

CHAPTER ONE

INTRODUCTION

1.1 General Background

Language, the sole property of all and only human beings, is a means of human communication through which human beings exchange their ideas, feelings, thoughts desires and emotions and so on. It is species-specific to mankind, which is assumed to be as old as human race itself. Lyons (1970, in Syal and Jindal 2005:5) says that "Languages are the principal systems of communication used by particular group of human beings within the particular society (Linguistic community) which they are members". Similarly, Richards et al.(1990:196) define language as "the system of human communication which consists of the structural arrangement of sounds (or their written representation) into larger units, e.g., MORPHEMES, WORDS, SENTENCES, UTTERANCES". Various scholars have defined language differently but none of these definitions are absolutely complete in themselves. So defining a language depends upon the perspective a person tries to base on. However, different definitions given by different scholars surely share some common characteristics of a language. On the basis of the common characteristics found in different definitions, a language can be considered as an arbitrary voluntary vocal system of human communication.

Simply speaking, language teaching means teaching of language for communication. In the past, language was viewed as a set of rules, and language teaching was regarded as the teaching of those rules. Before the advent of modern linguistics, the students who were studying a foreign language were given the rules of foreign language in their mother

tongue and were asked to follow them. They were expected to translate the literary text from or into the mother tongue. Writing was given priority than speaking.

The book of Ferdinand de Saussure (1916) entitled ' Cours de linguistique generale' brought a change in the view of language. The change in the view of language also brought change in language teaching as well. Speech was given more emphasis than the written form of language, and consequently, direct method came into existence. Similarly, audio lingual method was developed in the field of language teaching.

The publication of Chomsky's book 'Syntactic Structures' in 1957 brought a revolution in the field of linguistics. In the book, he defined language as "a set (finite or infinite) of sentences, each finite in length and constructed out of a finite set of elements". Chomsky had demonstrated that the current structural theories of language were incapable of accounting for the fundamental characteristics of language– the creativity and uniqueness of sentences (quoted in Richards and Rodgers, 1986).

1.1.1 Semantics

While talking about different systems of linguistics, we put Phonology in the beginning which is followed by Grammar and then Semantics at the end. However, Semantics has been one of the most neglected fields in linguistics since only recently serious interest has been taken in its various problems.

The history of Semantics goes back to the American Philological Association which introduced a paper entitled 'Reflected Meanings: a

point in Semantics'. In 1990, there appeared Breal's book, 'Semantics: Studies in the Science of Meaning'. It was one of the earliest books on linguistics which treated Semantics as the science of meaning, and that it was not primarily concerned with the changes of meaning from a historical point of view. Yet, the term 'Semantics' did not catch on for sometime. Later in 1923, two more books were published on Semantics: 'The Meaning of Meaning' by C. K. Ogden and I. A. Richards, and 'The Problem of Meaning in Primitive Language' by the anthropologist Malinowski. Semantics was popularized in the 1930s and 1940s by a school of thought—that of general semantics— which holds that the study of communicative processes can be a powerful force for good in the resolution of human conflict, whether on an individual, local or international scale (Leech, 1981). However, Semantics is neither just the study of change in word meaning through time as people falsely think nor it is something that can be used to mislead people. Instead, Semantics is a systematic study of what meaning is and how it operates. Today it has been a full-fledged discipline because without taking account of meaning the study of language is incomplete.

1.1.2 Pragmatics

The origin of Pragmatics goes back not to linguistics but to the philosophical writings. Morris used the term in 1938 for the first time when he distinguished three different branches of inquiry under semiotics (semiotic as Morris preferred). The three branches were syntactics (syntax), being the study of 'the formal relation of signs to one another', semantics, the study of 'the relations of signs to the object to which the signs are applicable', and pragmatics, the study of 'the relation of signs to interpreters' (Levinson, 1994). Morris (1938) took Pragmatics in much

broader sense than today because according to him, Pragmatics deals with the entire psychological, biological and sociological phenomenon which occurs in the function of signs.

The broader scope of pragmatics was narrowed down by Carnap (1938) (quoted in Levinson 1983:3). He tried to distinguish the area of Syntax, Semantics and Pragmatics. According to him, "If in an investigation explicit reference is made to the speaker, or to put it in more general terms, to the user of the language, then we assign it (the investigation) to the field of pragmatics...if we abstract from the user of the language and analyze only the expression and their designata, we are in the field of semantics. And, finally, if we abstract from the designata also and analyze only the relations between the expressions, we are in (local) syntax". Thus, he emphasized the importance of participant's role in a piece of conversation, and consequently, those aspects of language were emphasized which were needed for understanding the role of a speaker and setting or situation.

1.1.3 Language Testing

Teaching and testing are like two sides of a coin. Testing is as important as teaching. They are taken as inseparable phenomena. In other words, teaching and testing are so closely interrelated that it is virtually impossible to work in either field without being constantly concerned with the other (Heaton, 1988:5). 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 (Khaniya, 2005:1). In order to ensure that the teaching is effective and if not, put more efforts to make it effective, testing is used in the classroom or after the classroom teaching.

Language teaching and testing function like the combination of a pick and shovel to dig deep into the language education. Language testing plays a very important role in language teaching. It will help to locate the precise areas of difficulty encountered by the class or by the individual student. Unless the teacher is able to identify and analyze the errors a student makes in handling the target language, he or she will be in no position to render any help at all through appropriate anticipation-remedial work and additional practice.

Although the history of language testing goes back to the history of language teaching, it was not taken as a separate discipline in the past. There were different factors influencing in language teaching and testing. The emphasis on what is to be tested has been changing through different stages over the years. How it has been emphasized differently over time can be discussed as different approaches to language testing.

The pre-discrete point approach to language testing was based on the assumption that no special expertise was required for testing i.e. any teacher could do it. Language testing was viewed as entirely subjective in the sense that the total assessment was based on the subjective judgment of the examiner. Aspects of language to be tested are the abilities to translate and write open ended essays.

The discrete point approach to language testing came into existence since the pre-discrete point test was severely criticized for not being reliable and valid. This approach is based upon the assumption that "knowledge of the elements of a language is equivalent to knowledge of the language" (Morrow, 1979:145 quoted in Khaniya 2005: 13). It is due to the influence of structuralism in language testing. Language was considered as a set of habits. This approach to testing put emphasis on

reliability more than needed, may be at the cost of validity. The format of the test is composed of short answers and multiple choice items.

The integrative approach to testing emerged when discrete point test was followed by its shortcomings. The emergence of integrative approach to language testing is based on the assumption that "knowledge of a language is more than just the sum of a set of discrete parts", and also on the belief that it is only the integrative test that gives a true measure of language ability (Spolsky, 1978:viii). Although different approaches to language testing overlap in terms of time, it has been confined 1970s as a decade of this approach. Oller (1979) and some other came up with 'Unitary Competence Hypothesis' of language testing. This hypothesis asserts that language ability can not be divided into discrete items. The sum of discrete elements never equals to the language as a whole. So this is a wholistic approach to language testing. Oller (1979) (quoted in Khaniya 2005: 15-16) argues that language elements interact with each other for meaning, and if language is broken into pieces as in discrete point testing, crucial properties of a language are lost. Therefore, testing language elements is different from testing language itself. Close tests and dictation were exercised during the period.

