CHAPTER-ONE

INTRODUCTION

1. General Background

This study is about content validity of compulsory English grade eight test. The first chapter consists of general background, language, English language, language teaching and testing, qualities of a good language test, validity, types of validity, literature review, objectives of the study, the significance of the study and definitions of key terms.

1.1 Language

Language, a divine gift of God, to mankind helped to distinguish human being from all other creatures in the world. Language is a means of communication through which human beings exchange their ideas feelings, thoughts, desires, emotions, expressions and share information. Language is the most common system of communication. Every normal human, uses language in order to communicate with other people. Human communication system is uniquely and vastly different from the communication system carried by non-human species. That is to say, every normal human being is born with the capacity in his brain to acquire language. Thus, language is said to be 'species specific'. It is limited only to human languages; only human languages enable their users to produce and understand an infinite number of sentences. In such case we can required the definition of language.

Sapir (1971, p. 8) says, "Language is a purely human and non-instinctive method of communicating ideas, emotions and desires

by means of a system of voluntarily produced symbols." It is perhaps the most significant assest of human life by which we can perform every things.

Similarly, Richards et al (1985, p. 153) mention, "Language is the system of human communication by means of a structural arrangement of sounds to form longer units." That is to say, it is made up of smaller units which are related to each other and perform particular functions.

Likewise, Robins (as cited in Yadava,2001, p. 4) clarifies, "Language is a symbol systems based on pure or arbitrary conventions infinitely extendable and modifiable according to the changing needs of the speakers." The language is spoken as the speaker desired. That's why, human language is unique and it does not remain static.

In whatever ways it is defined, it is a system of arbitrary vocal symbols by means of which human beings communicate. Language is not an isolated phenomenon but it is a social phenomenon, which is used in our society to establish the human relationship. It allows people to talk to each other either in spoken form or in written form. For the purpose of communication, we can find the dignity of more languages in the world, but it seems impossible to use all these languages simultaneously. There is not a single language used in day to day communication in the world. However a common language enables human beings to work together in an infinite variety of ways.

1.1.1   The English Language

From the different books of linguists', we found that there are now around three thousand different languages in the world. These

2

languages are spoken under certain language families. These language families are Hamito-semitic, Ural-Altaic, Japanese-Korean, Dravidian, Niger-Congo, Malayo-Polynesion, Indo-European ,Sino-Tibetan etc. The Germanic language, Indo-Iranian language Greek language, Celtic and Italic language, Baltic and Slavonic language etc, are the branches of Indo-European language family. Thus, Indo-European language family is an enormous and complicated language family. The Indo-European family is by far the biggest language family in the world. The English language belongs to the same branch of Indo-European language family, Germanic language, the most wide spread, the second most spoken language in the world (see Barber 1999).

English Language is an international language for Nepali learners. It has been used as a dominant language all around the world. The use of English language can be found in every sectors where language is needed, for example, communication, commerce, business, industry, literature, politics, education, science and technology etc. Without English language we cannot interact with other and thus it has been a means of communication. Nowadays, the need and scope of English language has been increasing day by day. The improvement in educational policy easily reflects the importance of English language in our context.

1.1.2   English Language Teaching

Teaching of English language is taken in a global perspective. The rapidity of global spread of English language has come as a surprise and so more dramatic than anticipated. English language is taught and learnt as a foreign language in Nepal. It was first introduced in Durbar school (1910 B.S.) by Janga Bahadur Rana who was influenced by British Education system, Sharma (1999, p. 36). That

3

is why, it can be said that teaching of English language started since Rana regime. Nowadays, English language is taught as a compulsory subject in most of the educational institutions. It is one of the complicated tasks for Nepali learners. However, the attraction of pupils towards English language can be easily seen due to the establishment of boarding schools these days. The main purpose of teaching English language at school level is to make familiar via the basics of English language and to provide learners with a linguafranca that give them an access to the international contacts.

The National Education Policy Commission(1992) laid a great emphasis on the regular updating and improvement of the English language curriculum. On the basis of lower secondary level English curriculum, teaching of English language in school level has two main purposes. Firstly, to enable pupils to exchange ideas with people who speak or write English. Secondly, to expose them to the vast treasures of knowledge and pleasure available in English, both written and spoken. Similarly, the general objectives of teaching English in the lower secondary level are as follows:

) To enable students to communicate in speech and writing with other people who speak English,

) To enable students to understand and use the language functions prescribed in the curriculum,

) To enable students to gain skills for developing competency in reading different types of written materials,

) To enable students to develop an enthusiasm for and appreciation of reading both for information and enjoyment.

For the purpose of achieving these objectives we need to mention the course contents of English grade 8, which is given in appendix I.

Teaching of English language includes the teaching of language four skills: listening, speaking, reading and writing. If a language is taught and learnt, the learner must be competent in these four skills. Therefore, we need a procedure in order to judge the learners' achievement, which are evaluation/ testing.

1.1.3 Language Testing

Testing is an integral part of teaching. We cannot separate testing from teaching. Teaching and testing are the two parts of the same coin. Testing is inevitable for we have to assess the achievement of the students as well as the effectiveness of the programme. Heaton (1975, p. 5) mentions: "Language testing, in the past, was considered as a separate entity from teaching. But both testing and teaching are so closely interrelated that it is virtually impossible to work in either field without being constantly concerned with the other". Testing is conducted for finding out the performance of the students of the purpose of comparison, selection, gradation etc. It is used to find out how far learners have learned what the teacher wishes them to learn, whether the learner understands what has been taught or not.

With the help of language testing we hope to obtain information about students' achievement. That is why, the teacher will be able to compare, select, grade and provide certificate among the students with the help of language testing. We can also evaluate the effectiveness of teaching style, syllabus, methods, approaches, techniques, materials etc. with the help of testing. Therefore language testing almost never takes place in isolation. It is done for particular purposes.

⟩ To measure students proficiency in certain language course;

5

⟩ To discover how far students have achieved the objectives of course of study;

⟩ To discover students' strengths and weakness, to identify what they know and what they do not know;

⟩ To assist placement of students by identifying the stage or part of a teaching programmers most appropriate to their ability;

⟩ To motivate the students towards their study;

⟩ To provide the students with an opportunity to show their ability to recognize or produce correct form of the language;

⟩ To determine their readiness for instructional programmers';

To fulfill these objectives, students are required to take part in examination. Examinations are administered for some purposes too. In order to serve the purposes for which exams are conducted, they must be of good quality. The quality of a test is examined in light of the extent to which it serves the purpose for which it is administered (Khaniya, 2005, p.101). Therefore, it is necessary to mention the qualities of a good test here.

1.1.4 Qualities of a Good Language Test

Several things should be taken into consideration while constructing a test. If so, what are the qualities of a good test? There are different views on what makes a test good. Some scholars point only a few qualities and some other point many qualities but there are some qualities, which are common to all. These qualities are as follows:

a. Validity
b. Reliability
c. Administrability

d. Scorability

e. Economy

f. Backwash/washback

1.1.4.1 Validity

Validity is an essential characteristics of a good test. It refers to the degree of accuracy of a test. 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 purposes for which it is administrated (Khaniya, 2005, p.103).The detailed description of this quality can be seen later on, which is taken for the present study.

1.1.4.2 Reliability

Reliability is another important characteristics of a good test. It 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 retest of students and marking and remarking by the examiners. If the scores of two tests are similar or almost similar at that time we can say that the test is more reliable "In order to be reliable a test must be consistent in its measurement" (Heatons, 1975, p.162).It is assumed that the performance of the examinee remains more or less the same if the examinee is asked to take the same exam of the similar types. "The more similar the scores would have been, the more reliable the test is said to be" (Hughes 1989, p. 29).Therefore, in order to find out the reliability of any test we need to employ some particular methods. The methods are test-retest method, split-half method, parallel-forms method, inter-intra-marker method, kuder

Richardson and Cranach alpha method. As for example in the examination of ELT at M. Ed level (within a reasonable time), the paper checked by P.N. Campus lecturer gets 65 marks out of 100 marks and if the same paper rechecked by Tahachal Campus lecturer secures 50 marks out of 100 marks then that examination does not have the quality of reliability. Personal judgement may differ in the range of 2-4 marks, but vast difference in obtained mark, does not reflect the quality of reliability. To be a test reliable, it needs consistency in the score whoever is involved in checking.

1.1.4.3 Administrability

The third quality of a test administrability, also called practicality refers to the degree to which a test is adaptable in varying situations. It also indicates the well and advanced organization of how long the test will take, what special arrangements have to be made, what equipments are needed and how the test should be practically fit for the situation so that it can be easily administered. 'In order to develop a test with a reasonable degree of practicality, it is necessary to pay attention to human resource, material resource and time' (Khaniya, 2005, p.120). For example, using video for testing at primary level in present context of Nepal is not practical.

