

# **CHAPTER I**

## **INTRODUCTION**

This is the general introduction of the whole proposal. Under this heading, there are many sub-headings as described in the following paragraphs in detail.

### **1.1 General Background**

Language is a special gift for human beings since it is commonly defined as the voluntary vocal system of human communication. Furthermore, it is the highly developed and frequently means of human communication through which we express our feelings, thoughts, wants, desires and so on. According to Hockett (1958 as quoted in Paneru 2000), "Man does not live by bread alone, his other necessity is communication." Similarly, Hornby (2005) defines language as, "The use by humans of system of sounds and words to communicate."

In short, language is as human beings: it takes its birth, grows up and dies in course of time. English, Nepali and Sanskrit are the examples of fully developed, still growing up and dead languages respectively.

English is one of the international language in the world. It is a principal language for international communication and a gateway to the world. No language in the world is as popular as English. It is a lingua franca into the world. It is the most popular language used in international meetings and conferences. Not only this, but English is also serving as an important vehicle for the transmission of civilization and culture from the western world to the eastern world and vice-versa. Thus, English is growing as a global language because it has become the language of

business, internet, sports, civil aviation, medicine, science and technology.

English has become an appropriate international language in Nepal and a vital tool for any student to become successful in local, national and international communication. In the context of Nepal, the gravity of teaching English language was realized from the beginning, and it was taught as a compulsory subject up to the bachelor level. In addition, it is used as an access language, a library language and as a means of instruction and evaluation at the higher levels of education. Regarding all these things into consideration, English was given a high priority over other international languages in the world.

### **1.1.1 Language Testing**

Testing is a process of determining the quality or genuineness of something. Testing in a broad sense has always been an inherent part of teaching. Testing is as old as teaching. It is a process of scrutinizing how far learners have learned what the teacher wishes them to learn. In this process, the teacher usually makes queries during or after this teaching intending to understand whether or not the whole or part of his teaching was imparted and received by the learners effectively as well as to determine, if necessary, remedial measures. Davies (1968: 5) says that the good test is an obedient servant since it follows and apes the teaching. Khaniya says that testing is a joy killer of the testees so nobody wants to be tested. Anyway, test is any procedure for measuring ability, knowledge or performance. A good test may be primarily constructed having the qualities of a good test to reinforce learning and motivate the students. In addition, it is widely accepted that testing offers useful inputs to the teacher to be aware of the effect of his teaching, and also some insights on whether he should continue the way he teaches or change it in

order to make his teaching more effective. So, we can say that testing and teaching are interrelated and independent facets of an academic programme. Anyway, testing is a process of gathering information in order to make decision.

Carroll as quoted in Davis (1968: 46) states that a psychological or educational test is a procedure designed to elicit certain behaviour from which one can make inferences about certain characteristics of an individual. Thus, the fundamental use of testing in an education programme is to provide information for making decisions, that is, for evaluation.

Testing is an integral part of teaching and go side by side in the sense that teaching and learning activities remain incomplete in the absence of testing. Better understanding of testing techniques, no doubt, will lead the teacher to perform effectively in the classroom. Testing and teaching are thus regarded as an integral part of education. It is quite essential to say that whenever there is change in language teaching, there is a change in language testing and vice-versa. One is supplying information and the other is assessing it. Thus, one is supplementary to the other.

Testing plays an important role in teaching. Form the educational point of view, testing is an integral part of teaching. They have the relationship of head and tail. So, wherever head goes, tail also goes. Teaching and testing are the two sides of the same coin. Without testing, teaching is like the horse without rein. Testing is as important as teaching. So, it is said that whatever is taught should be tested. In this context, Heaton (1988: 5) says, "Testing and teaching are so closely interrelated that it is virtually impossible to work in either field without being constantly concerned with the others." Furthermore, teaching without

testing is like pointing without colour. With the help of testing, we can find the areas that need more attention and we can do the remedial works.

### **1.1.2 Qualities of a Good Test**

Tests are administered for some purposes. In order to serve the purposes for which tests are conducted, they must be of good quality. The quality of an exam is examined in light of the extent to which it serves the purposes for which it is administered. Putting this in another way, the quality of a test is examined in light of its usefulness. Therefore, while designing a test, the usefulness of the test, or the efficiency of the test must be considered. For example, a discrete point test can be good or useful for ascertaining grammatical competence, but not for communicative competence.

There are different views on what makes a test good or what are the qualities or requirements which are supposed to fall under a good test. Such as, Harrison (1991: iii) has said that the qualities of a good test are reliability, validity, practicability, and comparison and discrimination. Similarly, Khaniya (2005: 102) has said that in essence, validity, reliability and practicality constitute the quality of a good test. In this regard, Bachman and Palmer (1996 as quoted in Khaniya 2005: 101-102) argue that test usefulness involves reliability, construct validity, authenticity, interactiveness, impact and practicality. However, all these views share some common characteristics or qualities regarding the qualities of a good test. They are: validity, reliability and practicality (i.e. scorability, administrability and economy). Therefore, these characteristics should be taken into consideration while writing the tests otherwise it becomes just as it is impossible to play chess without knowing how a knight moves across the board. So, it is pointless to try to write tests without a basic understanding of the principles behind them as

Harrison said (1991: 10). Since the researcher's main concern is validity (i.e. content validity), validity is dealt with greater emphasis.

### **i. Reliability**

Reliability is one of the essential qualities of a test which refers to the consistency of score or performance of the same or similar test administered within a reasonable time. No matter whoever the scorer is; the score must be the same and if this is the case there is the reliability of the test. Reliability can be dealt with at two levels: test and retest of students and marking and remarking of the examiners. It is a statistical concept. The degree of consistency of measurement is determined by carrying out some statistical analysis. It is reported in terms of correlation coefficient. The higher the correlation, the higher the reliability of the test. In this regard Hughes (1995: 29) has said that the more similar the scores would have been, the more reliable the test is said to be. Harrison (1991: 10) has said that the reliability of a test is its consistency. In short, it is defined as consistency of measurement.

There are different factors that contribute to the reliability of the test. For example, homogeneity of items, discriminating power, variability of group-students, sufficient test taking time, freedom of choice, unambiguous item, objectivity in scoring, length of the test, methods of estimating reliability, etc.

### **ii. Practicality**

Practicality is the inclusive term for scorability, administrability and economy. It is different from other qualities of a test. Absence of this quality in a test will lead the test to be of no use. It is the next most important quality of a good test in the absence of which even a valid and

reliable test can be of no use. So, it is not a good test unless it is applicable. The test should be fairly straight forward to administer.