When close tests and dictation were criticized for not being communicative, the functional approach to testing emerged. It was argued that the nature of language knowledge is best captured by detailing the various uses to which the language can be put (Spolsky, 1985: 182). The functional approach to testing is based on the assumption that knowledge of language should be seen in terms of language related functions, not in terms of underlying grammatical structures. This approach places importance on performance rather than on the linguistic ability of the examinee. The advocates of this approach see language as being

composed of a series of functions that the learner should acquire in order for him to be able to perform different speech acts in various contexts and situations. It appears that performance is the demonstration of competence which is only measure of competence. However, this approach could not satisfy the thrust for determining what is to be tested. Experts involved in this field further explored the possibility of determining what it is that we need to test while testing language. Experts argued that this approach to language testing also could not offer a complete account of what is to be tested in language, and consequently another approach, i.e. communicative approach came into existence.

The communicative approach views language as communication and language learning as developing communicative competence which is essential for enabling learners to use language in the multiple functions it serves in the real life. Language test should evaluate not only the learner's knowledge of the elements and skills but also their ability to comprehend and produce utterances that are both situationally and contextually appropriate. The first communicative model was developed by Hymes (1972), which consisted of both linguistic and socio-linguistic elements. Later in early 1990s, language testing models such as those proposed by Morrow (1997), Canale and Swain (1980) and Bachman (1990) advocated that language testing should be concerned with both competence and performance, i.e. with (a) what the learner knows about the form and about how to use it appropriately, and (b) the extent to which s/he actually demonstrates language knowledge in meaningful performance.

1.1.4 Qualities of a good test

Any test is administered for some purposes. In order to serve the purposes for which a test is conducted, it must be of good quality. In other words, the quality of a test is examined in light of the extent to which it serves the purposes for which it is administered. It is strongly argued that a test without being congruent with the purpose for which it is administered, it is not going to be useful. Therefore, while designing a test, usefulness of the test or the efficiency of the test must be considered (Khaniya, 2005:93).

There are different views on what makes a test good. Some experts say that there are three constituents as the qualities of a good test : 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 frame work for designing a good test. Bachman and Palmer (1996, quoted in Khaniya 2005:93) argue that test usefulness involves reliability, construct validity, authenticity, interactiveness, impact and practicality. The presentation of Bachman and Palmer appears to be a bit elaborate.

Though various scholars may mention various points as qualities of a good test, some of the points are common to all. At least validity, reliability, and practicality are the fundamental qualities that a test should possess to prove itself as a good test. They are discussed below.

i) Validity

Validity is one of the important qualities of a good test "The validity of a test is the extent to which it measures what it is supposed to measure and nothing else" Heaton (1988:159). It refers to the degree to

which a test actually measures what it is designed to measure. It means if a test which is designed to measure pronunciation actually measures the same and nothing else, it is said to be a valid test of pronunciation. Regarding this concept, Davies et al. (1999) assert "A measure is valid if it does what it is intended to do . . .". The validity of a test is measured on the basis of how far the information it provides is accurate, concrete, and representative in light of the purpose for which it is administered (Khaniya, 2005:95).

There are different types of validity. They are:

- a. Content validity
- b. Criterion-related validity
 - Concurrent validity
 - Predictive validity
- c. Construct validity
- d. Face validity
- e. Wash back validity

a. Content Validity

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). Content validity is "the extent to which a test measures a representative sample of the subject matter content" Hatch and Farhady (1982:251). Similarly, according to Richards et al. (1999:81), "Content validity is a form of validity which is based on the degree to which a test adequately and sufficiently measures the particular skill or behaviour it sets to measure. For example, a test of pronunciation skills in language would have low content validity if it

tested only some of the skills which are required for accurate pronunciation."

Brown (1976:122-123) defines content validity as "whether the items composing the test do, in fact, constitute a representative sample of the content domain of concern". For him, content validity is "a measure of the adequacy of sampling". Similarly, Anastasi (1982:131) (quoted in Khaniya 2005: 96) 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."

In addition to this, 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.

Content validity is important from a washback point of view. Following the arguments presented above that an exam can not avoid influencing teaching and learning, and if an examination requires the examinee to exhibit the ability envisaged in the course objectives, the

washback effect of the examination can be beneficial. An exam based on communicative tasks will encourage the students to use language by providing learning opportunities as well. It is argued that an exam of this type can be used for educational change.

Content validity is essential to find out whether all the teaching contents are covered in the tests or not. The greater a test's content validity, the more likely it is to be an accurate measure of what it is supposed to measure. A test in which measure areas identified in the specification are underrepresented or not represented at all is likely to have a harmful washback effect. The areas which are not tested are likely to become areas ignored in teaching and learning. So content validity depends on a careful analysis of the language being tested and of particular course objectives. The test should be so constructed as to contain a representative sample of the course, the relationship between the test items and the course objectives always being apparent (Heaton, 1998:160).

Thus, it is necessary to analyze two aspects, viz. content relevance and content coverage for the investigation of the content validity of a test.

i. Content Relevance

The content validity of a test is examined in relation to its relevance to the given course of study. What it means is whether or not the test tasks included in a test are relevant to the language activities that are expected to be exercised under the given course (Khaniya, 2005:96). For example, if the course of study emphasizes communicative language ability, the test tasks must require the testees to demonstrate such abilities in order to achieve high content validity. The investigation of contents relevance requires the specification of the behaviour domain in question

and the attendant specification of the task or domain. 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, 1961). The amount or the way the subject matters are closely related with the course is content relevance. The more test items are constructed, the more content validity the test paper will have.

ii. Content coverage

It is the extent to which the tasks required in the test adequately represent the behaviour domain in question. In a two or three hour test, one can not use all the contents from the syllabus or course. Therefore, the selection of tasks to be included in the test is indispensable. The basic question is, therefore, whether the test items that compose an exam constitute an appropriate representative sample of behaviour domain under consideration (Brown, 1976:124). Similarly, Hughes (1989) views that content validity depends on 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.

b. Criterion -Related Validity

Criterion related validity is a statistical concept (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. It is important that a test must also be empirically validated 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. The external criterion, however hard to find and however difficult to operationally quantify, remains the best evidence of a test's

validity. All other evidence, including reliability and the internal validity is eventually circular (Davies, 1983:141).

A measure of criterion-related validity of an examination is the correlation between the performance in the exam and the performance in the criterion. It 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 predicting 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.

Concurrent Validity

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 the two tests is said to be the concurrent validity of the new test. 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; or
- 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 with the extent to which the test can predict the future performance of the testees. 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. This type of validity is obtained as a result of comparing the results of the test with results of some criterion measure such as :

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

C) Construct Validity

In language testing, construct validity is defined as the extent to which a test represents an underlying theory of language learning. In other words, a test is said to have construct validity if it is capable of measuring certain specific characteristics 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 inadequate 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 (Heaton, 1988:161).

d) Face Validity

Face Validity is defined as "What it appears superficially to measure" (Anastasi 1982:136). If a test item looks right to other testers, teachers, moderators, and testees, it can be described as having at least face validity (Heaton, 1988:159). Unless the learners genuinely accept a test as a real test, their performance cannot be genuine. Though face validity is regarded as fake or pseudo validity it appears to be important. In this sense, it is important that the test must genuinely look like a test for colleagues and learners.

e) Washback validity

A test's washback effect will be negative if it fails to reflect the learning principles and/ or course objectives to which it supposedly relates; and it will be positive if the effects are beneficial; and encourage the whole range of desired changes. Pilliner (1973:4) maintains the view that the most important requirement of a good test is that it should be adequately beneficial. Morrow (1986:6) terms this effect of a test 'Washback validity' (i.e. positive influence on teaching), and considers this the most important criterion for a good test. So washback effect is inherent in an exam.

ii) Reliability

Reliability is a necessary characteristic of any good test: for it to be valid at all, a test must first be reliable as a measuring instrument. It refers to the consistency of scores or performance of the same or similar test administered within a reasonable time. If the test is administered to the same candidates on different occasions (with no language practice work taking place between these occasions), then to the extent that it

produces differing results, it is not reliable. It is a statistical concept. It is reported in terms of correlation coefficient. Reliability is concerned with examining consistency in the performance of the examinee. In short, it is defined as consistency of measurement.