1.1.4.4 Scorability

Scorability is one of the qualities of a good test. It refers to the proper allocation of marks regarding calculation as well. It is important to consider while designing the test how far it is practical and possible to look for important points for marking the students' performance." Can the test be scored with ease so that the users may be able to handle it?"
(Lado, 1965, p.31).It is accepted that subjective tests are not easy to score and objective tests are easy to mark. Whether the test are

subjective or objective we should consider all the factors while designing a test. As for example, if there are 5 sub questions in one item of 3 marks, then it is very difficult to score.

1.1.4.5 Economy

Economy is one of the practical criteria of a good test. Regarding this, here is one question, "Does the test measure what we want it to test in a reasonable time considering the testing situation? If it does, the test is practical and economical" (Lado, 1965, p. 31).Any test to be a good one, should be economical in all senses. Economy of a test refers to the construction administration and scoring of test at minimum cost and labour.

1.1.4.6 Washback/ Backwash

Washback is the effect of examination in teaching and learning. Hughes, (1959, p. 1) says, "The effect of examination on teaching and learning is known as washback. It can be harmful or beneficial". It is said that washback is an inherent quality of an examination. Whatever is done all along the way of examination preparation is the washback effect of the examination. The reason is that the exam will shape the preparation of students. It also influences students' learning activities outside the classroom in the sense that the way teachers prepare students for examinations at home, are not unaffected by examinations. (Khaniya, 2005, p. 56-60)

Any test is said to be having good backwash effect if it exerts a good influence on the teaching and learning that take place before the administration of the test. For example, rote learning was encouraged by the old S.L.C. English Examination System in Nepal and that could be regarded as the bad washback effect but now no

such washback effect is there. In the current system of S.L.C. examination in Nepal students should develop actual language skills for the preparation of examinations. Therefore, they are encouraged to learn the language rather than to prepare for the examination. This can be regarded as a good washback effect.

That is why, washback effect influences on students' learning behavior and it helps to provide feedback to all who are involved in teaching sector. So it has become one of the essential and inherent qualities of examination.

This study is about a type of validity of grade eight test. So, it is needed to describe about this topic here.

1.1.5 Validity

Validity is one of the important qualities of a test. The validity of a test is the extent to which it measures what it is supposed to measure. In order words, validity is the degree to which the test actually measures what it is intended to measure. "Validity refers to the appropriateness of the interpretation and use made of the results of an examination procedure for a given group of individuals not to the procedure itself"(Linn and Miller, 2008, p. 72). We sometimes speak of the "validity of a test" for the sake of convenience but it is more correct to speak of the validity of the interpretation and used to be made of the results. It involves an overall evaluative judgment. It requires an evaluation of the degree to which interpretations and uses of examination results are justified by supporting evidence and in terms of the consequences of those interpretations and uses. It is always specific to some particular use or interpretation for a specific population of test takers no examination is valid for all purposes'.

As has been explained (Linn and Miller, 2008, p. 73), it is a matter of degree; it does not exist on an all-or-none basis. Validity is best considered in terms of categories that specify degree such as high, moderate and low validity. Validity is a unitary concept. Although evidence may be accumulated in many ways, validity always refers to the degree to which that evidence supports the inferences that are made from the scores. The inferences regarding specific uses of a test are validated, not the test itself.

Validity is the most important consideration for the selection and use of any testing procedure. A valid test reflects the real image of test what is intending to measure. Every test, whether it be a short, informal classroom test or a public examination, should be as valid as the setter can make it. The test must aim to provide a true measure of the particular skill which it is intended to measure to the extent that it measures external knowledge and other skills at the same time, it will not be a valid test. For example, the following test item is invalid if we wish only to measure writing ability; 'Is photography an art or a science? Discuss'. It is likely to be invalid simply because it demands some knowledge of photography and will consequently favour certain students (Heaton, 1975, p. 159)

According to Patel (2004, p. 167) there are some factors which affect the validity of any test. These factors are:

ʃ unclear direction;
ʃ reading vocabulary;
ʃ difficult sentence construction;
ʃ poorly constructed test items;
ʃ use of inappropriate items;
ʃ medium of expression;
ʃ difficulty level of items;

) influence of extraneous factors;

) inappropriate time limit;

) inadequate coverage; and

) Inadequate weightage;

With the purpose of reducing these factors, we can classify validity into four types.

1.1.5.1 Types of Validity

There are four types of validity, they are:

- Content validity
- Construct validity
- Criterion-related validity
- Face validity

1.1.5.2 Content Validity

Content validity is a kind of validity which is related to the objectives of content. Patel (2004, p. 168) mentions that, "content validity refers to the process how well the sample of assessment tasks represent the domain of tasks to be measured and how it emphasizes the important content". This kind of validity depends on a careful analysis of the language being tested and of the particular course objectives. Heaton(1915, p.160) says, "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". It is a matter of analyzing the content and tasks included in the measuring instrument and the domain of outcomes to be measured and judging the degree of correspondence between them. A test is said to have content validity if it constitutes a representative sample of the language skills, structures with which it is meant to be concerned. According

to Linn and Miller(2008, p. 73), the goal in the consideration of content validation is to determine the extent to which a set of examination tasks provide a relevant and representative sample of the domain of tasks about which interpretations of examination results are made. Content validity should answer the question. "Does the test measure the instructional objectives?" A comparison of test specification and test content is the basis for judgments to content validity Harrison,(1991, p. 51) argues that "The content validity can be achieved by considering the purposes of the assessment and then drawing up a content list known as a content specification."

It is necessary to mention here that content validity includes content coverage and content weighting. Comparison between course objectives/contents and test contents is the key to find out the content validity in terms of content coverage of the test. Content coverage is the extent to which the tasks required in the test adequately represent the behavioral domain in the questions. Similarly comparison between course content weighting or evaluation scheme or specification grid and test content weighting are the keys to find out the content validity in terms of content weighting of the test. Weighting is the distribution of marks according to the specified contents. In order to find out the content weighting of the test, we need something i.e the evaluation scheme or the specification grid of the relevant subject.

The specification grid for grade 8 is given in appendix II.

1.1.5.2 Construct Validity

Construct validity refers to how well the test items are designed so as to measure the particular skill or aspect. Construct validity is determined by finding whether test results correspond with scores

on other variables as predicted by some rationale. Heaton (1975, p. 161) argues that, "If a test has construct validity, it is capable of measuring certain specific characteristics in accordance with a theory of language behavior and learning". For example, if a communication approach to language teaching and learning has been adopted throughout a course, a test comprising chiefly multiple-choice items will lack construct validity

1.1.5.3 Criterion -Related Validity

Criterion-related validity is also called empirical validity. Empirical validity is of two types; concurrent validity and predictive validity, depending on whether the test scores are correlated with subsequent or concurrent criterion measures.

Concurrent   Validity

Concurrent validity is a kind of criterion-related validity. "Concurrent validity refers to the process of determining the validity against the set criterion at the same time. The correlation between the two tests is said to be the concurrent validity of the new test." (Khaniya,2005,p.108). It is concerned with how closely test scores are correlated with present criterion behavior. For example, if an examinee gets 70% marks in 1st examination and how much marks would he obtain in 2nd examination that is concurrent empirical validity

Predictive Validity

Predictive validity of a test is concerned with the degree to which a test can predict the future performance of the candidates. This type of validity is established by comparing test results. It is related to how well test scores predict future criterion behavior. For example, if an examinee gets 60% marks in an examination, on the basis of the

scores, how well he would do in teaching performance or in other skill that is predictive validity.

1.1.5.4 Face Validity

If a test item looks right to the other testers, teachers, moderators and testers, it can be described as having face validity. Face validity simply refers to the cosmetic value of a test. It is the validity of a test in layman's eye. Harrison (1991, p. 53), argues that face validity is concerned with what teachers and students think of the test. Does it appear to them a reasonable way of assessing the students or does it seem trivial or too difficult or unrealistic? The concept of face validity is relatively new. Most of the designers regard face validity as the most important of all types of test validity. If a test has not face validity it may increase hesitation, boredom and non-interest to them who are involved in this test.

1.2   Review of the Related Literature

This research is not directly related to other topics which were previously carried out. In the department of English language Education. Some of the studies are nearer which are reviewed below:

Batala (2004), carried out a research on "A Study on the Validity of the SLC Examination English Question Paper". The objective of his research was to examine the predictive and content validity of the SLC examination English question paper. He administered three different sets of questionnaires, one to the secondary level English teachers, the next to the students and the last to the question setters. He found out that the SLC examination English question papers had very low predictive validity and the content representation of the test is high.

Similarly, **Bhandari (2004),** conducted research entitled, "A Descriptive and Attitudinal Study on SLC English Questions and Specification Grid 1999". The sample population of his study were the students of SLC examination of that year. He collected the question of English paper and dealt with specification grid. When he analyzed the questions he found that the questions are less adequate in terms of levels and interest of students.