Generally, practicality involves the cost, ease of administration and scoring. It should be practically fit in the intended situation in all these respects. It means a test to be practically fit, it must be scorable (i.e. easy to score in terms of points), administrability (i.e. easy to administer in terms of length of the test, arrangement of test, necessity of equipments, etc) and economic (i.e. test should not be costly in terms of money, human resources and time).

This aspect of test efficiency is important because failing to achieve the practicalities would be difficult to convince the people who would actually implement the exam. In order to achieve the practicality of the exam, the designer must keep a close look on the situation which the exam is supposed to fit into.

### **iii. Validity**

Validity is the most important but complex quality of a good test. The term 'validity' has been defined variously by various scholars. Validity can be applied in different fields or areas in a different sense but here the term has been used in a specific sense of testing. Validity is a relative term and can be defined as the degree of accuracy of a test. Since validity is a relative term, there is no single perfect definition of it. However, some of the important definitions of validity are as follows:

Regarding validity, Davies et al (1999 as quoted in Khaniya 2005: 103) defines it as, "A measure is valid if it does what it is intended to do . . . ." It means, 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 administered.

Heaton (1988: 159) says, "The validity of a test is the extent to which it measures what it is supposed to measure and nothing else."

Lado (1961: 30) explains the concept of validity as asking a question: "Does the test measure what it is intended to measure? If it does, it is a valid test. Validity is not general but specific. If a test of pronunciation measures pronunciation and nothing else, it is a valid test of pronunciation; it would not be a valid test of grammar or vocabulary because it does not test grammar or vocabulary."

In the same way, Hughes (1995: 22) defines validity as "a test is said to be valid if it measures accurately what it is intended to measure." Harrison (1991: 11) also stands at right behind of Hughes in defining the term 'validity' as, "the validity of a test is the extent to which the test measures what is intended to measure."

Likewise, Khaniya (2005: 103) states validity as, "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 administered."

Lastly, American psychological Association (1985) explains validity as, "the most important quality of test interpretation or use is validity, or the extent to which the inferences or decisions we make on the basis of test scores are meaningful, appropriate and useful." It adds that if test scores are affected by abilities other than the one we want to measure, they will not be meaningful indicators of that particular ability. If, for example, we as students to listen to a lecture and then to write a short essay based on that lecture, the essays they write will be affected by both their writing ability and their ability to comprehend the lecture. Ratings of their essays, therefore, might not be valid measures of their writings ability.

In examining validity, we must also be concerned with the appropriateness and usefulness of the test score for a given purpose. A score derived from a test developed to measure the language abilities of monolingual elementary school children, for example, might not be appropriate for determining the second language proficiency of bilingual children of the same ages and grade levels. Similarly, scores from a test designed to provide information about an individual's vocabulary knowledge might not be particularly useful for placing students in a writing programme.

While reliability is a quality of test scores themselves, validity is a quality of test interpretation and use. Reliability and validity are both essential to the use of tests. Reliability is a quality of test scores, while validity is a quality of the interpretation or uses that are made of test scores. Unless a test score is reliable, it can not be valid.

Though none of the definitions mentioned above are perfect in themselves, since validity is language and author specific (relative concept). However, on the basis of these above mentioned definitions, it can be concluded that validity is one of the most important qualities of a good 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 administered. It is concerned with the reflection of the objectives of its administration. Anyway, validity refers to the extent to which a test actually measures what it is intended to measure and nothing else. Moreover, every valid test must be meaningful, appropriate and useful. Reliability is necessary ingredient of validity but it is not sufficient to ensure validity. Unless the test scores measure what the test user intends to measure the scores will not be valid.



Validity in language tests depends on the linguistic content of the test and on the situation or technique used to test this content. It means if a test tests or measures what is actually supposed to be measured, then it becomes a valid test and the situation of such test becomes its validity.

### **Types of Validity**

As there is not a fixed or clear cut definition of validity, there is not a single and universal classification of it too. However, the following are the main recognized types of validity:

- i. Content validity
- ii. Construct validity
- iii. Criterion related validity
  - a. Concurrent validity
  - b. Predictive validity
- iv. Face validity
- v. Washback validity

Content and construct validity are said to be conceptual, and concurrent and predictive validity are said to be statistical validity. Face validity is said to be pseudo validity, and washback validity is said to be an emerging necessary element for a good test.

### **Construct Validity**

Construct validity is the extent to which a test represents an underlying theory of language learning. Putting it in simple way, a test can be said to have construct validity if it measures just the ability which it is expected to measure and nothing else. Here, the ability refers to theoretical construct or the theoretical explanation or proposition of a trait.

Regarding construct validity, Heaton (1988: 161) states, "If a test has construct validity, 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." Similarly, Brown (1976: 128) (Citing from APA 1974: 29) writes: "Construct validity is implied when one evaluates a test or other set of operations in light of the specified construct."

Actually, construct validity is a "comprehensive concept" which includes content and criterion related validities for Anastasi and content and predictive types of validity for Davies.

Thus, a test is said to have construct validity if it measures just the ability which it is expected to measure and nothing else. For example, if we are testing reading ability of an individual, we just test reading ability and nothing else. For this, we may ask him/her to read the text or to go through the text but not to write the answer of a question. It is because if we ask the testee to write long answer of a question in the name of testing reading, in fact, it is testing writing; and thus, this test doesn't have construct validity since it demands writing in the name of testing reading ability. Here, the examiner has confusion that whether the testee has the problem of reading or writing ability. But I do not mean that we should not administer question for reading test, we can administer or use just the objective questions (i.e. multiple choice questions). Moreover, construct validity is the inclusive term which includes almost all the validities of a good test, therefore, it is said that if a test has construct validity, it is automatically said to have all the other validities as well.

## **Criterion-related Validity**

Criterion-related validity is also called empirical validity or statistical 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. Anyway, the validity of a test established by comparing with a set criterion measure is called criterion-related validity.

Criterion-related validity can be established by giving the students an established test with similar nature which has proved to be valid. The test can be administered at the same time or in a short gap ensuring that no additional learning opportunity is given. Criterion-related validity is important where the test is to be used to decide entries for public examination. Criterion-related validity is of two types: concurrent validity and predictive validity. The main difference between the two types of validation procedure is time interval.