The degree of consistency of measurement is determined by carrying out some statistical analysis. For that purpose, two sets of scores are obtained from the performance of the same sample of examinees. A correlation of the two sets of scores is said to be the correlation coefficient of the examination, which is interpreted as the coefficient reliability of the exam. Other things being equal, the higher the reliability, the better the exam.

iii) Practicality

Practicality is another quality of a test, though non-technical, absence of this quality in a test leads the test to be of no use, no matter how valid and /or reliable the test is. Heaton's (1988:167) explanation of practicality is that the exam "must be fairly straight forward to administer,"

In order to achieve the practicality of the exam, the designer must keep a close look at the situation which the exam is supposed to fit into. Otherwise, the current literature based on sophisticated situations may lead the designer to be highly ambitious, and to forget the practical problems which are likely to occur at the time of implementation.

It is necessary to pay attention to the following issues in order to develop a test with a reasonable degree of practicality: human resource, material resource, and time. Here, human resource refers to test writers, markers, test administrators, and clerical support. Material resource

refers to space, equipment, and materials. Time refers to the time available for the development, implementation, time given for students to perform the given tasks, the time for scoring and analyzing.

1.1.5 Course Structure of the Master's Degree in English Language Education

There are altogether eleven (11) papers, and out of them five (5) papers carry 100 marks each and six (6) papers carry 50 marks each. These eleven (11) papers are divided into two groups. There are ten (10) papers in group 'A', and Group 'B' has four elective papers out of which 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.No.	Course title no.	Subject	Marks
1	Eng. Ed. 511	Phonetics and Phonology	100
2	Eng. Ed. 512	Grammar : Theory & Practice	100
3	Eng. Ed. 513	Psycholinguistics & 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	English Language Teaching Methods & Practices	100
10	Eng. Ed. 599	ELT Practicum	50

Table No. 2 : Electives

S.No.	Course title no.	Subject	Marks
1	Eng. Ed. 571	English Literature & its pedagogy	50
2	Eng. Ed. 572	Advanced Reading & Writing	50
3	Eng. Ed. 573	Translation: Theory & Practice	50
4	Eng. Ed. 575	Second Language Acquisition	50

Source : CDC, TU.

From the above data and description we know that at M.Ed. first year there are three specialization papers: English Ed. 511(Phonetics and Phonology), Eng. Ed. 512 (Grammar: Theory and Practice) and Eng. Ed. 513 (Psycholinguistics and Sociolinguistics). In second year, there are seven 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 English Ed. 574 Discourse Analysis), Eng. Ed 590 (English Language Teaching Methods and Practices), and Eng. Ed 599 (ELT Practicum). From the elective group, one paper is adopted out of four papers.

1.1.6 The Syllabus of Semantics and Pragmatics at M. Ed Second Year

This is a course on Semantics and Pragmatics. It consists of three units. The first unit is an introduction on some basic concepts. The second and third concentrate on Semantics and Pragmatics respectively.

Course objectives

On completion of this course the trainees will–

- be acquainted with the basic concepts in Semantics and Pragmatics.
- distinguish the relations between Semiotics and Linguistics, Semantics and Pragmatics.
- comprehend the different kinds of meaning and different types of sense relations in language.
- get an insight into field theory of Semantics.
- be acquainted with the ethnography of speaking and with the types of speech acts.

In addition they will be able to

- carry out componential analysis
- carry out conversational analysis.

The contents of this paper can be divided into three units as follows:

Unit I: Introduction

Unit II: Semantics

Unit III: Pragmatics.

(The more details of this course contents is given in Appendix I).

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

Table No. 3 : Unitwise Weighting and Time Allotment

Unit	Marks	Time
I	10	20 hrs
II	15	25 hrs
III	25	30 hrs

Source: CDC, T.U.

In the case of question setting, only skeleton framework of syllabus does not work itself. It depends on the question setter to what extent he tries to follow the syllabus dictating phenomena while setting a test. So content validity depends upon the extent to which the test represents its course and balances the test items within it.

1.2 Literature Review

Although there are some research works in the field of language testing none of the researches are carried out to investigate the content validity of Semantics and Pragmatics paper in master's degree in English Language Education. This will be the first work to find out the content validity of the exam of the above mentioned course. Some of the researches carried out in the field of language testing are reviewed below.

Khaniya(1990) conducted a research on "Examination as Instruments for Educational Changes: Investigating the Washback Effect of Nepalese English Exams" and comes to the conclusion that 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 does not encourage students and teachers to focus on language skills entailed in the objectives. Finally he concluded that:

- I) Washback is an inherent quality of an exam.
- II) Ingredients of an exam determine whether the washback is negative or positive, and
- III) Teaching for final exam is inevitable.

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 examination was + 0.1. On the other hand, in terms representativeness, objectives and itemwise analysis, the SLC English examination had good content validity. But in terms of weighting, the same test papers had low content validity.

Neupane (2004) carried out a research on "Washback Effect of Examinations: A Case of Communicative English." She found that the most favourable technique for the preparation as well as theoretical examination was memorization of exponents. She also found that examination had failed to follow the genuine spirit of the course objective .

Aryal (2005) carried out a research work on "Quality of English Exam: A Case of Content Validity of Grade Twelve Compulsory English." He analyzed the question paper from different angles (by rubrics, length, difficulty level, and content coverage). The study concluded that the rubric of all the questions was simple except few items; the length of the question matches to the allotment of time for the examinees; the questions of English are of moderate difficulty level; the

questions related to 'Heritage of words' lacked content validity. They didn't cover all genres equally. However, the content validity of 'Meanings Into Words' was nearer to the coverage of units and teaching items in comparison with "Heritage of words;

Bhattarai (2005) carried out a research work on "The Content Validity of Compulsory English Textbook for Grade Eight". She concluded that the contents were applicable and the book had content validity in terms of content coverage. Skills and functions were less valid but language structures had high content validity.

Neupane (2005) carried out research on "The Content Validity of English Textbook for Grade Seven" In his research he found that some of the contents were valid and some were less valid. He conducted the research on the basis of content coverage, selection, gradation, and language skills, but he did not mention the language functions and languages structures.

Timsina (2006) studied on" Testing the Test: Investigating the Content Validity of Language Testing Test at M.Ed. Level." He analysed the five years' question papers (2058-2062) of language testing in terms of content coverage and content weighting. In his study, he found that the language testing tests had low content validity in terms of both content coverage and content weighting.

1.3 Objectives of the study

The objectives of the study were as follows:

- i. to examine the content validity of Semantics and Pragmatics test paper of M.Ed. second year in terms of

- a) content coverage
 - b) content weighting
- ii. to suggest some pedagogical implications for the betterment of the test for future use.

1.4 Significance of the Study

As a distinct research work from the rest in the Department of English Language Education, this study will provide information on whether the administered Semantics and Pragmatics question papers had content validity or not. For the test designers, it will provide some guidelines to make a test valid. Thus, this study will be beneficial to all those who are concerned with language learning, teaching and testing, more particularly to the teachers and test designers who are involved in the teaching of Semantics and Pragmatics at M.Ed. second year. Moreover, hopefully this research will be significant to policy makers, curriculum designers, students and those who are directly or indirectly involved in the field of teaching & testing.