Ojha (2005), conducted research entitled "Content Validity of ELT Theories and Methods Exam at B.Ed level". The objective of his research work was to examine the content validity of ELT Theories and Methods tests. To fulfill the objective, he analyzed the test papers administered in the annual examination from 2057 to 2061. The findings of his research work is that in terms of content coverage the test paper have good content validity but in terms of content weighting the test papers lack content validity.

Likewise, **Bhattarai (2005),** studied on "The Content Validity of Compulsory English Textbook for Grade Eight". The objective of her research work was to examine the content validity of English text book for grade eight. She has found that the textbook has less content validity in terms of interest and level, listening text or objectives, speaking, reading, writing exercise and communicative function of language.

Furthermore, **Subedi (2006),** carried out a research on "Content Validity of B.Ed Questions, A Case of English Sounds and Structures(302)". To accomplish the specified objectives a set of questionnaire was prepared for the purpose of collecting the teachers' responses. The sample population consisted of 20 teachers; 15 from kathmandu,2 from Bhaktapur, 2 from Kavre and 1 from Lalitpur district. The result of the study indicated that the

question papers of the written examination of the course are not as valid and objectives oriented as they should be. They do not seem to follow the evaluation scheme of the syllabus properly.

In the same way, **Nepal (2006),** carried out the research on "Content Validity of Examination; A Case of Fundamentals of Language and Linguistics at B.Ed Level". To accomplish this task the question papers administered in the annual examination during 2057 to 2062 were selected for the collection of data and analyzed in terms of content coverage and content weighting. He found that in terms of content coverage the tests have high content validity but in terms of content weighting the tests have low content validity.

Similarly, **Timilsina (2007),** carried out a research on "Content Validity of Phonetics and Phonology Exam at M.Ed Level". He analyzed the five years question papers from 2058 to 2062 of phonetics and phonology. The major findings of his research are that the tests have low content validity in terms of content coverage and low content validity in the weightings principle as well.

Gurung (2008), studied on "Content Validity of Semantics and Pragmatics Examination at M. Ed Level". He analyzed the five year question papers from 2058 to 2062. His major findings were in terms of coverage the content validity of semantics and pragmatics was in average (neither low nor high) but in terms of weightage it has poor content validity.

Moreover, **Karki (2009),** carried out a research on "Content Validity of General Linguistics Course of Grade XI". He found that according to the coverage or representativeness principle, general linguistic question papers have high content validity and low content validity in terms of content weighting. He concluded that mark distribution of each unit is not given in the syllabus.

Evaluation scheme chart is given but it was not duly followed by the question setter. And no any standard procedure was followed while allocating marks to the test items.

Finally, **Sharma (2010),** studied on "An Appraisal of Grade Five English Textbook in Terms of Content Validity". He found that the textbook has content validity in terms of content coverage and applicability to everyday work. He concluded that the contents of grade five textbook are well graded and there are also varieties in the selection of information in the textbook.

The above mentioned research works are related to the present topic in one or another way. Firstly, all of them are on testing and evaluation. Secondly, some of them are on content validity of textbook and thirdly, some of them are on content validity of examinations or questions. But this research work is different from others because no research has been done yet to find out content validity of district level questions of compulsory English grade eight.

1.3   Objectives of the Study

The objectives of the proposed study were as follows:

a. to examine the content validity of compulsory English grade eight test in terms of :
−content coverage;
−content weighting
b. to suggest some pedagogical implications.

1.4   Significance of the Study

The present study provides information on the content validity of compulsory English grade 8 examinations. This research work is a distinct research work in the Department of English Language Education because no research has been done yet in the present topic. So, it will be most important and useful for the department.

So, the proposed study will be useful to all those who are concerned with language teaching, testing and learning. Specifically, the teachers and test designers who are involved in teaching and testing will get insight into the validity level of the administered tests in district level questions. Moreover, this research work will be significant to question setters, students, examiners, examination administrators and so on. Furthermore, this research work will be significant to other researchers who want to carry out research works in this field.

1.5 Definition of Key Terms

Validity: Validity is the degree to which a test measures what is supposed to measure, or can be used successfully for the purposes for which it is intended (Richards et. al., 1985, p. 304).

Content Validity: 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." (Richards et. al., 1985, p. 61).

Washback: Washback is the effect of testing on teaching and learning. It can be beneficial or harmful.

Content coverage (Representativeness):

It is the extent to which the tasks required in the test adequately represent the behavior domain in the questions.

Content Weighting: It is the distribution of marks according to the specified grid.

Reliability: It is a measure of the degree to which a test gives consistent results ((Richards et. al., 1985, p. 243).

CHAPTER-TWO

METHODOLOGY

To accomplish this study she has used secondary sources of data on the basis of judgmental sampling procedure.

2.1   Sources of Data

There are two types of sources of data viz. primary and secondary but she has used only secondary sources of data for this proposed study.

2.1.1   Secondary Sources of Data

In order to fulfill the intended objectives, the researcher used only the secondary sources of data. The researcher used the question papers of compulsory English grade 8 from the years 2063 to 2066. Besides this, she consulted the syllabus of lower secondary level education, various books, articles journals and research reports. Some of them include Heaton (1975), Richards (1985), Khaniya (2005), Hughes (2003), Nepal (2006), Karki (2009).

2.2   Sampling Procedure

She employed judgmental sampling which is one of the most useful non-probability sampling designs. For this study, the researcher collected four question papers which are administered in district level from (2063 to 2066). So, 4 question papers of concerned examinations were collected judgmentally and were taken for the analysis of the study.

2.3   Tools for Data Collection

The researcher prepared the checklist for data collection. In that checklist, she prepared the areas of contents on the basis of specification chart and syllabus

of the given course. Four question papers of the test used previously were analyzed.

2.4   Process of Data Collection

The researcher collected question papers administered in the district level examination of compulsory English from 2063 to 2066 and the syllabus. Then, she categorized the test items (in terms of grammar, reading skill and writing skill) based on the course contents given in the syllabus and prepared a checklist. To find out content coverage of the course, she compared the test items with the course contents in the checklist and listed the instance of the question papers and she counted them. Finally, to find out content weighting, she has prepared the checklist according to specification grid and observed the test weightage.

2.5   Limitations of the Study

The study has the following limitations.

a. The study was limited on district level compulsory English questions of grade 8.

b. The study was limited to the question papers administered from the years 2063 to 2066.

c. Only the question papers of Kaski district were analyzed.

d. Only tables and percentage were used as the statistical tools for data analysis.

e. The questions were analyzed only (in terms of grammar, reading skills and writing skills) content validity of questions.

CHAPTER - THREE

ANALYSIS AND INTERPRETATION

This chapter is the central part of the study since it is concerned with the analysis and interpretation of the raw data obtained from the study to examine the content validity of 'compulsory English grade 8'. For this purpose, it has been divided into two parts. The first part deals with the analysis of the content validity of compulsory English grade 8 question papers in terms of coverage and the second part deals with the content validity of the same question papers in terms of weighting.

3.1 Analysis and Interpretation of Content Representativeness (Coverage)

The question papers of compulsory English grade 8 administered in district level examination from 2063 to 2066 are analyzed in terms of their coverage. Content representativeness/coverage is the comparison between course contents/objectives and test contents. So, the researcher examined the test papers to examine whether the test items are the representative sample of the course objectives or not.

The researcher has analyzed the question papers only in terms of two skills reading and writing skills. So the listening and speaking skills are beyond analysis which bears 30 percent marks out of 100 marks. Regarding reading skill, writing skill and grammar carries 70 percent marks out of 100 marks. Therefore, in this study the questions with 70 full marks have been analyzed.

The course content and the specification grid for grade 8 and the four years' question papers are given in appendices.

3.1.1 Examining Course Representativeness Regarding Grammar

On the basis of course content, there are 10 grammatical items. She compared these items with the test items whether they are covered or not in these four years' (2063 to 2066) examinations. Let's check by this table.

Table 1

Grammar: Course Coverage

Course contents		Test contents			
		Test Items Represented			
		2063	2064	2065	2066
Grammar	Tense	✓	✓	✓	✓
	Preposition	✓	✓	✓	✓
	Articles	✓	✓	✓	✓
	Relative clauses (who, which, that etc)				
	Voice	✓	✓✓	✓✓	✓✓
	Reflexive pronouns	✓	✓	✓	✓
	Reported speech	✓✓	✓	✓✓	✓
	Question (yes/no, tag-question, wh-question)	✓✓	✓✓	✓	✓
	Command				✓
	Statement	✓	✓	✓	✓

Note : ✓ = covered the language contents by test contents

✓✓ = double coverage

Table 1 shows that in case of grammar in grade eight there are ten grammatical items. These ten items are: tense, preposition, article, relative clauses, voice, reflexive pronouns, reported speech, question (yes/no question,

tag-question, wh-question) command and statement. Within these four years (2063 to 2066) question papers, relative clauses had no place in the tests.