By name, 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. The established test can take a form of either a well known test or the rating of the teachers. It is used to decide entries of the new tests for public examination.

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 criterion. One of the difficulties in dealing with predictive validity is to find a satisfactory criterion with which the exam results are to be correlated. To establish the predictive validity of an exam, some standard measure of performance must be pre-specified. It is this pre-specified performance which serves as a criterion. It is used to predict the future performance of the candidates. The TOEFL and IELTS – are being used as selection devices on the assumption that there is a prominent role of language in academic success.

### **Face Validity**

Face validity is what it appears superficially to measure. A test is said to have face validity if it looks as if it measures what it is supposed to measure. Face validity is hardly a scientific concept, yet it is very important. Moreover, face validity is not a real test. It is just a matter of acceptability. The question is: unless the learners genuinely accept a test as a test, how can their performance be genuine? In the testing literature, face validity is often considered as fake or pseudo-validity. However, it is believed that if the examinee does not consider an exam a valid one, the information collected from it may not be genuine. Anyway, if a test item looks right to other testers, teachers, moderators, and testees, it can be described as having at least face validity. It is, therefore, often useful to show a test to colleagues and friends.

Thus, if a test is accepted as a real test by other testers, teachers, moderators and testees, it is said to have face validity or if a test looks as if it measures what it is supposed to measure, then it is said to have face validity. Though, in testing literature, face validity is often considered as pseudo-validity, it is very important because unless the learners genuinely

accept a test as a real test, their performance can not be genuine. So, to collect genuine information any test should be taken as a real test by the testee. If a test lacks face validity in spite of having all the other validities, it is not considered as a good test. That is why, face validity is a pre-requisite for other validities.

### **Washback Validity**

The effect of testing on teaching and learning is known as washback. The washback or educational effect of the exam may be either positive or negative and the positive washback of testing in the classroom or teaching learning process is termed as washback validity which is the contribution of Khaniya (1990) in his Ph.D. work. He has concluded that it is an inherent quality of an exam.

If a test has positive effect on testees then it is said to have washback validity. The difference between washback validity and content validity is simply one of the direction - content validation proceeds from the classroom to the test and washback validation proceeds from the test to the classroom. But in both cases, the aim is the same-to ensure that test content and classroom behaviour are related.

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

Morrow (1986: 6 as quoted in Khaniya 2005: 113) argues that one of the principal responsibilities of testing or examining boards and institutions, which operate in the public domain is to provide, "package" with a powerful and positive washback effect into the classroom.

### **Content Validity**

Content validity of a test is the extent to which the test actually measures what it is intended to measure. In case of a final achievement test, it is said to have content validity if its test items constitute a representative sample of the tasks as can be seen in the course objectives. It is connected to the representation of the objectives of the syllabus and the universe of subject matter in a test. It also includes the relevance of the test content in terms of the related course content which is often referred to as content relevance. The more a test reflects these aspect, the more it is said to have content validity.

There is not a fixed clear cut definition of content validity. It has been defined variously by various schoalrs. Some of the vital definitions are as follows:

Brown (1976: 122-123 as quoted in Khaniya 2005: 103) defines content validity as, "whether the items composing the test do, in fact, constitute a representative sample of the content domain of concern."

Hughes (1995: 22) defines content validity as, "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." He further explains that in order to judge whether or not a test has content validity, we need a specification of the skills, structures, etc that it is meant to cover since a comparison of test specification and test content is the basis for judgements as to content validity. Ideally,

these judgements should be made by people who are familiar with language teaching and testing but who are not directly concerned with the production of the test in question.

He also shows the importance of content validity that the greater a test's content validity, the more likely it is to be an accurate measure of what it is supposed to measure . . . . Too often the content of tests is determined by what is easy to test rather than what is important to test. The best safeguard against this is to write full test specifications and to ensure that the test content is a fair reflection of these. This definition emphasizes that content should be in accordance with a specification chart and relevant items should be included in the test items.

Heaton (1988: 159) focuses on the reflection of course objectives and the representation of the course content in a test for the maintenance of content validity and says, "This kind of validity depends on a careful analysis of the language being tested and of the particular course objective. 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." He further explains that when embarking on the construction of a test, the test writer should first draw up a table of test specification, describing in very clear and precise terms the particular language skills and areas to be included in the test and the test should achieve content validity and reflect the component skills and areas which the test writer wishes to include in the assessment.

Similarly, Messick (1975: 961 as quoted in Bachman 1998: 247) states content validity as, "content validity is like the barker outside a circus tent touting two bowing aerialists, a waving clown, and a poster of a lady riding a unicorn as a sample of the show you will see inside. It is not just that the sample is not representative of the variety of individuals

and animals inside or even an accurate portrayal of them, it is that you do not see any performances."

Harrison (1991: 11) argues "Content validity is concerned with what goes into the test. The content of the test should be decided by considering the purposes of the assessment and then drawn up as a list known as a content specification."

Likewise, Anastasi (1982: 131 as quoted in Weir 1990: 25) defines content validity as: "essentially the systematic examination of the test content to determine whether it covers a representative sample of the behaviour domain to be measured."

On the basis of the different views of the authors mentioned above, it can be concluded that content validity is not absolute but a relative feature of a test. Content validity is very essential quality of any good test because of the fact that the greater a test's content validity, the more likely it is to be an accurate measure of what it is supposed to measure.

In short and in fact, a test is said to have content validity if its content constitutes a representative sample of the skills, structures, etc with which it is meant to be concerned and such test tasks included in an exam are relevant to the language activities that are expected to be exercised under the given course. It is the broadest type of validity among all since it covers almost all the criterion that are to be tested. For example, it is concerned with the reflection of course objectives, content coverage, content relevance, and balance of weightage for the areas or units of the course. Thus, the researcher has selected the area to examine the content validity of "Grammar: Theory and Practice Test (Eng. Ed. 512)" because in the absence of content validity, any test becomes useless.



## **Guidelines for Establishing Content Validity:**

As there is not a single fixed definition of content validity, there are not hard and fast guidelines to examine content validity of a test so some of the useful guidelines to measure content validity are as follows:

Anastasi (1982: 132) provides the following useful guidelines for establishing content validity:

1. The behaviour 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 was prepared.
3. Content validity depends on the relevance of the individual's test responses to the behaviour area under consideration rather than on the apparent relevance of item content.

What emerges on the basis of these guidelines is the concern of content validity with the priori stage of exam construction. It means what we should be, to make an exam educationally beneficial, is to make the exam a mirror of the course objectives and behaviour domain in order to make the people understand that what it is expected by them.