CHAPTER TWO

METHODOLOGY

The researcher has followed the following methodology during the study in order to achieve the objectives specified.

2.1 Sources of Data

The researcher has used only secondary sources of data:

2.1.1 Secondary Sources of Data

Question papers of the course entitled 'Semantics and Pragmatics' from the year 2058 to 2063 B. S. at M. Ed. Second year were the secondary sources of data. Apart from this. New English Syllabus of T.U., different books on testing, Semantics and Pragmatics, reports, journals, articles related to the topic have been taken as the secondary sources of data, for example, Khaniya (2000, 2005), Batala (2004), Timsina (2006), etc.

2.2 Process of Data Collection

The researcher collected six years' question papers, from 2058 to 2063 B. S., of the subject entitled ' Semantics and Pragmatics' at M .Ed. second year, major English, which were administered at T. U. containing both subjective and objective questions to judge whether the test items had content validity or not in terms of coverage and weighting.

2.3 Limitations of the Study

The study had the following limitations:

- i. The study was limited to the content validity of the test on 'Semantics and Pragmatics' of M. Ed. (Second year).
- ii. The study was limited to the question papers from 2058 to 2063 B. S. asked by Tribhuvan University.
- iii. The study was limited to the testing of content validity regarding the written test of Semantics and Pragmatics since there is no provision of oral test in M. Ed. Second year test on the subject in question.

CHAPTER THREE

ANALYSIS AND INTERPRETATION

3.1 Analysis of Content Validity

This chapter deals with the analysis and interpretation of the raw data used in the study. The main concern of the present research work was to find out the content validity of 'Semantics and Pragmatics' question papers asked at M. Ed. Second year. For this purpose, this chapter has been divided into two parts. The first part deals with the analysis of the content validity of 'Semantics and Pragmatics' question papers in terms of coverage and the second part deals with the content validity of the same subject in terms of weighting.

Here, in the first part, the question papers of Semantics and Pragmatics which were administered in Tribhuvan University Examinations during a period of six years, from 2058 to 2063 B. S. were analyzed in terms of coverage. The researcher tried to find out whether these test papers had content validity or not.

There are altogether three units in the subject entitled 'Semantics and Pragmatics' at M. Ed. (Second year). It is a theoretical subject, 50 is its Full marks; and 20 is its Pass marks. Course contents and its weighting are clearly specified in the syllabus which is given in Appendix I. The researcher has mainly analyzed subjective as well as objective questions of those academic years. The question papers of those six examinations are given in Appendix II, III, IV, V, VI and VII.

Though there are different views on content validity, Hughes (1995) and Bachman (1998) have emphasized two things for content validity: content representation and content relevance. But here, the

researcher has used mainly Heaton's(1988) views on content validity which argue that there are two components through which we can judge the test papers to find out whether they have content validity or not. They are: 'representative sample of the course' and 'weighting of the course'.

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

3.1.1 Representative Sample/Content Coverage: Comparison between course contents and test contents.

For the purpose of investigating content validity of 'Semantics and Pragmatics' test at M. Ed. Second year during six years (2058 to 2063 B. S.), the researcher compared the test contents in relation to the course contents. In other words, he has examined whether the test contained a representative sample of the whole course. If the representative sample of the content is above 60 percent, it is assumed that the test paper is nearer to the high content validity. If it is below 50 percent, it is supposed to have low content validity. So the higher the representative sample of the content, the higher the content validity the test will have. So, to find out the content validity of a test in terms of course representativeness, the researcher examined and analyzed the 'Semantics and Pragmatics' question papers from the year 2058 to 2063 B. S. including all subjective and objective questions. The researcher used only descriptive and tabulation methods to analyze the data.

a. Examining Course Representativeness in Unit One.

Table No. 4 : Representation of test contents in terms of course content in unit one

S.N.	Course contents Course items	Test Contents Test items represented					
		2058	2059	2060	2061	2062	2063
Unit							
I	Introduction						
1.1	Semiotics and Linguistics	O5	O7	O1		O1	
1.2	Form and Substance			S1			S5
1.3	Form and Meaning	S1					
1.4	Form and Function		S1				
1.5	Semantics and Pragmatics				O2,O7	S1	O1
	Total no. of questions asked from Unit I.	1-O 1-S	1-O 1-S	1-O 1-S	2-O	1-O 1-S	1-O 1-S

Note: O1= Objective questions: question no. 1

S1= Subjective short questions: question no. 1

L6= Subjective long questions: question no. 6

The above table indicates that in unit one, there are 5 language items from 1 to 1.5. If we see diachronically among these 5 language items, 1.1 (Semiotics and Linguistics) is the most representative language item, repeated for four years from 2059 to 2063 B. S. Language item 1.5 (Semantics and Pragmatics) has been represented in three years' question

papers out of six. Language item 1.2 (Form and Substance) has been represented in two years' question papers. Language item 1.3 (Form and Meaning) and 1.4 (Form and Function) have been represented only in one year's question paper out of the six.

If we see synchronically, two questions (1 objective and 1 subjective) were asked in 2058. The objective question was asked from 1.1 (Semiotics and Linguistics); and the subjective question was asked from 1.3 (Form and Meaning). In 2059, two questions (1 objective and 1 subjective) were asked. The objective question was asked from 1.1 (Semiotics and Linguistics); and the subjective question was asked from 1.4 (Form and Functions). In 2060, two questions (1 objective and 1 subjective) were asked. The objective question was asked from 1.1 (Semiotics and Linguistics) and the objective question was asked from 1.2 (Form and Substance). In 2061, two questions (2 objectives) were asked, and both of them were from 1.5 (Semantics and Pragmatics). In 2062, two questions (1 objective and 1 subjective) were asked. The objective question was asked from 1.1 (Semiotics and Linguistics), and the subjective question was asked from 1.5 (Semantics and Pragmatics). In 2063, two questions (1 objective and 1 subjective) were asked. The objective question was asked from 1.5 (Semantics and Pragmatics), and the subjective question was asked from 1.2 (Form and Substance).

b. Examining Course Representativeness in Unit Two

Table No. 5 : Representation of test contents in terms of course content in unit two

S.No.	Course Contents	Test Contents					
	Course Items	Test items Represented					
		2058	2059	2060	2061	2062	2063
Unit II Semantics							
2.1	Types of Meaning						
2.1.1	Linguistic meaning and speaker meaning						
2.1.2	Intended meaning and interpreted meaning						
2.1.3	Sentence meaning and utterance meaning			S2			
2.1.4	Word meaning and sentence meaning		O5				
2.1.5	Lexical meaning & Grammatical meaning	O6	S2				
2.1.6	Sense and denotation	S2			S2		O2
2.1.7	Denotative meaning and connotative meaning						
2.1.8	Literal and figurative meaning						
2.2	Sense relations						
2.2.1	Synonymy		L6	O2		L6	
	- partial and absolute synonymy - stylistic synonymy - dialectal synonymy - register synonymy						

	- intralingual and interlingual synonymy						
2.2.2	Antonymy	L6					O5
	- gradable antonymy				O5		
	- complementarity						
	- converseness		O1			O7	
	- incompatibility				O1		
2.2.3	Hyponymy	O1					
	- hyperonymy and hyponym - co-hyponyms						
2.2.4	Homonymy			S3			
	- homophony and homography						
	- lexical and grammatical homonymy						
	- homonymy and polysemy				S3		S2
2.3	Componential Analysis					S3, O2, O5	O7
2.3.1	Semantic fields			S4			
2.3.2	Hierarchical structure in the vocabulary						
2.3.3	Semantic features		O2	O5			
	Total No. of questions asked from unit II	2-O. 1-S. 1-L.	3-O. 1-S. 1-L.	2-O. 3-S.	2-O. 2-S.	3-O. 1-S. 1-S.	3-O. 1-S.