One question was asked from 'tense item' in all these four years similarly, one question from preposition, articles, reflexive pronouns and statement was asked in all these four years' tests. In case of 'voice item' in the year 2063, only one question was asked but to make passive into active and active into passive both type (two questions) were asked in the years. 2064, 2065 and 2066. Similarly, 'reported speech item' to make direct speech into indirect speech and vice-versa (two questions) were asked in the years 2063, 2065 and 2066 examination but only one question in 2064 test. Likewise, other grammatical items, question formation (yes/no question) was asked in the years 2063 and 2065. But wh-question was asked in the years 2064 and 2066 examinations.

Thus, it is found that, there are ten grammatical items according to course content/grid but the representation of the test items are nine grammatical items. One grammatical item was neglected while designing test items. It means the coverage of course contents in test contents of grammatical item is 90 percent. But 10 percent of course of contents was not covered in the question papers. To sum up, grammatical part has occupied very good place in content validity because it (question papers) covers 90 percent course contents.The above description can be presented in the pie-chart as follows:
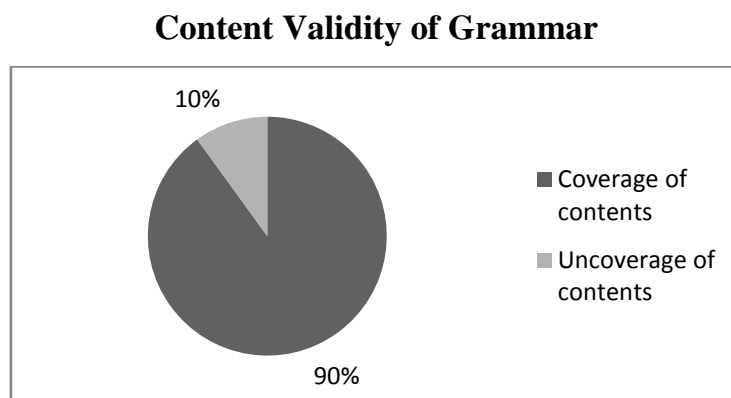
Content Validity of Grammar

Figure 1

Thus, it is found that, 90% course contents of grammar were covered and other 10% course contents were not covered.

3.1.2 Examining Course Representativeness in Reading Skills

On the basis of course content, there are 20 reading topics. In order to examine content validity of test she compared these reading topics and test items by the help of this table.

Table 2
Reading Skill: Course Coverage

Course Contents			Test Contents			
			Test Item Represented			
			2063	2064	2065	2066
Reading skill	Comprehension Question form Textbook	⟩ Short answer questions	Tp 3rd Tp 5th Tp 8th Tp 14th Tp 15th Tp 9th Tp 18th	Tp 3rd Tp 5th Tp 16th Tp 5.5th Tp 6th Tp 7th Tp 13th	Tp 2nd Tp 3rd Tp 7th Tp 8th Tp 13th Tp 19th Tp 18th	Tp 3rd Tp 8th Tp 13th Tp 14th Tp 15th Tp 18th Tp 20th
			(2-opt)	(2-opt)	(2-opt)	(2-opt)
	Seen Reading Passage	⟩ Short answer questions	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓
		⟩ Multiple choice			✓✓✓✓	
		⟩ Fill in the gaps				
		⟩ Matching				✓✓✓✓
		⟩ Ordering				
	Unseen Reading Passage	⟩ Short answer questions		✓✓✓	✓✓✓✓	
		⟩ Multiple choice	✓✓✓✓	✓✓✓✓		✓✓✓✓
		⟩ Fill in the gaps				
		⟩ Vocabulary	✓✓✓✓	✓✓✓✓✓	✓✓✓✓	✓✓✓✓
		⟩ Matching	✓✓✓	✓	✓✓✓✓	✓✓✓✓
		⟩ True/false item				
	Vocabulary	⟩ Use of words/phrases in sensible sentences	✓✓✓✓✓ ✓	✓✓✓✓✓ ✓	✓✓✓✓ ✓✓	✓✓✓✓ ✓✓
			(4-com.)	(4-com.)	(4-com.)	(4-com.)

Note:-

Tp = reading topic

✓ = question coverage

2-opt. = two optional questions

4-comp. = four compulsory questions

Course objective: to read text extensively, show understanding of the main theme and guess unknown words by context.

Table 2 shows that there are altogether twenty reading topics. Out of these twenty reading topics four reading topics (1st , 4th, 12th, 17th) have not been represented in any years' examination during that four years. The third reading topic was more representative item repeating in all the four years' examinations. Similarly, the 5th, 7th, 8th, 13th, 14th, 15th, 18th reading topics were partially represented in the four years' examinations. The 9th, 19th, 20th and 16th reading topics have been represented only once in four years' question papers. From the side of comprehension questions from the textbook, there were altogether seven-short-answer questions, out of them five questions were compulsory and other two were optional.

In case of seen reading passages of 2063, 2064, 2065 and 2066 four short-answer questions were asked respectively from 11th, 14th, 13th and 19th reading topics. But in case of 2065 'fill in the gaps' item was asked and in the year of 2066 'ordering' item was also asked.

In case of unseen reading passage, 'matching item' was asked in all the four years examination. 'True false item' was also asked in three years examinations beside 2064. Similarly, 'fill in the gaps' item was also asked in the three years' examinations beside the year of 2065. But short-answer questions from unseen reading passage were asked only in the year 2064 and 2065 but not in 2063 and 2066 examination.

Likewise, reading skill also includes vocabulary item. These vocabulary items were not asked from seen and unseen reading passage in these four year

questions. Four words/phrases have been asked to use in sensible sentences out of 6 options. In place of vocabulary items, the same type of question was set in all the four years' examinations.

In conclusion, from the above table and description the researcher found that there are altogether 20 reading topics according to the course contents. Out of 20 reading topics, 16 reading topics were set in the tests during four years. But four reading topics (1st, 4th, 12th, 17th) were ignored while designing test items. It means, the coverage of the course contents in test contents of reading skills is 80 percent. But 20 percent of course contents were not covered in the question papers. To sum up, reading skill has high content validity because it (questions papers) covers 80 percent course contents.

The above description can be shown in the pie-chart as follows:
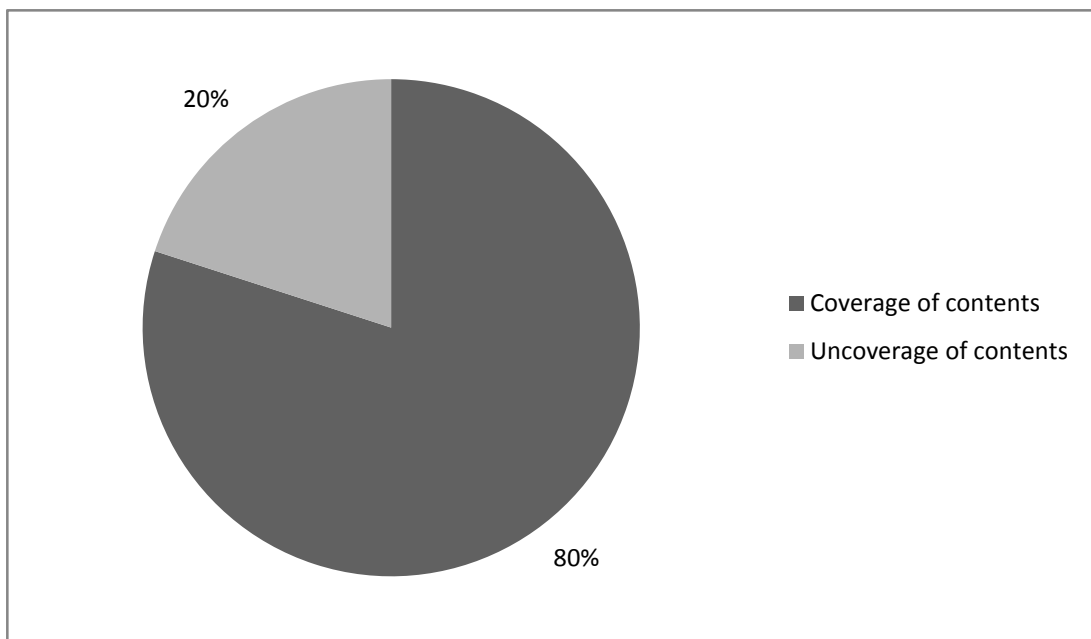
Content Validity of Reading Skill

20%

80%

■ Coverage of contents

■ Uncoverage of contents

Figure 2

Thus, it is found that, 80% course content of reading skill covered and other 20% course content were not covered.