Bachman (1998: 244) has provided two guidelines with the help of which we can examine the actual content validity of a test. They are: *content coverage* and *content relevance*.

Harrison (1991: 11) argues that the content of the test should be decided by considering the purposes of the assessment and then drawn up as a list known as a content specification.

Similarly, Hughes (1995: 22) puts his view that in order to judge whether or not a test has content validity, we need a specification of the skills, structures, etc that it is meant to cover . . . . A comparison of test *specification* and *test content* is the basis for judgements as to content validity. Ideally these judgements should be made by people who are familiar with language teaching and testing, but who are not directly concerned with the production of the test in question.

In the same way, Heaton (1988: 160) forwards his view that content validity depends on a careful analysis of language being tested and of the 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.

By observing the above mentioned different views and guidelines by different scholars to measure content validity for the English test, the researcher has selected Bachman's (1998: 244) guidelines as the more common, useful and logical guidelines to measure the content validity of "Grammar: Theory and Practice Test (Eng. Ed. 512)" by analyzing the test papers which were administered in 2057 to 2063.

As we mentioned above, Bachman (1998: 244) has provided two guidelines with the help of which we can examine the actual content validity of a test. They are: content coverage and content relevance.

**Content Coverage:** Comparison between course contents and test contents.

Content coverage is one of the two guidelines for examining test content without which there is no content validity. It is the extent to

which the tasks required in the test adequately represent the behavioural domain in question.

It is explicit that in a three or four hours 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 but the test items that compose an exam should constitute an appropriate representative sample of behavioural domain under consideration because the more the test coverage the more the test valid and vice-versa. Especially in an achievement test, the emphasis will be mainly on the coverage of the subject matter. It implies that the test designers must specify what skills the exam is designed to cover. Furthermore, 'to have good content validity a test must reflect not only the content of the course, but also demonstrate the balance of test items in terms of weighting given to each unit or area' (Khaniya 2005: 104). Harrison (1991: 11) claims that content validity is established by considering the purposes of the assessment and then drawing up a content list.

Nothing is, in fact, absolute and even a test can not be fully valid or fully invalid. Therefore, a test is judged as having content validity when the test items represent the course contents and the course objectives. The more test items are constructed, the more chances of having content validity. If all teaching units are covered in the test, it is supposed to have content validity.

### **Content Relevance**

Content relevance is another guideline for examining test content without which there raises a question in content validity. As we know, in addition to coverage, content validity of a test is examined also in relation to its relevance to the given course of study. In fact, content relevance means whether or not the test tasks included in an exam are relevant to

the language activities that are expected to be exercised under the given course. So, the test items should be congruent with the objectives that are supposed to be fulfilled. Moreover, for a test to be valid we expect the content and conditions to be relevant and then there will be no irrelevant problems. The investigation of content relevance requires 'the specification of the behavioural domain in question and the attendant specification of the task or test domain' (Messick 1980: 1017 as quoted in Bachman 1998: 244). While it is generally recognized that this involves the specification of the ability domain, what is often ignored is that examining content relevance also requires the specification of the test method facets that define the measurement procedures. In short, the way that test items are closely related with the course is content relevance. The more the test items are constructed, the more content coverage the test paper has. If most of the contents are covered in test paper, there is the establishment of content relevance.

Validity is essentially a matter of relevance. Is the test relevant to what it claims to measure? For example, if the course of study emphasizes communicative language ability, the test tasks must require the testee to demonstrate such abilities in order to achieve high content validity in terms of content relevance. In case of the course with linguistic competence in focus, a discrete point test will serve the purpose, not the communicative one.

Thus, a test can have high content validity if it meets the two requirements or guidelines for establishing content validity by designing such a test which covers the whole course and its test items are related with the course objectives. But the problem with language tests, of course, is that we seldom have a domain definition that clearly and unambiguously identifies the set of language use tasks from which

possible test tasks can be sampled, so that demonstrating either content relevance or content coverage is difficult. However, in a nutshell, the examination of content relevance and content coverage is a necessary part of the validation process.

### **1.1.3 Specification for Maintaining Content Validity of Test**

Test specification is a blue print which provides the test designers all the necessary information related with the test. So, it is said that development of purposeful, effective, successful and valid test begins with the test specification. It is also called 'Grid'. A test specification grid reflects the assessment and evaluation part of the curriculum. It not only shows the marks allocation to different areas and skills to be measured as mentioned in the curriculum but also shows all the relevant factors which are helpful while constructing test. For example, it shows the purpose of test, type of test, medium of test, content of test, time of test, type of test item, test method, weightage of test item, number of test item, type of questions, rubrick of test, level and age of the learners, text type, what kind of skills to be measured, scoring scheme, etc. Thus, the test developers get guidelines on all relevant aspects of consideration for the preparation of a valid test from specification grid.

In the context of Nepal, the new specification grid was introduced in 1999. Only then the SLC question papers of English from 2000 were developed in accordance with the new specification grid.

Since there is no provision of specification grid above SLC, we have no grid for "Grammar: Theory and Practice Test (Eng. Ed. 512)" too. The examination questions are developed on the basis of the concerned curriculum, syllabus, textbooks and sample questions.

## **Is the Test Specification Grid Necessary Even in Master's Degree?**

While I was describing test specification, a question came in my mind that 'Is specification necessary even in Master's Degree?', then I tried to collect some views from teachers to answer this question. Actually, I found two views: First view is in favour of specification and claimed that there must be specification to maintain content validity otherwise the irresponsible teachers and test designers carelessly design test and the test lacks content validity.

On the contrary, second view is against the specification and claimed that there is not necessary of a grid. Specification does not help to maintain content validity rather it confines the testees into the limited area. It may be helpful in school level and even in campus level to some extent but not in university level. In Master's Level, testees should be ready to tackle in every field so that they can make their mind broad and master.

However, I support the first view since in Nepal, the questions administered in the examinations of different levels are often criticized for not being proportional representative of the concerned courses of study and sometimes, they are blamed to have included some test items even beyond the specified course by the examinees. So, if there is specification grid, the test designers and testees can not go out of track which it is supposed to follow otherwise there will be the freedom to go any where they like while constructing test as in Nepal and most of the tests will have low content validity.