The above table shows that in unit two, there are 18 language items from 2 to 2.3.3. If we see diachronically among these 18 language items, there is not any most representative language item repeating each and every year from 2058 to 2063 B. S. Language item 2.2.2

(Antonymy) has been represented in five years' question papers out of six. Language items 2.1.6 (Sense and denotation), 2.2.1 (Synonymy), and 2.2.4 (Homonymy) have been represented in three years' question papers out of six years. Language items 2.1.5 (Lexical meaning and grammatical meaning), 2.3 (Componential Analysis) and 2.3.3 (Semantic features) have been represented in two years' question papers. Language items 2.1.3 (Sentence meaning and utterance meaning), 2.1.4 (Word meaning and sentence meaning), 2.2.3 (Hyponymy) and 2.3.1 (Semantic fields) have been represented in only one year's question paper out of six years. Beside them, the language items which are mentioned in the table have not been represented in any year.

If we see synchronically, 4 questions (2 objectives, 1 subjective and 1 subjective long question) were asked in 2058. The objective questions were asked from 2.1.5 (Lexical meaning and grammatical meaning) and 2.2.3 (Hyponymy). The subjective short question was asked from 2.1.6 (Sense and denotation). Regarding subjective long questions, it was asked from 2.2.2 (Antonymy). In 2059, 5 questions (3 Objectives, 1 subjective short and 1 subjective long) were asked. Regarding objective questions, they were asked from 2.1.4 (Word meaning and sentence meaning), 2.2.2 (Antonymy) and 2.3.3 (Semantic features). The subjective short question was asked from 2.1.5 (Lexical meaning and grammatical meaning), and the subjective long question was asked from 2.2.1(Synonymy). In 2060, 5 questions (2 objective and 3 subjective short) were asked. The objective questions were asked from 2.2.1 (Synonymy) and 2.3.3 (Semantic features). Regarding the subjective short questions, they were asked from 2.1.3 (Sentence meaning and utterance meaning), 2.2.4 (Homonymy) and 2.3.1 (Semantic fields). In 2061, 4 questions (2 objective and 2

subjective short) were asked. Both of the objective questions were asked from 2.2.2 (Antonymy). The subjective short questions were asked from 2.1.6 (Sense and denotation) and 2.2.4 (Homonymy). In 2062, 5 questions (3 objectives, 1 subjective short and 1 subjective long) were asked. One objective question was asked from 2.2.2 (Antonymy) and 2 objective questions were asked from 2.3 (Componential Analysis). Regarding the subjective long question, it was asked from 2.2.1 (Synonymy). In 2063, 4 questions (3 objective and 1 subjective short) were asked. In the case of objective question, they were asked from 2.1.6 (Sense and denotation), 2.2.2 (Antonymy) and 2.3 (Componential Analysis. The subjective short question was asked from 2.2.4 (Homonymy).

c. Examining Course Representativeness in Unit Three

Table No. 6: Representation of test contents in terms of course contents in unit three.

S.N.	Course Contents	Test Contents					
		Test items Represented					
		2058	2059	2060	2061	2062	2063
Unit III Pragmatics							
3.1	The scope of Pragmatics						S1
3.1.1	The origin of the term Pragmatics						
3.1.2	Defining Pragmatics						O6
3.1.3	Current interest in Pragmatics						
3.2	Deixis			L6			

3.2.1	Introduction						
	- person deixis						
	- place deixis						
	- time deixis						
	- discourse deixis	O7	O8		O4		
	- social deixis		O3			S2	
3.3	Conversational Implicature			O7			
3.3.1	Introduction						
3.3.2	The co-operative principle	S3	S3	O3	L6		S3
	- the maxim of quality				O6		
	- the maxim of quantity						
	- the maxim of relevance						
	- the maxim of manner						
3.3. 3	The politeness principle	O2			S5	O4	
	- maxim of tact						
	- maxim of generosity						
	- maxim of approbation				O8		
	- maxim of modesty					O6	O3
3.4	Speech Acts			S5			
3.4. 1	Austin's speech act (theory)	O8				S5	S4
	- locutionary act					O8	
	- illocutionary act						
	- perlocutionary act						O8
3.4.2	Searle's speech acts	O3	S4				
	- representatives						
	- directives					O3	O4

	- commissives						
	- expressives						
	- declarations						
3.4.3	Enthnography of speaking	S4			S1		
	- Setting: physical, psychological						
	- participants: addresser = speaker, addressee = hearer						
	- Ends: goals, outcome						
	- Acts: forms, contents						
	- Key: tone				O3		
	- Instruments: channels						
	- Norms						
	- Genre						
3.4.4	Conversational Analysis	S5		O4, O6			L6
	- Opening						
	- Turns and turn-taking: self selection, other selection/nomination	O4	S5			S4	
	- Adjacency pairs: first part pairs, second part pairs, chaining						
	- Timing: overlap, pause						
	- move						
	-Topic: Introduction/nomination, continuation, change, conflict						
	-side-sequence:		O6	O8			

	cohesion, coherence						
	- Pre-closing, closing						
	Total No. of questions asked from unit II	5-O. 3-S.	4-O. 3-S.	5-O. 1-S. 1-L.	4-O. 3-S. 1-L.	4-O. 3-S.	4-O. 3-S. 1-L.

The above table shows that in unit three, there are 15 language items from 3.1 to 3.4.4. If we see diachronically among these 15 language items, language item 3.4.4 (Conversational Analysis) is the most representative one, repeating each and every year from 2059 to 2063. Language item 3.3.2 (The co-operative principle) has been represented in five years' question papers out of six. Language item 3.4.1. (Austin's speech act) has been represented in three years' question papers out of six. Language item 3.4.3 (Ethnography of speaking) has been represented in two years' question papers out of six. Language items 3.1 (The scope of Pragmatics), 3.1.2 (Defining Pragmatics), 3.2. (Deixis), 3.3. (Conversational implicature), and 3.4 (Speech acts) have been represented in only one year's question paper out of six. Besides them, the language items which are mentioned in the table above have not been represented in any year.