3.1.3 Examining Course Representativeness in Writing Skills

Regarding writing skill there are three writing processes concerning with all the 20 reading topics. Here, she compared these language items with test items by the help of this table.

Table 3

Writing Skill: Course Coverage

Course Contents			Test Contents			
			Test Item Represented			
			2063	2064	2065	2066
Writing skill	Punctuation	⟩Full stop	✓✓		✓✓	✓✓
		⟩Question mark				
		⟩Commas	✓✓	✓✓	✓✓✓	✓
		⟩Exclamation				
		⟩Spelling	✓	✓✓		✓
		⟩Inverted comma	✓	✓✓	✓	✓
		⟩Apostrophe	✓✓	✓✓	✓✓	✓✓
		⟩Capitalization				
	Guided writing	⟩Story from pictures or clues	✓	✓	✓	✓
		⟩Completing a dialogue				
		⟩Note making				
		⟩Description of movement				
	Free writing	⟩Application				
		⟩Greeting cards				
		⟩Paragraph writing				
		⟩Reporting facts & events				
		⟩Letter writing	✓	✓	✓	✓
		⟩Thank-you note				
		⟩Writing instructional leaflets				
		⟩Essay writing	✓	✓	✓	✓

Note: ✓ = question coverage

Table 3 shows that, on the basis of the specification grid grade 8, there are altogether three processes of writing skill. The three process of writing skill are concerned with all the twenty reading topics as the course content suggests. If we see diachronically, among these three writing process (punctuation, guided writing and free writing) punctuation was the most

representative skill. Regarding punctuation, capitalization, inverted commas and commas were asked in all the four years' examinations. Full stop was also asked in the examinations excluding 2064. Similarly, correction of spelling was required in 2063, 2064, 2066 examination besides 2065. Likewise, apostrophe was asked only in the year 2066. Besides them, 'the question mark' and 'exclamation' which were mentioned in the table, were not asked in any years' examination during these four years.

In case of guided writing, 'story on the basis of given outlines' was asked in all the four years examination. Similarly, from the side of free writing, 'letter writing' and 'essay writing' were asked in all the four years' examinations. Besides them, other written items, which were mentioned in the above table have not represented in any years examination during these four years.

Thus, it is found that there are altogether three processes of writing to reflect the course objectives 'to write: a simple short story with the help of given clues, simple short notes, brief newspaper report, short essays, personal letter, and to prepare a dialogue on a given situation'. These three processes of writing skill are somehow concerned with all the twenty reading topics as the course content suggests.

According to the specification grid grade eight, there are altogether twenty writing items, out of them only 10 writing items were represented in the tests during four years. It means, the coverage of course content in test content is 50 percent of writing skill. 50 percent writing items were not covered in the question papers. To sum up, the writing skill has average content validity because it (question paper) covers 50 percent course content.

Which can be shown in the pie-chart as follows:
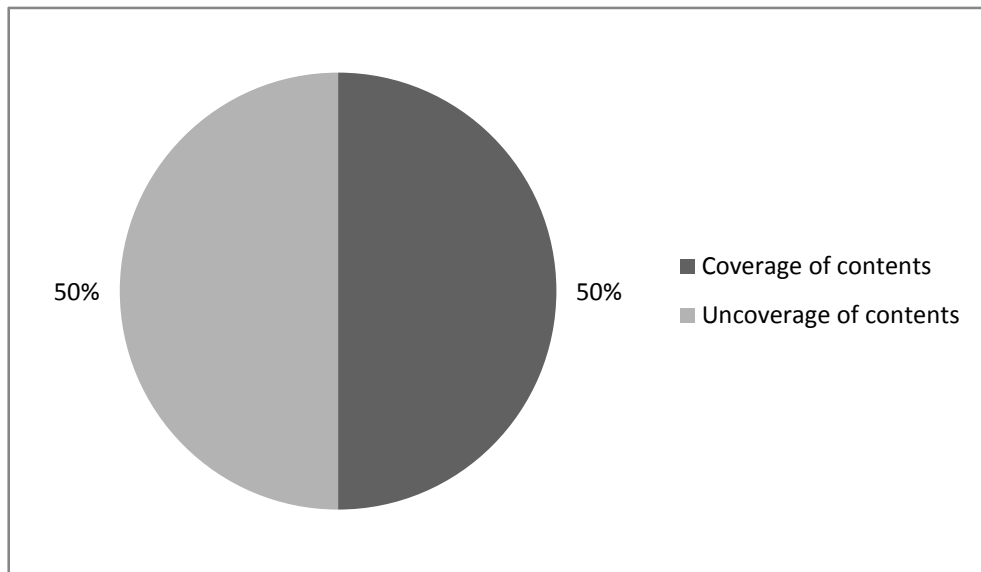
Content Validity of Writing Skill

Figure 3

Thus, in case of writing skill, 50% writing items were covered and other 50% writing items were not covered.

3.1.4 Examining Content Validity of the Question Papers in General in Terms of Representativeness Coverage

Table 4

Course Coverage as a Whole

S.N.	Aspect/skill	Course contents	Test contents	Ignore items	Test coverage
1	Grammar	10	9	1	90.0
2	Reading skill	20	16	4	80.0
3	Writing skill	20	10	10	50.0
	Total	50	35	15	70.0

The above table shows that on the whole regarding grammatical items, reading skill and writing skill, there are altogether 50 languages items, the representative of the test content language items are 35. And 15 language items were neglected while designing the test items. It means that, coverage of contents in the whole aspect/skill is 70 percent but 30 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:

Content Validity in General

Legend:
- Coverage of contents
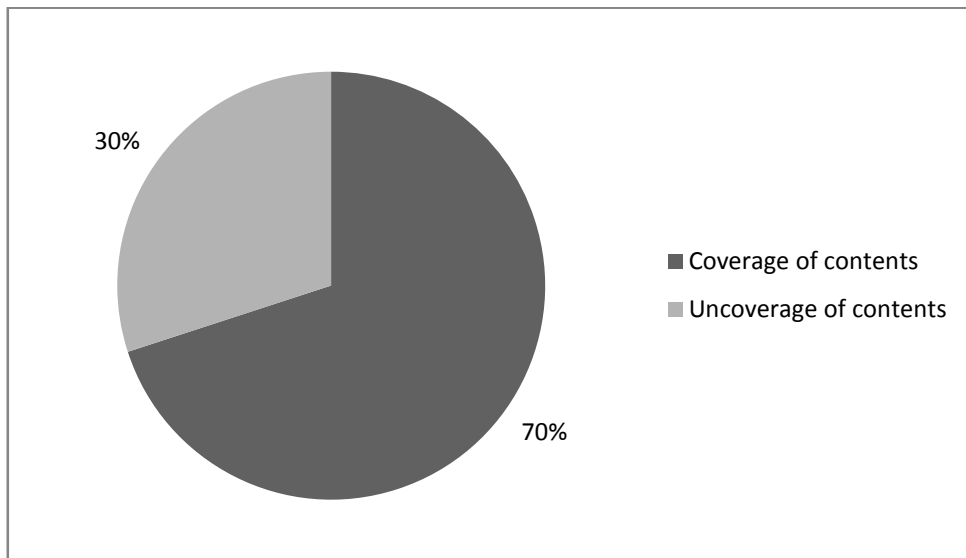- Uncoverage of contents

30%

70%

Figure 4

After analyzing the above given figure, the researcher concludes that the content validity of the test papers, on the whole, in terms of content coverage is high because it covers 70 percent of the course content and other remaining 30 percent course is not covered. That is to say, these test papers (somehow except writing skill) have really tested what they were supposed to test in the question papers.

3.1.5 Comparison of the Four Years' Question Papers in Terms of Content Representativeness/Coverage

The comparison of grammar, reading skill and writing skill within these four years' examinations, shows that the least representative skill is writing skill one which covers only 50 percent of the course contents. But reading skill 80 percent and grammatical items were represented 90 percent of the course contents.