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

There are altogether eleven (11) papers, out of them five (5) papers carry 100 marks each and six (6) papers carry 50 marks each. These

eleven papers are divided into two groups. There are ten papers in group A and group B has elective one paper. In electric group there are four papers but only one paper will be offered as prescribed by the concerned subject committee in the campuses for teaching learning processes.

**Table 1: Specialization**

S.N.	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	Psycholinguistic and Sociolinguistics	100
4	Eng. Ed. 504	Research Methodology in Language Education	50
5	Eng. Ed. 505	Language Testing	50
6	Eng. Ed. 551	Semantics and Pragmatics	50
7	Eng. Ed. 552	Applied Linguistics	100
8	Eng. Ed. 589	Thesis/Eng. Ed. 574 Discourse Analysis	50
9	Eng. Ed. 590	English Language Teaching Methods and Practices	100
10	Eng. Ed. 599	ELT Practicum	50

**Table 2: Electives**

S.N.	Course title and No.	Subject	Marks
1	Eng. Ed. 571	English Literature and Its Pedagogy	50
2	Eng. Ed. 572	Advanced Reading and Writing	50
3	Eng. Ed. 573	Translation: Theory and Practice	50
4	Eng. Ed. 575	Second Language Acquisition	50

*Source: CDC, T.U.*

From the above data and description we know that at M.Ed. first year three specialization papers: Eng. Ed. 511 (Phonetics and

Phonology): Eng. Ed. 512 (Grammar: Theory and Practice) and Eng. Ed. 513 (Psycholinguistics and Sociolinguistics). In second year 7 papers from the specialization group: Eng. Ed. 504 (Research Methodology in Language Education); Eng. Ed. 505 (Language Testing); Eng. Ed. 551 (Semantics and Pragmatics) Eng. Ed. 552 (Applied Linguistics); Eng. Ed. 589 (Thesis or Eng. Ed. 574 Discourse Analysis); Eng. Ed. 590 (English Language Teaching Methods and Practices); Eng. Ed. 599 (ELT Practicum) are adopted. From the elective group one paper is adopted out of the four papers.

Generally in our testing system, most of the schools, colleges and universities follow the only annual examination system. Written, oral and practical examinations are administered according to the nature of courses. In M.Ed. 1<sup>st</sup> year in the subject "Grammar: Theory and Practice" only annual written test is used to be administered upto the year 2062 but from the year 2063, 20 marks is assigned for practical exam or internal assessment to measure the students' understanding. The full marks of this subject is 100 and 40 is its pass marks. Subjective (long and short questions) as well as objective questions are administered in the written test. Sixteen objective questions carrying 16 marks (i.e. one mark for each item) are asked in the objective test. Similarly, in the case of subjective questions eight short questions are asked carrying 6 marks for each question and three long questions carrying 12 marks for each one. It is hoped that 16 objective questions, 8 short questions and 3 long questions can represent the whole course as well as their weighting suits according to its contents. So the purpose of the study is to look into whether the test papers represent the whole course or not, and whether the content relevance of the given test is appropriate or not.



### **1.1.5 The Syllabus of Grammar: Theory and Practice (Eng. Ed. 512)**

Grammar: Theory and Practice (Eng. Ed. 512) course deals with various models of grammar and their practical applications to English Language Teaching [ELT]. It contains six units altogether. The first four units present different theoretical models of grammar. The fifth unit deals with pedagogic grammar and the final unit deals with grammatical practice specially geared to developing pedagogic skills. The course carries 100 full marks. Out of which, first unit carries 5 marks, second unit carries 15 marks, third unit carries 20 marks, fourth unit carries 15 marks, fifth unit carries 5 marks and sixth unit carries 40 marks.

This course is primarily designed for Master Level students in English Education. However, students studying language or language teaching in other departments or faculties, language teachers, researchers, grammarians and general readers as well can be benefited from it.

#### **Course Objectives:**

On completion of the course the students will be able to:

- develop insights into formal aspects of language with special emphasis on structural and generative models.
- develop insights into English grammar and its pedagogic applications and implications.

The course contents of this subject are divided into six units as follows:

Unit I: Traditional Grammar

Unit II: Structural Grammar

Unit III: Generative Grammar

Unit IV: Other Models of Grammar

Unit V: Pedagogic Grammar

Unit VI: Grammatical Practice

Note: The more detail of this course contents is given in the appendix-1.

Source: CDC, T.U.

## 1.2 Literature Review

Up to now different researches have been carried out in the field of language testing. Some of them are as follows:

Khaniya (1990 as quoted in 2005: 79) conducted a research on "Examination as Instrument for Educational Changes: Investigating the Washback Effect of Nepalese English Exams", and came to the conclusion that the SLC exam failed to assess the language skills that the SLC English course intended to develop in students . . . because of its textbook and previous exam paper oriented nature, it did not encourage students and teachers to focus on language skills entailed in the objectives. Finally, he has concluded the following things:

- i. Washback is an inherent quality of exam,
- ii. Ingredients of the exam determine whether the washback is negative or positive, and
- iii. Teaching for final exam is not only inevitable, but desirable as well.

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. He found out that the predictive validity of the SLC English examination was very low because the coefficient of correlation between the two sets of scores on the SLC and grade 11 English examinations is +0.1. On the other hand, the content validity of the SLC examination English question paper was high in terms of content representation, objectives and itemwise analysis of it.

Aryal (2005) carried out a research work on "Content Validity of Grade Twelve Compulsory English Exams 2061." He analyzed the question papers from different angles (i.e. by rubrics, length, difficulty

level, and content coverage) and found that the rubric of the all questions was simple and scientific except few items; the length of the question matches to the allotment of time for the examinees; the questions of English are moderate in difficulty level; the questions related to Heritage of Words lack content validity and unscientific.

Bhattarai (2005) carried out a research on "Content Validity of the English Textbook for Grade Eight." She found out that the contents are applicable and the book has content validity in terms of content coverage and applicability. Skills and functions are less valid but language structures have the high content validity.

Neupane (2005) carried out a research on "The Content Validity of English Textbook for Grade Seven." The main objective of this study was to find out the content validity of English Textbook prescribed for Grade Seven in relation to the specified objectives. He found out that the textbook of Grade Seven has less content validity.

Ojha (2006) carried out a research on "Content Validity of ELT Theories and Methods Exam at B.Ed. Level." The main objective of this research study was to examine the content validity of ELT Theories and Methods Exam at B.Ed. Level. For this purpose, the researcher has analyzed 5 years question papers from 2057 to 2061 of ELT Theories and methods in terms of 'content coverage' and 'content weighting' and found out that the ELT Theories and Methods tests have high content validity in terms of coverage but low content validity in terms of weighting.