If we see synchronically, 8 questions (5 objective and 3 subjective short) were asked from this unit in 2058. In the case of objective questions, they were asked from 3.2.1. (Deixis: discourse deixis), 3.3.3 (The politeness principle), 3.4.1. (Austin's speech act theory), 3.4.2 (Searle's speech Acts), and 3.4.4. (Conversational implicature: Adjacency pairs). The subjective short questions were asked from 3.3.2 (The co-operative principle), 3.4.3.(Ethnography of speaking) and 3.4.4 (Conversational Analysis). In 2059, 7 questions (objective and 3

subjective short) were asked. Out of 4 objective questions, two were asked from 3.2.1 (Introduction Deixis), and two of them were asked from 3.4.4 (Conversational implicature) related to the subheadings- Turns and turn taking, and side sequence: cohesion, coherence. The subjective questions were asked from 3.3.2. (The co-operative principle), 3.4.2.(Searle's speech acts), and 3.4.4 (Conversational Analysis). In 2060, 7 questions (5 objective, 1 subjective short and 1 subjective long) were asked. Regarding objective questions, 3 were asked from 3.4.4 (Conversation Analysis); and 1 question was asked from 3.3 (Conversational Implicature) and 3.3.2 (The co-operative principle) each. The subjective short question was asked from 3.4 (Speech Acts). The subjective long question was asked from 3.2. (Deixis). In 2061, 8 questions (4 objectives, 3 subjective short and 1 subjective long) were

asked. In the case of objective questions, they were asked from 3.2.1 (Deixis: discourse deixis), 3.3.2 (The co-operative principle: the maxim of quality), 3.4.3 (Ethnography of speaking: key) and 3.3.3 (The politeness principle: maxim of approbation). The subjective short questions were asked from 3.3.3 (The politeness principle), 3.4.3. (Ethnography of speaking), and 3.4.4 (Conversational Analysis: Adjacency pairs). The subjective long question was asked from 3.3.2 (The co-operative principle). In 2062, 7 questions (4 objective and 3 subjective short) were asked. Regarding objective questions, two were asked from 3.3.3 (The politeness principle), one from 3.4.1 (Austin's speech act theory), and one from 3.4.3. (Searle's speech acts). In the case of subjective short questions, they were asked from 3.2.1 (Deixis: Social

deixis), 3.4.1 (Austin's speech act theory), and 3.4.4.(Conversational Analysis). In 2063, 8 questions (4 objectives, 3 subjective short and 1 subjective long) were asked. Regarding objective questions, they were asked from 3.1.2 (Defining Pragmatics) 3.3.3 (The politeness principle), 3.4.1 (Austin's speech act theory), and 3.4.2 (Searle's speech acts). The subjective short questions were asked from 3.1 (The scope of Pragmatics), 3.3.2 (The co-operative principle), and 3.4.1 (Austin's speech act theory). Long question was asked from 3.4.4 (Conversational Analysis).

d. Examining Content Validity of the Test papers on the whole in terms of coverage

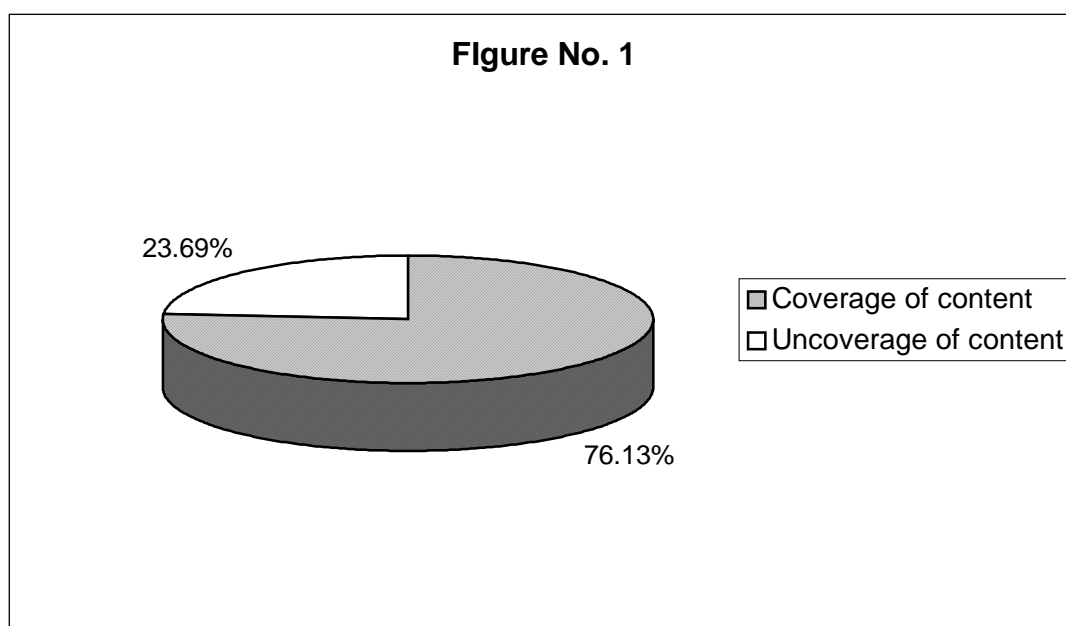
Table No.: 7 : Examining content validity of the test papers on the whole in terms of coverage

S.No.	Units	Course contents language items	Test contents language items	Test coverage (in percentage)
1.	1	5	5	100
2	2	18	12	66.66
3	3	15	12	80
	Total	38	29	76.31

The above table indicates that on the whole, there are 38 language items in the Semantics and Pragmatics course at M.Ed. second year. Out of 38 course contents, 29 language items were represented in test contents. 9 language items were neglected while constructing the test items. It means the coverage of contents in whole units is 76.31 percent.

23.69 percent contents were not covered in the question papers from 2058 to 2063.

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



As the above description mentions that the content validity of the test papers as a whole is satisfactory because it has covered more than 60 percent course contents.

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 Semantics and Pragmatics test papers. For the purpose of examining content validity of the Semantics and Pragmatics test, the researcher has examined the Semantics and Pragmatics question papers from the years 2058 to 2063 to find out

whether or not the marks weighting in the course contents according syllabus is proportional with the weighting of the test contents.

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 weightage	2058	2059	2060	2061	2062	2063
1	1	10	1+6	1+6	1+6	2	1+6	1+6
2	2	15	2+6+12	3+6+12	2+18	2+12	3+6+12	3+6
3	3	25	5+18	4+18	5+6+12	4+8+12	4+18	4+18+12
	Total marks	50	8+30+12	8+30+12	8+30+12	8+30+12	8+30+12	8+30+12
	Total	50	50	50	50	50	50	50

Note: In 2058 (from unit 1) 1+6=1 objective question which carries 1 mark and 1 subjective short question which carries 6 marks. Here, 12 means a subjective long question which carries 12 marks. In other words, the first number, in the sequence of marks, is related to objective question, second to subjective short question and third to subjective long question.

The above table indicates that out of 50 marks of the whole 'Semantics and Pragmatics' course at M.Ed. level, unit 1 carries 10 marks according to syllabus. While comparing the weighting of the course

contents with that of weighting of the test contents, the researcher has found that though unit 1 is supposed to carry 10 marks, it is not found to have followed the marking schedule in test papers because in 2058, there was asked only 7 marks (1 mark for 1 objective question and 6 marks for 1 subjective short question). Similarly 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 7 marks (1 mark for 1 objective question and 6 marks for 1 subjective short question). In 2061, questions carrying only 2 marks (2 marks for 2 objective questions) were asked. In 2062, questions carrying 7 marks (1 mark for 1 objective question and 6 marks for subjective short question) were asked. In 2063, questions carrying 7 marks (1 mark for 1 objective question and 6 marks for 1 subjective short question) were asked.

In unit 2, this syllabus makes a provision of weighting 15 marks but the researcher has found that the weighting of the course contents has not been followed in question papers. In 2058, questions carrying 20 marks (2 marks for 2 objective questions, 6 marks for 1 subjective question and 12 marks for 1 subjective long question) were asked. In 2059, questions carrying 21 marks (3 marks for 3 objective questions, 6 marks for 1 subjective question and 12 marks for 1 long question) were asked. In 2060, there was asked 20 marks (2 marks for 2 objective questions and 18 marks for 3 subjective short questions). In 2061, questions carrying 14 marks (2 marks for 2 objective questions and 12 marks for 2 subjective short questions) were asked. In 2062, there was asked 21 marks (3 marks for 3 objective questions, 6 marks for 1 subjective short question and 12 marks for 1 subjective long question). In 2063, there were the questions carrying 9 marks (3 marks for 3 objective questions and 6 marks for 1 subjective short question).