This can be presented in the bar diagram as follows:
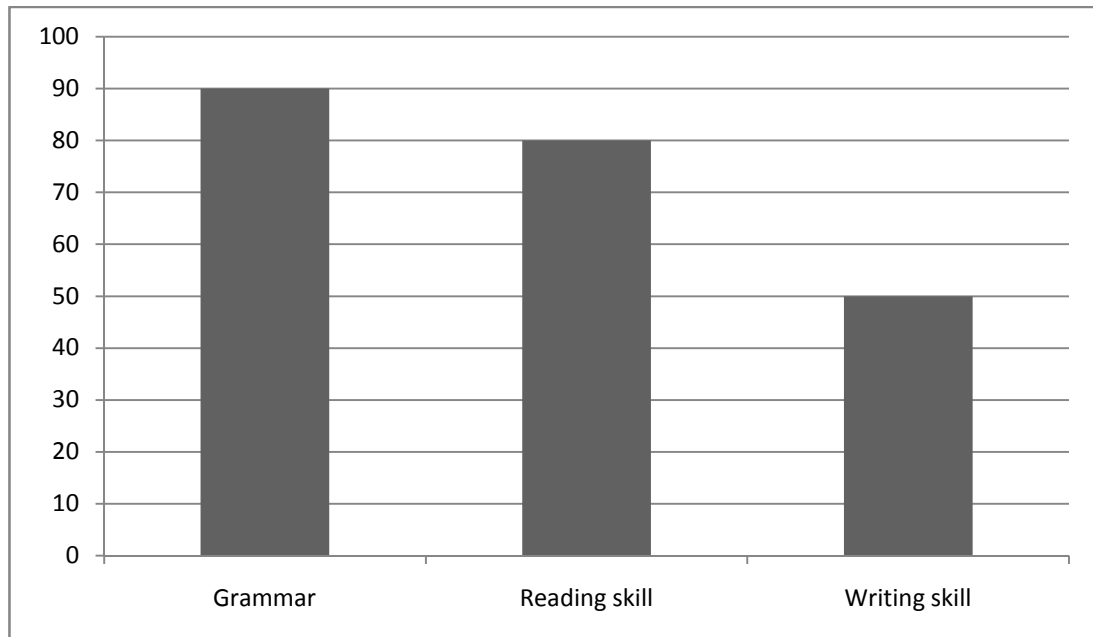
Comparison of Content Coverage

Figure 5

It reflects that, the content validity (coverage) of grammar is high, the coverage of reading skill is only 80% and the writing skill has only 50% coverage of the contents. It means writing skill has average course coverage.

3.2 Content Weighting

This is the second part of this chapter which deals with content weighting. Content weighting is the comparison between weighting of the test contents and the weighting of the course contents. Here, for the purpose of examining content validity of the tests, the researcher has examined whether or not the marks allocated to each topic of the course content/in the grid is proportional to the weighting of the test contents administered during 2063 to 2066. It is believed that if test contents strictly follow the specification grid, question papers have high content validity in terms of weightage.

According to the specification grid the weightage of grammar, reading skill and writing skill is as follows:

Table 5

Content Weighting

Aspect/skill		Weightage
Grammar		10
Reading skill	⟩Text based question	10
	⟩Seen passage	8
	⟩Unseen passage	8
	⟩Vocabulary	4
Writing skill	⟩Punctuation	4
	⟩Guided writing	6
	⟩Free writing	20
Total		70

Table 5 indicates that the weightage of grammatical items, reading skill and writing skill according to the specification grid is 70 marks. As a whole, grammatical items cover 10 marks, reading skill represents 30 marks and writing skill also 30 marks. Therefore, the question papers were analyzed only up to 70 percent marks of the course.

3.2.1 Examining the Four Years' Question Papers Grammar in Terms of Content Validity (Weightage)

In order to examine content weightage, she has compared the specification grid and allocated marks for grammar given in the tests by the help of this table.

Table 6

Content Weighting of Grammar

Weightings of the course contents			2063			2064			2065			2066		
			No. of question	Weightage	Total weightage	No. of question	Weightage	Total weightage	No. of question	Weightage	Total weightage	No. of question	Weightage	Total weightage
Grammar (10 marks)	Full weighting	Fill in the gaps												
		Multiple choice	5	1	5	5	1	5	4	1	4	4	1	4
		Transformation	5	1	5	5	1	5	6	1	6	6	1	6
Total			10	1	10	10	1	10	10	1	10	10	1	10

Table 6 shows that out of 10 marks, from grammatical items 10 questions can be asked according to the specification grid. Researcher compared the weighting of the question papers with the weighting of the specification grid and found that the weightage of grammatical items was exact. The question designer has kept in mind the weightage of the grammatical items while constructing question papers during these four years period.

3.2.2 Examining the Question Papers Reading Skill in Terms of Content Validity (weightage)

She has mentioned the allocated marks for reading skill given in the tests and compared with specification grid in this table.

Table 7
Content Weighting of Reading Skill

Weightings of the course contents			Weighting of the tests											
			2063			2064			2065			2066		
	Full weighting		No. of question	Weightage	Total weightage	No. of question	Weightage	Total weightage	No. of question	Weightage	Total weightage	No. of question	Weightage	Total weightage
Reading skill	30 marks	) Textbook based short-answer questions	5	2	10	5	2	10	5	2	10	5	2	10
		Total	5		10	5		10	5		10	5		10
		Seen passage												
		) Short answer questions	4	2	8	4	2	8	4	1.5	6	4	1.5	6
		) Multiple choice							4	0.5	2			
		) Vocabulary												
		) Matching item												
		) True/false item												
		) ordering										4	0.5	2
		Total	4		8	4		8	4		8	8		8
		Unseen passage												
		) Short-answer questions				3	1	3	4	1	4			
		) Fill in the gaps	4	1	4	4	0.5	2				4	1	4
		) Multiple choice												
		) Vocabulary												
		) Matching item	4	0.5	2	6	0.5	3	4	0.5	2	4	0.5	2
		) True/false item	4	0.5	2				4	0.5	2	4	0.5	2
		Total	12		8	13		8	12		8	12		8
		Vocabulary												
		) Use of words/ phrases in sensible sentences	6 (2-opt)	1	4	6 2-opt)	1	4	6 2-opt)	1	4	6 2-opt)	1	4
		Total	6		4	6		4	6		4	6		4

Note: (2-opt) = 2 optional choice

Table 7 shows that the full mark of reading skill is 30 marks out of full marks, according to the skillwise mark allocation given in the specification grid. Obviously, the researcher compared the weighting of the question papers with the weighting of the specification grid and found that the weightage of the reading skill matched. In which 10 marks for text based short-answer questions, 8/8 marks for seen and unseen reading passage and other left 4

marks for vocabulary items. Out of 30 marks short answer questions have been asked with 10 marks weightage in all the four years, which was proportional weighting according to the grid.

From seen reading passage, only short-answer questions of 8 marks have been asked in the year 2063 and 2064. And short-answer questions (6 marks) as well as fill in the gaps (2 marks) and ordering items (2 marks) have been asked respectively at the time of 2065 and 2066 examination.

Similarly, in case of unseen reading passage, the questions have been asked with different weighting in different years. In the year of 2063 examination, matching items (2 marks), fill in the gaps (4 marks), true/false item (2 marks) have been asked with proportional weighting. But at the time of 2064 examination from unseen reading passage matching items (3 marks), fill in the gaps item (2 marks) and short answer questions (3 marks) have been asked. Similarly, short-answer questions (4 marks), true/false item (2 marks) and matching items (2 marks) have been asked with proportional weighting in 2065 examination. 2066 examination also carries proportional weighting in case of unseen reading passage in which true/false items (2 marks), fill in the gaps (4marks) and matching item (2 marks) have been asked.

As the grid suggests other remaining 4 marks was set vocabulary items in which 6 words/phrases have been asked to use in sensible sentences with two options in all these four years. That is also a proper mark allocation in reading skill.

Thus, it is found that whether the questions are short-answer questions, fill in the gaps, true/false item or whatever else, asked in these four years' examinations reflect proportional weighting of the reading skill. That is to say, the test designer has kept in their mind the proper mark distribution of reading skill while designing the test in those four years. Therefore, it can be concluded that the reading skill represents really exact weighting of the course during those four years' examinations.

3.2.3 Examining the Question Papers Writing Skill in Terms of Content Validity (weightage)

By the help of this following table she has compared the mark weighting given in the tests and compared with specification grid.

Table 8
Content Weighting of Writing Skill

Weightings of the course contents			2063			2064			2065			2066		
Writing skill	Full weighting	**Punctuation**	No. of question	Weightage	Total weightage	No. of question	Weightage	Total weightage	No. of question	Weightage	Total weightage	No. of question	Weightage	Total weightage
		⌉ Full stop	2	0.5	1				2	0.5	1	2	0.5	1
		⌉ Question mark												
		⌉ Commas	2	0.5	1	2	0.5	1	3	0.5	1.5	1	0.5	0.5
		⌉ Exclamation												
		⌉ Spelling	1	0.5	0.5	2	0.5	1				1	0.5	0.5
		⌉ Inverted commas	1	0.5	0.5	2	0.5	1	1	0.5	0.5	1	0.5	0.5
		⌉ Capitalization	2	0.5	1	2	0.5	1	2	0.5	1	2	0.5	1
		⌉ Apostrophe										1	0.5	0.5
		Total	**8**		**4**	**8**		**4**	**8**		**4**	**8**		**4**
	30 marks	**Guided writing**												
		⌉ Story from picture or clues ⌉ Completing a dialogue ⌉ Note making ⌉ Description of movement	1	6	6	1	6	6	1	6	6	1	6	6
		Total	**1**		**6**	**1**		**6**	**1**		**6**	**1**		**6**
		Free writing												
		⌉ Application ⌉ Greeting cards ⌉ Paragraph writing ⌉ Reporting facts & events ⌉ Letter writing	1	10	10	1	10	10	1	10	10	1	10	10
		⌉ Thank-you note ⌉ Essay writing ⌉ Writing instructional leaflets	1	10	10	1	10	10	1	10	10	1	10	10
		Total	2		20	2		20	2		20	2		20

The above table indicates that writing skill covers 30 percent marks out of full marks as the skillwise mark allocation given in the specification grid. All the questions are required to give answer in written form in examination but all of them cannot represent the writing skill. That is why, according to the grid, writing skill covers 30 marks weightage in which free writing covers 20 marks (2 long-questions) and guided writing bears 6 marks (1 short-question). The remaining 4 mark is allocated for punctuation marks.