Timilsina (2006) carried out a research on "Content Validity of Language Testing Test at M.Ed. Level." The main objective of this research was to examine the content validity of Language Testing Exam papers at M.Ed. level. For this purpose, the researcher has analyzed the 5 years question papers from 2058 to 2062 B.S. of Language Testing in

terms of 'content coverage' and 'content weighting' and concluded that the language testing tests have low content validity in terms of coverage principle and weighting principle.

The reviews above show that none of the researches have been carried out on "Content Validity of Grammar: Theory and Practice Tests (Eng. Ed. 512)." So, the researcher has selected this topic for the study.

### **1.3 Objectives of the Study**

The objectives of the proposed study were as follows:

- i. To examine the test papers of "Grammar: Theory and Practice Test (Eng. Ed. 512)" in terms of content coverage and content relevance of the tests.
- ii. To suggest some pedagogical implications and recommendations for the improvement of the test.

### **1.4 Significance of the Study**

As a distinct research work from the rest in the Department of English Education, this study will be useful for the department. This study provides information on whether the administered "Grammar: Theory and Practice Tests" had content validity or not. Thus, this study will be useful for those who are directly or indirectly involved in the field of language teaching and learning, especially in language testing. It will be equally useful for the policy makers, curriculum designers and subject experts. No doubt, it will be helpful for the students of the English language (Applied Linguistics). Particularly, it will be valuable for the test designers and teachers as well to make them aware of their mistakes hidden in the tests for the betterment. Hopefully, this research will have a global significance too.

## **CHAPTER II**

### **METHODOLOGY**

This chapter describes the plans and procedures of the study. The researcher followed the following methodology during the study to achieve the desired objectives of the study.

#### **2.1 Sources of Data**

The researcher used only secondary sources of data since primary sources seemed insignificant in this study.

##### **2.1.1 Secondary Sources of Data**

The written test papers of "Grammar: Theory and Practice (Eng. Ed. 512)" that were administered from 2057 to 2063 and the syllabus including the course objectives of the same subject were the main sources of data. The syllabus and the questions are given in appendices-I and II respectively. Apart from these, different related books with the study, journals, theses approved in the Department of English Education and articles related to the study were the other sources of data.

#### **2.2 Process of Data Collection**

The researcher collected all the seven years (2057-2063) subjective and objective test papers on course entitled "Grammar: Theory and Practice (Eng. Ed. 512)". Then, he also collected all the other supporting details or sources, such as, books, theses, journals, etc. from the respected fields. Then after, he studied, analyzed and judged them in order to know whether the test papers have content validity or not in terms of content coverage and content relevance.

### **2.3 Limitations of the Study**

This study had the following limitations:

- i. This study was limited to the content validity of the test papers on "Grammar: Theory and Practice (Eng. Ed. 512)."
- ii. This study was limited to the annual written test papers from 2057 to 2063 of Grammar: Theory and Practice Test of M.Ed, syllabus and course objectives of the same subject.
- iii. This study was limited only in annual written tests.
- iv. The researcher used only tabulation and percentage tools of statistics to present the data.

## **CHAPTER III**

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

This chapter deals with the analysis, interpretation and presentation of the raw data which has been collected. The main purpose of this study was to find out the content validity of "Grammar: Theory and Practice Tests (Eng. Ed. 512)" by analyzing the annual written test papers of seven years of the same subject in terms of content coverage and content relevance of the test items. For this purpose, data analysis is presented into two parts. The first part is concerned with the analysis of content coverage of test papers and the second part with the content relevance.

The researcher analysed the annual test papers of "Grammar: Theory and Practice Test" which were administered in T.U. during seven years (2057-2063) to find out content coverage in all the six units separately (i.e. unitwise) and then as a whole. Then after, the researcher studied and analyzed the test items of "Grammar: Theory and Practice (Eng. Ed. 512)" in relation to the course objectives to find out the content relevance since the test items are said to be relevant if they are congruent with the objectives. Anyway, the researcher analysed the content validity of annual written test papers (i.e. test items) of "Grammar: Theory and Practice (Eng. Ed. 512) Tests" administered in the tests of seven years (2057-2063) using Bachman's two guidelines: content coverage and content relevance to examine the actual content validity among these all.

#### **3.1 Analysis of Content Coverage**

For the purpose of analyzing content coverage of test papers in question, the researcher compared the test contents (i.e. test items) in

relation to course contents or he examined whether the test contained a representative sample of the whole course or not. Since content validity is logical or rational property of a test, there are no any numerical criteria for measuring the adequacy of content validity. The experts of the subject determine how much of the particular course content should be represented in a test so that it can be said to have adequate representation of the course content. But generally, we assume that the test content should be the representative of the whole course content as far as possible and we can say, for example, that a question paper is 100% valid in terms of content coverage if it has represented all the contents of the course. However, it is almost impossible to get such a test which has full coverage. So, the researcher made the criteria to examine content validity in terms of coverage that if the representative sample of the coverage of contents is above 60% then it is believed that the test paper is nearer to content validity but if it is below 60% then it is supposed to lack content validity. Moreover, if more than 70% course contents are covered in the test, then, it is supposed to have high content validity since the more the test items are constructed, the more chances of having content validity. So, to find out the content validity of a test in terms of course representativeness (course coverage), the researcher examined and analysed the "Grammar: Theory and Practice Test Papers" including both subjective and objective questions from the year 2057 to 2063. For this purpose, the researcher dealt with unitwise analysis in question of the test contents (i.e. test items) in relation to course contents. The researcher has used both descriptive method and statistical method (i.e. tabulation and percentage) to analyze the data.



### 3.1.1 Analysis of Content Representativeness (Coverage) in Unit One

**Table 3: Representation of Test Contents in Terms of Course**

#### **Contents in Unit One**

S.N.	Course Contents	Test Contents						
	Language Items	Representation of Test Items						
Unit-1	Traditional Grammar	2057	2058	2059	2060	2061	2062	2063
1.1	Overview			S <sup>1</sup>	S <sup>1</sup>	O <sup>2</sup> , O <sup>11</sup> , O <sup>13</sup>	S <sub>1</sub>	S <sup>7b</sup> (p: opt.)
1.2	General Characteristics		O <sup>13</sup>			O <sup>15</sup> S <sup>1</sup>		
1.3	Limitations	S <sup>5</sup>						
	Total No. of Asked Questions in Unit One	O-0 S-1 L-0	O-1 S-0 L-0	O-0 S-1 L-0	O-0 S-1 L-0	O-4 S-1 L-0	O-0 S-1 L-0	O-0 S-1(P: Opt.) L-0
	Total Unit Marks: 5	6	1	6	6	10	6	3 (Opt.)

