the best answers.**

4

- |  |
|--|
| 1. Which one of the following is not the rest of grammar?<br>a. Rearrangement item<br>b. Error recognition item<br>c. Composition item<br>d. Completion item   |
| 2. Phoneme discrimination test is a test of<br>a. reading skill                      b. writing skill<br>c. listening skill                      d. speaking skill   |
| 3. A psychometric test is<br>a. a test of grammatical awareness<br>b. a test for testing psycholinguistics<br>c. a test for testing one things at a time<br>d. a test for testing two or more things at a time     |
| 4. There is nothing wrong in preparing students for examination if<br>a. the exam are practical                      b. the exam are good<br>c. the exams are short                      d. the exams are relevant |

**B. Fill in the gaps with appropriate works.**

4

- |  |
|--|
| 5. The power of a test item in separating the more form the less capable students on some latent attitude is called..... |
| 6. .... is used to refer to the most frequent measure of variability.  |
| 7. If the test covers a reprehensive sample of the skills and abilities of the course given the test is said to have.    |
| 8. If a test measures what it is supposed to measure and nothing else, the test is said to be.                           |

# TRIBHUVAN UNIVERSITY

2062

Master Level /II year/ EDUCATION

Full Marks: 50

Language Testing (Eng.Ed.505)

Times: 3 hrs

*Candidates are required to give their answers in their own words as far as practicable.*

*The figures in the margin indicate full marks.*

**Attempt all the questions.**

## **Group "B"**

**5 × 6 = 30**

1. Write a short note on Achievement Test.
2. What is test specification? Discuss why a test specification is necessary for developing a good test?
3. What are the methods of testing the speaking skill? Discuss briefly.
4. Distinguish between discrete point test and integrative test.
5. Discuss why it is important to test the communicative ability of the learners.

## **Group "C"**

**12**

6. What is testing the listening skill? How is it tested? Discuss the practicality of testing the listening skill in the Nepalese context.

## **APPENDIX-VI**

### **The Syllabus of Language Testing at M. Ed. Second Year**

#### **Course Contents in Detail**

##### **Unit I: Introduction**

- 1.1 Language education and testing
- 1.2 Testing language competence
- 1.3 Testing language performance
- 1.4 Historical perspective
  - 1.4.1 Traditional approach
  - 1.4.2 Structural approach
  - 1.4.3 Communicative approach

##### **Unit II: Classification of Language Tests**

- 2.1 Goal-based tests
  - 2.1.1 Proficiency test
  - 2.1.2 Achievement test
  - 2.1.3 Diagnostic test
  - 2.1.4 Prognostic test/aptitude test
  - 2.1.5 Placement test
- 2.2 Medium-based test
  - 2.2.1 Oral test
  - 2.2.2 Written test
- 2.3 Mode-based test
  - 2.3.1 Objective test
  - 2.3.2 Subjective test
- 2.4 Aspect-based tests
  - 2.4.1 Pronunciation test
  - 2.4.2 vocabulary test
  - 2.4.3 Communicative function test
- 2.5 Skill-based tests
  - 2.5.1 Test of listening
  - 2.5.2 Test of speaking

2.5.3 Test of reading

2.5.4 Test of writing

2.6 Approach-based tests

2.6.1 Discrete-point test

2.6.2 Integrative test

2.7 Reference-based Test

2.7.1 Norm-reference test

2.7.2 Criterion-reference test

### Unit III: Considerations in Designing Tests

3.1 Validity

3.1.1 Content validity

3.1.2 Criterion validity

3.1.3 Construct validity

3.1.4 Face validity

3.1.5 Empirical validity

- concurrent validity

- predictive validity

3.2 Reliability

3.2.1 Methods of determining reliability

- test-retest method

- alternative method

- split-half method

- internal consistency method

3.2.2 Factors influencing reliability

- length

- restriction of group performance

- methods of estimating reliability

- objectivity of scoring

3.3 Administrability

3.4 Scorability

3.5 Economy

3.5 Washback effects

### Unit IV: Constructing Tests

- 4.1 Stages of test construction
  - 4.1.1 Test specifications
  - 4.1.2 Contents
  - 4.1.3 Weighting
  - 4.1.4 Time
- 4.2 Criterial level of performance
- 4.3 Piloting tests
- 4.4 Improving test items
- 4.5 Developing tests for final administration

#### Unit V: Testing Language Competence

##### Testing sound system

- Testing segmental sounds
- Testing suprasegmental features

##### Testing grammatical system

- Testing grammatical units
- Testing grammatical structure
- Testing grammatical function
- Testing grammatical categories
- Testing grammatical transformation

##### Testing semantic system

- Testing vocabulary
- Testing other semantic features

##### Testing communicative functions

#### Unit VI: Testing Language Performance

##### Testing listening

- Testing discrimination
- Testing comprehension

##### Testing-speaking

- Testing discrimination
- Testing connected speech

#### Unit VII: Testing Reading

- 7.1 Testing reading aloud
- 7.2 Testing comprehension

## Unit VIII: Testing Writing

Testing at sentence level

Testing at supra-sentence level

Testing creative writing

## Unit IX: Testing Communicative Abilities

9.1 Testing communicative functions

9.2 Testing of the pragmatic sensitivity

## Unit X: Interpretive Language Test Scores

10.1 Frequency Distribution

10.2 Measurement of central tendency

10.3 Item-analysis

10.4 Moderating language test

10.5 Scoring techniques

10.6 Reading test scores

## **Instructional Technique**

Lecture and discussion

Demonstration

Explanation and illustration

Project work

Group and individual work

Self-study

## **Assessment Technique**

Written examination

# The End