In case of punctuation marks, every time 8 punctuation marks questions have been asked with 4 marks weightage (each with 0.5 weightage) during those four years' examinations, which was the correct wieighting in writing skill according to the grid. Similarly, in those four years' examinations, guided writing for e.g. 'story from given outlines' has been asked with 6 marks weightage which was also exact weighting in writing skill. As far as the free writing is concerned, 2 long questions have been asked with 10 marks weightage for each question which was also another correct weighting in writing skill.

Thus, she has found that there is no need to raise a question of mark weighting in case of writing skill during that four years' examinations. It is found that the test setters have strictly followed the specification grid while designing the test.

On the whole, the compulsory English grade 8 question papers have proportional content validity (weighting) in terms of grammar, reading skill and writing skill.

3.3 Comparison between Content Validity in terms of Coverage and Weighting

She has found that the compulsory English grade eight test has high content validity in terms of content coverage in which 70 percent course contents were covered in tests. It is believed that if more than 60 percent course

contents are covered by test contents it is supposed to have high content validity. It means, the test items have tested what they ought to test.

On the other hand, when the researcher has analyzed skillwise mark distribution, there is no need to raise a question about weighting of the test items. It is found that the norms and marks allocation followed distribution of marks given in the question papers. Therefore, compulsory English grade 8 question papers have exact content validity in terms of weighting.

The above description shows that a valid test from one perspective may not necessarily be valid from another perspective. Hence, the test papers having exact content validity in terms of content weighting, has only 70 percent content validity in terms of content coverage. So, tests having high content coverage may have low content weighting and vice versa.

CHAPTER - FOUR

FINDINGS AND RECOMMENDATIONS

This is the final chapter of this research, which deals with the findings and recommendations made by the researcher after analyzing the data. This chapter has been divided into two parts where the first part deals with the findings made by the researcher after analyzing the data and the second part deals with the recommendations made on the basis of the collected data and analyzed data and the findings made in the first part of this chapter.

4.1   Findings

After analyzing the test papers from different angles, the researcher has listed the following findings:

4.1.1 General Findings

It has found that the compulsory English grade 8 tests (2063-2066) have high content validity in terms of content coverage. It is because out of 50 language items 35 language items were represented or became the content of the test. It is 70 in percentage. Other remaining 30% course content is not given in the question papers.

In case of content weightage it has found that the compulsory English grade 8 tests (2063 to 2066) have exact content weightage. It is 100 in percentage.

4.1.2 Specific Findings

1   In Terms of Content Representativeness/Coverage

In grammar, there are altogether 10 course content language items, out of which 9 language items are represented and 1 language item is not covered

during four years examinations. It means, the coverage of course content is 90 percent. Therefore, grammatical items have high content validity.

About reading skill, there are altogether 20 reading topics, out of which 16 of them are represented and 4 topics are left during four years tests. It means, the coverage of course contents is 80 percent. It is less than grammatical items but it has also good content validity.

In writing skill, there are altogether 20 course content language items, out of which 10 items are represented and other 10 are ignored during four years examinations. It means, the coverage of course content is 50 percent. This is the skill having least content coverage but still it has average content validity.

2   In terms of Content Weighting

According to the mark allocation given in the specification grid, grammatical item carries 10 marks in which each question carries 1 mark. It is found that the question setter has followed the grid while designing the test during four years examination. Therefore, this grammatical item has exact content validity in terms of weighting.

According to the skillwise allocation of mark, as shown in the specification grid, reading skill carries 30 marks, out of 30 marks, 10 marks is for short-answer questions, 8/8 marks for seen and unseen reading passage and remaining 4 marks for vocabulary item. It is found that the test designer has followed the mark distribution in question papers during four years. That is why, reading skill has correct content validity in terms of weighting.

According to the skillwise mark allocation given in the specification grid, writing skill carries 30 marks. Out of that, 20 marks is for free writing, 6 marks for guided writing and remaining 4 marks for punctuation marks. In the question papers of those four years, the distribution of marks is proportional with the specification grid. Therefore, it is concluded that writing skill has exact content validity in terms of weighting.

3   In terms of content coverage, the researcher has found that no language items were represented in all the four years' tests.

4   The researcher has found that some of the reading topics are repeated time and again. For e.g. from 'the story of Icarus' (what warning did Deaduls give to Icarus-three times and why did Icarus fall on to the sea-one time) questions are asked all the four years examinations. Such repetition of same test item exerts harmful washback effect, encourage rote learning and reduce content validity (coverage).

5   The researcher has also found that the same type of test item is asked during four years. For e.g. 'story from given outlines' from that test item 2064 and 2065 questions are repeated.

6   The researcher has also found that, in case of free writing some test items are repeated time and again, for e.g. (essay writing 'importance of woman education'-twice), ('wonders of science'-twice), ('my favourite place'-twice) etc. Such repetition of the same test item time and again encourages rote learning and reduces content coverage.

7   Asking questions from the same language test item lack content validity in terms of coverage. For e.g. only essay writing and letter writing are asked and other writing skills are ignored. That is found during 2063-2066 examinations. Due to that reason, students prepare what is asked and do not prepare other items what is not asked.

8   According to the weightage given in the specification grid, this course has tested what has been supposed to measure because it strictly followed the procedure while designing the test and it is one of the beneficial tasks for the examinees.

9   The researcher has found that in case of grammatical items, some test items are focused. For e.g. 'reported speech' and 'voice' from which 2

questions have been asked in each year examinations besides 2064 and other items are neglected.

4.2   Recommendations

The researcher has made some recommendations on the basis of the analyzed data, the findings of the research work listed in the previous sub-chapter and the researcher's own intuitions which are listed below.

1   The researcher has found that the content representativeness of compulsory English grade eight tests during four years is 70 in percentage. Some items are left out. So, the test designers should design the tests ensuring 100 percent content representation as far as possible.

2   It has been found that some topics and language items are neglected and some others are over emphasized while designing the question papers. Therefore, the question setters should give equal emphasis on all the topics and language items.

3   Many test items are repeated in four question papers. Such repetitions encourage guessing and reduce content validity. Therefore, repetition of the same items time and again should be avoided.

4   There should be fixed criteria on how many and what types of test items to be asked from each language item.

5   If possible, in case of guided writing, completing a dialogue, note making and pictures to make a story should be given in place of story from given outlines. Because it helps to make variation in question papers and covers more course contents.

6   If feasible, tests of same language items only should not be made. For e.g. essay writing and letter writing. The test designer should keep in their mind to cover the whole writing skill contents.

7   The test has to be valid in terms of both content coverage and content weighting in order to match with course objectives.

8   The test designers should be trained and highly experienced. They should have the knowledge of language teaching, language learning and language testing.

9   The test designers have to do a minute study on the related course objectives, course contents and specification grid/evaluation scheme before constructing the question papers.

10 Finally, in maintaining the content validity, other types of validity should not be neglected.

References

Bachman, L. F. (1990). *Fundamental considerations in language testing.* Oxford: OUP

Barber, C. (1999). *The English language: A historical introduction.* Cambridge: Cambridge University Press.

Batala, K. B. (2004). *Validation of the SLC English question paper.* An unpublished M. Ed. Thesis, Tribuvan University, Kathmandu, Nepal.

Bhattarai, J. (2005). *The content validity of compulsory English textbook for grade Eight.* An unpublished M. Ed. Thesis, Tribhuvan University, Kathmandu, Nepal.

Burns, A. (2005). *Teaching English from a global perspective.* USA: Crofton Maryland.

Gurung, R. C. (2008). *Content validity of semantics & pragmatics Examination at M. Ed. Level.* An unpublished M. Ed. Thesis, Tribhuvan University, Kathmandu, Nepal.

Harrison, A. (1991). *A language testing handbook.* Longman: ELTS.

Heaton, J. B. (1975). *Writing English language test.* London: Longman

Hughes, A. (2003). *Testing for language teachers.* Cambridge: Cambridge University Press.