The above table shows that there are altogether 3 language items stretching from 1.1 to 1.3 and the whole language items were represented within seven years (2057-2063). Diachronically speaking, among these three language items, most of the questions were represented from the language item 1.1 (Overview) since the five years' questions out of seven years were represented as similar as the same language item. The language item 1.2 (General Characteristics) was represented only in two years' questions out of seven years and the language item 1.3 (Limitations) was represented only in one year out of seven years. Moreover, no subjective long question was asked during seven years (2057-2063) from unit one. However, all the language items were represented in some of the years.

Synchronically speaking, no objective and long questions were asked from unit one, and only one subjective question was asked from 1.3 (limitations) in 2057.

In 2058, only one objective question from 1.2 (General Characteristics) was asked and no subjective question was asked from unit one.

Similarly, in 2059, only one subjective short question ( $S^1$ ) from 1.1 (Overview) was asked and neither long nor objective question was asked from this unit.

In the same way, only one subjective short question ( $S^1$ ) from 1.1 (Overview) was asked hence there were no long question and objective question from this unit in 2060.

Then, in 2061, four objective questions and one subjective short question were asked from this unit. In case of objective questions, the first, second and third ( $O^2$ ,  $O^{11}$  and  $O^{13}$ ) were asked from the same language item 1.1 (Overview), and the fourth ( $O^{15}$ ) was asked from 1.2 (General Characteristics). And in case of subjective question, the only one short question ( $S^1$ ) was asked from 1.2 (General Characteristics) but no long question was asked from this unit in 2061 too.

Likewise, only one subjective short question ( $S^1$ ) was asked from 1.1 (Overview) and no long question and objective question were asked from this unit in 2062.

In 2063, no subjective long question and objective question were asked, and only one subjective short question ( $S^{7b}$ ) which was both partial and optional from 1.1 (Overview) of this unit.

### **Repeated Items**

After analyzing the above data, the researcher also found that from unit one, with in seven years (2057-2063) there is not very good balance of repetition of the language items since some of the language items, for example, 1.1 (Overview) was most represented (i.e. five years out of

seven years), 1.2 (General Characteristics) was represented in two years out of seven and 1.3 (Limitations) was represented only in one year out of seven years. Though the questions were not exactly same but almost same in terms of language items represented in the test (i.e. both 2059<sup>S1</sup> and 2060<sup>S1</sup>) were from the same language items 1.1 (Overview). Similarly, 2058<sup>O13</sup> and 2061<sup>O15</sup> were from the same language item 1.2 (General Characteristics) etc. Anyway, all the language items have content coverage during seven years though the balance of the weightage of language items is just satisfactory.

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

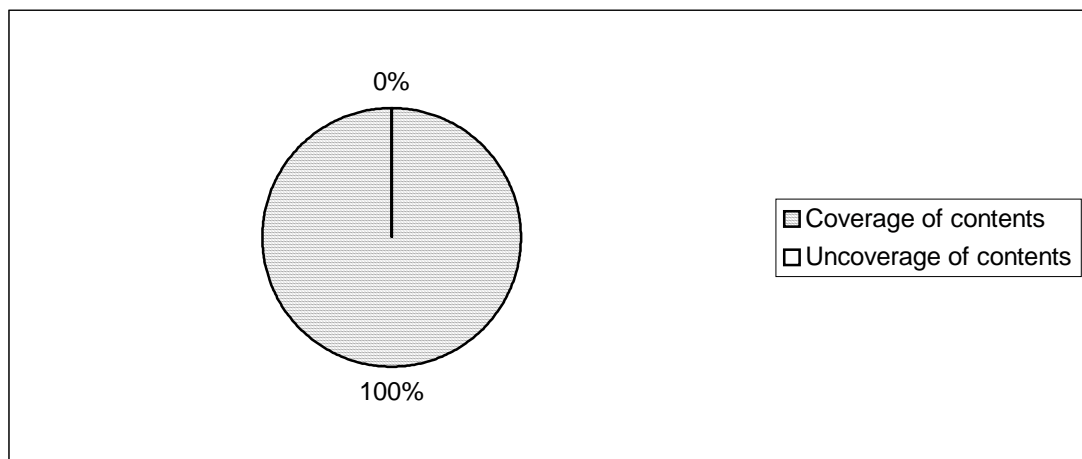
**Figure 1: Content Coverage from Unit One in Percentage**

Total coverage of contents = 3

Total uncoverage of contents = 0

Total areas of contents = 3

∴ Content coverage in percentage =  $\frac{3}{3} \times 100 = 100\%$



As shown in the above table and description, there are three language items in unit one from 1.1 to 1.3 according to course contents and all the language items were represented while constructing the test items during seven years. Therefore, the coverage of course contents in terms of test contents in unit one is 100%. That is why, the content

validity of the test papers of "Grammar: Theory and Practice (Eng. Ed. 512)" in unit one is very high since it has covered more than 70% course contents.

### 3.1.2 Analysis of Content Representativeness (Coverage) in Unit Two

**Table 4: Representation of Test Contents in Terms of Course Contents in Unit Two**

S.N.	Course Contents	Test Contents						
		Representation of Test Items						
Unit-1	Structural Grammar	2057	2058	2059	2060	2061	2062	2063
2.1	Various Senses and Dichatomics in Grammar		S <sup>3</sup>					
2.1.1	Broad Sense and Narrow Senses of Grammar						O <sup>1</sup>	
2.1.2	Notional Grammar and Formal Grammar	O <sup>14</sup>	O <sup>16</sup> , S <sup>3</sup>					
2.2	Notions of Grammatical Unit, Arrangement, Rank, Structure, Class, System, Function, Category, Process, Tool/Device, Realization	O <sup>1</sup>	S <sup>4</sup>	O <sup>12</sup>			O <sup>8</sup>	O <sup>13</sup>
2.3	Morphology and Syntax					S <sup>2</sup>		
2.4	Morph, Morpheme, Allomorph		O <sup>2</sup> , S <sup>1</sup>		O <sup>1</sup> , O <sup>14</sup>			
2.5	Types of Morph and Morpheme			O <sup>15</sup>			O <sup>3</sup>	
2.6	Types of Allomorphic Alternation							S <sup>1</sup>
2.7	Nida's Principle for the Identification of Morpheme			S <sup>2</sup>	S <sup>2</sup>		L <sup>10</sup> (Opt.)	
2.8	Word, Stem, Base, Root, Suffix							O <sup>5</sup>
2.9	Structural Classification of Words, Clauses and Sentences							