According to the syllabus, unit 3 is supposed to carry 25 marks, but the question papers have carried different weightage. In 2058, there was asked 23 marks (5 marks for 5 objective questions and 18 marks for 3 subjective short question). In 2060, questions carrying 23 marks (5 marks for 5 objective questions, 6 marks for 1 subjective short question and 12 marks for 1 long question). In 2061, there was asked 34 marks (4 marks for 4 objective questions, 18 marks for 3 subjective short questions and 12 marks for 1 subjective long question). In 2062, there were the questions carrying 22 marks (4 marks for 4 objective questions and 18 marks for 3 subjective short questions). Finally in 2063, there was asked 34 marks (4 marks for 4 objective questions, 18 marks for 3 subjective short questions and 12 marks for 1 subjective long question).

As a whole during the 6 years' period (2058 to 2063) in 'Semantics and Pragmatics' question papers at M.Ed. second year, the researcher has found that unit 1 should have carried 10 marks but the question setters have not followed the weightage schedule according to syllabus. In the six years' question papers the distribution of marks is under weightage. In the case of unit 2, it should have carried 15 marks, but the question papers in the six years' period have no matching with the weightage schedule. Regarding unit 2, the question papers in 2058, 2059, 2060 and 2062 have over weightage whereas those in 2061 and 2063 have under weightage. In the case of unit 3, it should have carried 25 marks according to the schedule in syllabus, but question papers in some years have overweightage and some have underweightage. Question papers in 2058, 2059, 2060 and 2062 have underweightage whereas those in 2061 and 2063 have over weightage. It shows the negligence in the part of question setters which may lead to the lack of content validity and negative washback as well. Though in each unit there is a fixed course content

marking (weighting) schedule, but it is not found to be followed in setting the question papers. It creates negative washback effect because the unit which is given least weightage in question papers in examination, is given very less priority by the students. They do not bother reading such unit because they know the neglected area in course. Consequently, it leads to the lack of fulfilment of course objective as well.

So far the researcher has observed the course content weighting and test weighting of language items in the course 'Semantics and Pragmatics' M.Ed. 2nd year, there is no satisfactory finding. Some course contents are found to have over weighting and some under weighting. It seems that the question setters followed no specific norm for setting questions. Thus, from the above data and description, the researcher concluded that Semantics and Pragmatics question papers lack content validity in terms of the weighting of the course contents.

3.3 Examining The Content Validity in Terms of Coverage and Weighting

It has been found that the Semantics and Pragmatics question papers at M.Ed. second year have high content validity in terms of coverage or representation of the course contents. It is so because out of 38 language items, the question papers represented 29 language items, i.e. 76.13 percent. But the same question papers have low content validity in terms of weighting of the course contents because test items did not follow the exact scheduled weighting mark as it was supposed to. This shows that the question paper having high content validity in terms of coverage does not necessarily have high content validity in terms of weighting. In other words, a question paper may have high content validity in terms of coverage but low content validity in terms of weighting and vice-versa.

CHAPTER FOUR

FINDINGS AND RECOMMENDATIONS

This chapter deals with the major findings of the study. After the analysis of the question papers from different angles, the researcher has come with the following findings.

4.1 Findings

The major findings of this research are as follows:

A. According to the coverage/representativeness principle, the Semantics and Pragmatics tests have high content validity. It is so because out of 38 language items in total, the test items have represented 29 language items, i.e. 76.31 percent during the 6 years (2058 to 2063).

On the other hand, according to the weighting principle, the Semantics and Pragmatics tests have low content validity. It is so because the question papers didnot follow any norm or tendency in the distribution of marks as it was given in the weighting schedule according to the syllabus. For example in 2061, from unit 1, the test item has carried only 2 marks in the test papers whereas its weightage, according to the syllabus, is 10 marks.

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

1. In terms of coverage/representativeness

Comparison between course contents and test contents

a. In unit one, out of 5 language items of course contents, test items represented 5 language items during 6 years (2058 to 2063). No item was untouched. It means the coverage of the

course contents is 100 percent. In conclusion Semantics and Pragmatics question papers have a very high content validity in unit one since 100 percent course contents have been represented in the question papers during the 6 years (2058 to 2063).

- b. In unit two, out of 18 language items of course contents, test items represented 12 language items during 6 years (2058 to 2063). Six (6) language items were untouched. It means that the coverage of course contents is 66.66 percent. In conclusion, Semantics and Pragmatics tests have high content validity in unit two because question papers contents have represented more than 60 percent course contents during the 6 years (2058 to 2063).
- c. In unit three, out of 15 language items of course contents, test items represented 12 language items. Three (3) language items were untouched. It means that the coverage of course contents is 80 percent. In conclusion, Semantics and Pragmatics question papers have high content validity in unit three because test contents have represented more than 60 percent course contents during the 6 years (2058 to 2063).

2. In terms of weighting

Comparison between course content weighting and test content weighting

- a. In unit one, out of 10 marks weighting according to syllabus, the question papers have carried different weightings in different years during the 6 years (2058 to 2063).

In 2058, unit one carried 7 marks, 1 for objective and 1 subjective short question. In 2059, unit one only carried 7 marks, 1 for objective and 1 subjective short question. In 2060, it only carried 7 marks, 1 for objective and 1 subjective short question. In 2061, it only carried 2 marks, 2 for objective questions. In 2062, it carried 7 marks, 1 for objective and 1 subjective short question. In 2063, it carried 7 marks, 1 for objective and 1 subjective short question.

Thus, as a whole in unit one, it has been found that in all the years the mark weighting in the question papers was not proportional to the weighting of its course contents because of its under weighting in the test papers. This shows that the 'Semantics and Pragmatics' question papers lack content validity in terms of weighting regarding unit one.

- b. In unit two, out of 15 marks weighting of course contents, according to the syllabus, the weighting of the test contents was as follows:

In 2058, unit two carried 20 marks, 2 for objective, 1 subjective short and 1 subjective long questions. In 2059, it carried 21 marks, 3 for objective, 1 subjective short and 1 subjective long questions. In 2060, it carried 20 marks, 2 for objective and 3 subjective short questions. In 2061, it carried only 14 marks, 2 for objective and 2 subjective short questions. In 2062, it carried 25 marks, 3 for objective questions, 1 subjective short question and 1 subjective long question. In 2063, it carried only 9 marks, 3 for objective and 1 subjective short question.

Thus, as a whole in unit two, it has been found that in 2058, 2059, 2060 and 2062, the marks in the question papers were over weighting. On the other hand, in 2061 and 2063, the marks in the question papers were under weighting. It has been found that in all the years, the mark

weighting was not proportional against the weighting of its course contents. This shows that the Semantics and Pragmatics question papers lack content validity in terms of weighting regarding unit two.

- c. In unit three, out of 25 marks weighting of course contents, the weighting of the test contents was as follows:

In 2058, unit three carried 23 marks, 5 from objective and 3 subjective short questions. In 2059, it carried 22 marks, 4 for objective and 3 subjective short questions. In 2060, it carried 23 marks, 5 for objective, 1 subjective short and 1 subjective long question. In 2061, it carried 34 marks, 4 for objective, 3 subjective short and 1 subjective long question. In 2062, it carried 22 marks, 4 for objective and 3 subjective short questions. In 2063, it carried 34 marks, 4 for objective, 3 subjective short and 1 subjective long question.