Karki, M. B. (2009). *Content validity of general linguistics course of grade XI.* An unpublished M. Ed. Thesis, Tribhuvan University, Kathmandu, Nepal.

Khaniya, T. R. (2005). *Examination for enhanced learning. Lalitpur:* Millennium Publication.

Kubiszyn, T. & Gary Borich (2003). *Educational testing and measurement.* India: New Delhi.

Lado, R. (1965). *Language testing.* London: Longman.

Linn, R . L. & M. David Miller, (2008). *Measurement and assessment in Teaching.* India: New Delhi.

Luitel, Dr. B. (2009). *Research in language education.* Kathmandu: Subhakamana Publication.

Nepal, P. (2006). *Content validity of examination: A case of fundamental of language and linguistics at B. Ed. level.* An unpublished M. Ed. Thesis, Tribuvan University, Kathmandu, Nepal.

Ojha, D. R. (2005). *Content validity of ELT theories and methods exam at B. Ed. Level.* An unpublished M. Ed. Thesis, Tribhuvan University, Kathmandu, Nepal.

Patel, Dr. R. N. (2004). *Educational evaluation theory and practice.* New Delhi: India.

Poudel, D. (2010). The origin of language, *Journal of perspectives*,1.1-75

Richards, et al. (1985). *Longman dictionary of applied linguistics.* London: Longman.

Sapir, E. (1971). *Language.* New York: Harcourt.

Sharma, G. (1999). *Educational history of Nepal.* Kathmandu: Lumbini Pustak Bhandar.

Sharma, L. (2010). *An appraisal of grade five English textbook in terms of content validity.* An unpublished M. Ed. Thesis, Tribhuvan University, Kathmandu, Nepal.

Subedi, R. (2006) *Content validity of B. Ed. questions*: *A case of English sounds and structures.* An unpublished M. Ed. Thesis, Tribhuvan University, Kathmandu, Nepal.

Thapa, B. (2010). *A study on testing reliability interms of Parallel Forms method.* An unpublished M. Ed. Thesis, Tribhuvan University, Kathmandu, Nepal.

Timilsina, T. P. (2007). *Content validity of phonetics and phonology exam at M. Ed. level.* An unpublished M. Ed. Thesis, Tribhuvan University, Kathmandu, Nepal.

Yadav, Y. P. (2000*). Linguistics: A basic course*. Kirtipur: New Hira Books.

	Reading	New language	Speaking	Listening	writing
	How it all began: notices conversation	Revision	Questions, short answers. Tag questions Describing and narrating	Descriptions of movements. Filling a form. Correct sentences	Competition, enter, prize, travel, favourite, report, prepositions of movement
	The sta	Repo	A di	A sto	Balloon, film

	rt of the adventure a letter	rted instructions, commands, nouns in apposition	scussion: picking out words	ry from pictures.	crew, studio, honour excite, mistake, congratulate, equipment, expedition, nationalities.
	The sto	'who'	Teleph	Creative	Boast, invent,

		ry of Icarus	clauses Descriptions using 'like,'	one conversation: checking details	descriptions A poem	centre, melt, sew, feathers, weep, accident various adjectives
		Animals in Chitwan: conversation and infor	'which', 'that', claus	Crocodiles: information	Giving information. A story from pictures	Antler, spot, spotted, rescue, breed dangerous, cliff, unconscious

		mation	es		
		Marco Polo and Geng his Khan: biograbpy. map	Reported statements	Dialogue to report	Reporting facts.	Travel, desert, sailor, conquer, murder, control, reach, serve explore, surprise
		In Janakpur: diary and narrative conversation.	'where' clauses Reported questi	Song Dialogues. Reporting. Giving short answer	Different dialogues to match to situations. Song to complete.	Describing pictures including dialogue.

		ons	. Discussing pictures .		
		Ellen's mountain : a longer story	Impersonal verb . Past perfect	Describing sensations	Identifying sounds Dialogues	A dialogue. A cloze
		Up to Gokyo: story infor	Time : when , du	Discussing. Describ	Volcanic eruption: factual	Cloze: An instructional leaflet

		mation, instructions	ration 'while', 'when ;', 'until', clauses: past continuous	ing time. Asking about time	report	
		Shyamtal	Pres. p	Asking g	Ordering ng	A tourist guide to

		ks about Australia information, mapwork	articiples as nouns. As soon as Until.	and answering (information) Making Suggestions	dialogue	Nepal
		Over Everst: a newspaper	Know, Remember Sta	Using know and remember. G	A radio programmer	Writing newspaper report. Clauses.

		ting purpose	iving reasons Discussion. Sports day in English		
	Explorers and strange animals: exciting history, facts	'too' and 'enough' (all for	A quiz Telephoning asking for	Kangaroos: information	Descriptive paragraph: the rhinoceros.

		ms)		repeats Dialogues with too/enough		
		Thak Khola: letter, personal memories famous poem	Habitual past	Describing former and present habits Sharing m	Vi: a personal story	First person paragraph. The/Now paragraphs. Paragraph about parents

56

			emories Poem.		
	Monkey Tricks in Muktinath: realistic adventure story	Reflexive pronouns. Each other	Using reflexive pronouns Quizzing Discussing	Radio quiz	Making notes A semi-formal letter. Correcting
	Annapurna: a true account	Conditions: (r	Dialogues with condi	Completing dialogue	Diary tense practice

		eal) Present perfect continuous Past perfect contin	tions Using continuous tenses Arranging, Persuading Giving conditions		

		uous			
	Treasures from the Past: guide book for Lumbini	Imaginary conditions Predicting future	Imagining Predicting Suggesting Ordering	China: radio programme	Descriptive paragraph>terracotta
	The wooden horse: famous us	Simple past p	Asking politely and d	A dialogue: imaginary ide	Telling story Passive sentences. Newspaper articles

	story	assive (all forms)	replying Using passive Telling story	as	
	Jhimruk Power Station: description and conversation	Present perfect passive. A g	Conversation and discussion Using perfect	Guided tour of power station	Formal letter

		uessing game	passive. A guessing game		
	Encyclopedia: how to use and interesting entries	Present passive Flow charts Getting th	Complaining Requesting Apologizing Using Getting do	Dialogue: complaining politely	Flow chart, Vocabulary Paragraph in passive

			ings done	ne		
		In Jajarkot: real life personal stories	Present continuous and future passive. Makin	Using future passive	A deal girl: parents talking	Newspaper report (use of passive) Description of actions

			g people do things. Make work		
		The end of the adventure: dialogue and short article	Revision	Leave taking Telling a story Poem and	Leave taking Beatle song	Revision exercises

			di sc us si o n		

Appendix II

Specification Grid for Grade 8

Gramm ar & Str uct ure 10 ma rks	Listening Skill 15 mark s	Speaking 15 marks	Readin g Ski ll 30 ma rks	Writing Skill 30 marks

Tense, preposition, articles, relative clauses(who, which, that etc.), active/passive, reflexive pronouns, reported speech (instructions, co	The teacher reads out the following listening materials or plays the audio cassette once or twice Materials: **Dialogues** (consisting polite and formal/informal expressions) **Telephone Conversation** **Listening**	−Interview −Describing object/person /place/event −Telephone Conversation (Dramatization) −Retelling story −Discussions **Types of Questions** −Oral questions −Asking to read a passage −Simulated telephone conversations −Group discussion on topic to interest **Oral Activities** −Interview 5 marks −Describe object/person/place or	−Comprehension Questions from textbook −Seen Reading Passage −Unseen Reading Passage **Types of Questions** −Short answer questions −Multiple cho	−**Punctuation** (full stop, question mark, exclamation, inverted , comma capitalization) −**Guided writing** (story from pictures or clues, completing a dialogue, note-making etc) −**Free writing** (application, greeting cards, paragraph, writing, reporting facts & events, letter writing, thank you note, writing instructional leaflets etc.) −**Types of Questions**

mmands, statements), **Types of Questions:** Fill up the gaps Multiple choice Transformations(negation, passive, interrogation) Indirect	**Comprehension** **Types of Questions:** True/False or Yes/No Questions Tick the correct Listen and follow Short Questions Any three types of questions given above should be asked with 5 items carrying 1 marks	retelling story 5 marks −Situational dialogue or simulated telephone conversation 5 marks	ice −Fill in the gaps −Vocabulary −Matching **Number of questions:** Shorts answer questions: 1.textbook based 5*2=10 2.Seen passage 4*2=8 3.Unseen passage 4*2	−Passage to be punctuated=4 marks −Paragraph/story/reports writing on given clues=6 marks −Questions for free writing 2*10=20 marks

spe ech	each.		=8 4.Vocab	
Total nu mb er of que stio n=1 0 and eac h que stio ns carr ies 1 mar ks			ular y 4*1 =4	