Thus, as a whole in unit three, it has been found that in 2058, 2059, 2060 and 2062, the marks in the question papers were underweighting. On the other hand, in 2061 and 2063, the marks in the question papers were overweighting. It means that in all the years the mark weighting was not proportional against the weighting of its course contents. This shows that the Semantics and Pragmatics question papers lack content validity in terms of weighting regarding unit three.

3. According to the coverage or representation of the course contents, the Semantics and Pragmatics question papers have really tested what they have been supposed to test in the testees because the test items represented more than 60 percent course items.

4. According to the weighting of the course contents, Semantics and Pragmatics question papers have not accurately tested what they have been supposed to test in the testees because they did not follow any weighting system of the course contents in the test contents. Weighting of the language items seems to be determined by what is easy to test rather than what is important to test following unitwise weighting system.
5. More variation has been found in asking all types of questions (objective, subjective short and subjective long) in the different years in the same unit as well. For example, in 2061, only 2 objective questions were asked from unit 1; whereas 1 objective and 1 subjective short questions were asked in all the years except in 2061. Similarly, from unit 3, no long question was asked in 2058, 2059 and 2060 whereas in 2061 and 2063, 4 objective, 3 subjective short and 1 subjective short questions were asked.

4.2 Recommendations

On the basis of the findings of the research work some recommendations are made which are listed below.

1. It has been found that Semantics and Pragmatics question papers have represented 80 percent course contents as a whole. Seemingly it is supposed to have good content validity. However, 20 percent course items have remained untouched. Semantics and Pragmatics question papers should try to cover those untouched course contents to have higher content validity.
2. It has been found that Semantics and Pragmatics question papers have not followed the scheduled weighting of the course contents.

Therefore, while setting the question papers, the question setter should strictly obey the weighting schedule to have high content validity.

3. For a test to have high content validity, it should follow not only the representative principle but weighting principle as well. Thus, Semantics and Pragmatics question papers should follow both principles - principles of representativeness and weighting-to have high content validity.
4. Asking different types of questions differently in different years creates harmful effects on the validity of the tests and testees. Semantics and Pragmatics question papers would have high content validity if, any fixed criteria, for example 'specification table', were prepared for asking different types of questions (objective, subjective short and/or subjective long) regarding different units. So Semantics and Pragmatics should have specification table and follow it very carefully to have high content validity.
5. It has been found that some test items seem to be determined by the insights that what is easy to test rather than what is important to test. Thus, Semantics and Pragmatics question papers should test those aspects or qualities which are supposed to be developed on the part of testees according to the course. For this purpose, Semantics and Pragmatics question papers should follow criteria of content validity - representativeness and weighting - very strictly.
6. Instead of repeating the same items in each and every year's examination, the tests should cover different language items in the examination. The repetition of the same language items should be excluded as far as possible so that it would increase the

representation of the course contents and have positive washback effect on the testees.

7. For a test to have high content validity, it should emphasize all the units proportionally according to the syllabus. So Semantics and Pragmatics tests should give proportional emphasis (neither more or less) to each and every unit as it has been scheduled in the syllabus.
8. If a test contains a greater number of test items, there will be the possibility of having higher content validity since the test may have higher coverage. So instead of including few test items having high weighting in the tests, many more test items having low weighting should be constructed to represent the course contents of the syllabus. For example, it can be done by asking short questions carrying 4 marks instead of 6 marks for each, and long questions having 8 marks instead of 12 marks for each.
9. For a test to have high content validity, the question setters should have deep insights regarding course objectives, course contents, qualities of a good test, and weighting of the course contents before developing the question papers. Highly experienced and trained teachers should design the question papers using their insight carefully.

BIBLIOGRAPHY

- Anastasi, A. 1982. *Psychological Testing*. Fifth Edition. New York: Macmillan Publishing Co., Inc.
- Aryal, K.P. 2005. *Quality of English Exam: A Case of Content Validity of Grade Twelve Compulsory English Exam 2001*. An Unpublished M.Ed. Thesis. Kathmandu: T.U.
- Batala, K.B. 2004. *Validity of the SLC English Examination*. An Unpublished M.Ed. Thesis. Kathmandu: T.U.
- Bhattarai, J. 2005. *The Content Validity of Compulsory English Textbook for Grade Eight*. An Unpublished M.Ed. Thesis. Kathmandu: T.U.
- Brown, F.G. 1976. *Principles of Educational and Psychological Testing*. New York: Holt Rinehart and Winston.
- Canale, M. and Swain, M. 1980. *Theoretical Bases of Communicative Approaches to Second Languages Teaching and Testing*. Applied Linguistics.
- Chomsky, N. 1957. *Syntactic Structures*. The Hague: Mouton..
- David, A. 1983. *The Validity of Concurrent Validation*. London: OUP.
- Davies et al. 1999. *Dictionary of Language Testing*. University of Melbourne.
- Hatch, E. and Farhady, H. 1982. *Research Design and Statistics for Applied Linguistics*. Rowley, Mass: NBHP, Inc.
- Heaton, J.B. 1988. *Writing English Language Tests*. Longman: ELBS.
- Hughes, A. 1995. *Testing for Language Teachers*. London: OUP.
- Hymes, D. 1972. *On Communities Competence*. Harmondsworth: Penguin.
- Khaniya, T.R. 1990. *Examinations as Instruments for Educational Change: Investigating the Washback Effects of Nepalese Exams*.

- An Unpublished Ph.D. Thesis. U.K.: University of Edinburgh.
- Khaniya, T.R. 2005. *Examination for Enhanced Learning*. Lalitpur: Millennium Publication (P.) Ltd.
- Kumar, R. 1996. *Research Methodology*. London: Sage Publication.
- Lado, R. 1961. *Language Testing: The Construction and Use of Foreign Language Tests*. London: Longman.
- Leech, G. 1981. *Semantics*. Harmondsworth: Penguin.
- Levinson, S.C. 1983. *Pragmatics*. Cambridge: CUP.
- Lyons, J. 1995. *Linguistic Semantics*. Cambridge: CUP.
- Morrow, K. 1986. *The Evaluation of Tests of Communicative Performance*. London: OUP.
- Neupane, M. 2004. *Washback Effect of Examinations: A Case of Communicative English*. An Unpublished M.Ed. Thesis. Kathmandu: T.U.
- Neupane, S. 2005. *The Content Validity of English Textbook for Grade Seven*. An Unpublished M.Ed. Thesis. Kathmandu: T.U.
- Ogden, C.K. and I.A. Richards 1923 (8th edition). *The Meaning of Meaning*, London: Routledge and Kegan Paul.
- Older, J.W. 1979. *Language Tests at School*. London: Longman.
- Pilliner, A. 1973. *Assessment Principles and Practice with Special Reference to Education in Pakistan*. The British Council.
- Richards et al. 1999. *Longman Dictionary of Language Teaching and Applied Linguistics*. London: Longman.
- Richards, J.C. and T. Rodgers 1986. *Approaches and Methods in Language Teaching*. New York: CUP.
- Saussure, F.de. 1966. *Course in General Linguistics*. New York: McGraw-Hill.

- Spolsky.B. 1978. *Advances in Language Testing Survey-2: Approaches to Language Testing*. Virginia: Centre for Applied Linguistics.
- Syal, P. and D.V. Jindal, 2005. *An Introduction to Linguistics, Language, Grammar and Semantics*. India: Prentice-Hall.
- Timsina, H.L. 2006. *Testing the Test: Investigating the Content Validity of Language Testing Test at M.Ed. Level*. An Unpublished M.Ed. Thesis. Kathmandu: T.U.