2.10	Functional Classification of Phrases, Clauses and Sentences							
2.11	Immediate Constituent (IC) Analysis	S <sup>2</sup>	S <sup>5</sup>			S <sup>3</sup> , O <sup>16</sup>		S <sup>2</sup>
2.12	Types of Syntactic Construction	S <sup>1</sup>	S <sup>4</sup>	O <sup>5</sup> , S <sup>5</sup>			L <sup>10</sup> (Opt.)	
2.13	Types of Syntactic Linkage	S <sup>1</sup> , O <sup>2</sup>		S <sup>3</sup>	O <sup>16</sup>			
	Total No. of Asked Questions in Unit Two	O-3 S-2 L-0	O-2 S-4 L-0	O-3 S-3 L-0	O-3 S-1 L-0	O-1 S-2 L-0	O-3 S-0 L-2	O-2 S-2 L-0
	Total Unit Marks: 15	15	26	21	9	13	15	14

The above table shows that there are altogether 15 language items in unit two stretching from 2.1 to 2.13. Within seven years (2057-2063), out of 15 language items, 13 language items were represented in the tests. Looking at the data diachronically, among the 15 language items, the most represented one was 2.2 (Notions of Grammatical Unit, Arrangement, Rank, Structure, Class, System, Functions, Category, Process, Tool/.Device, Realization) which was represented in five years out of seven years. Similarly, the language items 2.11 (Immediate Constituent (IC) Analysis), and 2.12 (Types of Syntactic Construction) were represented in four years; then, the language items 2.7 (Nida's Principle for the Identification of Morpheme and 2.13 (Types of Syntactic Linkage) were represented in three years; then after, the language items 2.1.2 (Notional Grammar and Formal Grammar), 2.4 (Morph, Morpheme, Allomorph) and 2.5 (Types of Morph and Morpheme) were represented in two years, and then the language items 2.1 (Various Senses and Dichotomies in Grammar), 2.1.1 (Broad Sense and Narrow Sense of Grammar), 2.3 (Morphology and Syntax), and 2.8 (Word, Stem, Base, Root, Suffix) were represented only in one year out of seven years. After

all, the other remaining language items which were not represented are 2.9 (Structural Classification of Words, Clauses and Sentences) and 2.10 (Functional Classification of Phrases, Clauses and Sentences).

Looking at the data synchronically, three objective questions and two subjective short questions were asked from this unit in 2057. In case of objective questions, the first (O<sup>14</sup>) was asked from 2.1.2 (Notional Grammar and Formal Grammar), the second (O<sup>1</sup>) was asked from 2.2 (*Notions of Grammatical Unit, Arrangement, Rank, Structure, Class, System, Function, Category, Process, Tool/Device, Realization*), and the third (O<sup>2</sup>) was asked from 2.13 (Types of Syntactic Linkage). Then, in case of subjective short questions, the first (S<sup>2</sup>) was asked from 2.11 [Immediate Constituent (IC) Analysis] and the next (S<sup>1</sup>) was collectively asked from 2.12 (Types of Syntactic Construction) and 2.13 (Types of Syntactic Linkage). Moreover, no subjective long question was asked from this unit during seven years (2057-2063) except in 2062.

In 2058, two objective questions and four subjective short questions were asked from this unit. In case of objective questions, the first (O<sup>16</sup>) was asked from 2.1.2 (Notional Grammar and Formal Grammar) and the next (O<sup>2</sup>) was asked from 2.4 (Morph, Morpheme, Allomorph). Then, in case of subjective short questions, the first (S<sup>3</sup>) was asked under 'Various Senses and Dichotomies in Grammar' from 2.1.2 (Notional Grammar and Formal Grammar), the second (S<sup>4</sup>) was collectively asked from 2.2 (*Notions of Grammatical Unit, Arrangement, Rank, Structure, Class, System, Function, Category, Process, Tool/Device, Realization*) and 2.12 (Types of Syntactic Construction), the third (S<sup>1</sup>) was asked from 2.4 (Morph, Morpheme, Allomorph), and the fourth (S<sup>5</sup>) was asked from 2.11 [Immediate Constituent (IC) Analysis].

In 2059, three objective questions and three subjective short questions were asked from this unit. In case of objective questions, the first (O<sup>12</sup>) was asked from 2.2 (*Notions of Grammatical Unit, Arrangement, Rank, Structure, Class, System, Function, Category, Process, Tool/Device, Realization*), the second (O<sup>15</sup>) was asked from 2.5 (Types of Morph and Morpheme), and the third (O<sup>5</sup>) was asked from 2.12 (Types of Syntactic Construction). Then, in case of subjective short questions, the first (S<sup>2</sup>) was asked from 2.7 (Nida's Principle for the Identification of Morpheme), the Second (S<sup>5</sup>) was asked from 2.12 (Types of Syntactic Construction), and the third (S<sup>3</sup>) was asked from 2.13 (Types of Syntactic Linkage).

In 2060, three objective questions and only one subjective short question were asked from this unit. In case of objective questions, both the first (O<sup>1</sup>) and the second (O<sup>14</sup>) were asked from the same language item 2.4 (Morph, Morpheme, Allomorph), and the Third (O<sup>16</sup>) was asked from 2.13 (Types of Syntactic Linkage). Then, in case of subjective short question, only one subjective short question (S<sup>2</sup>) was asked from 2.7 (Nida's principle for the Identification of Morpheme).

Similarly, in 2061, only one objective question and two subjective short questions were asked from this unit. In case of objective question, only one objective question (O<sup>16</sup>) was asked from 2.11 [Immediate constituent (IC) analysis]. Then, in case of subjective short questions, the first (S<sup>2</sup>) was asked from 2.3 (Morphology and Syntax) and the next (S<sup>3</sup>) was asked from 2.11 [Immediate Constituent (IC) Analysis].

Likewise, in 2062, three objective questions and two subjective long questions were asked from this unit. In case of objective questions, the first (O<sup>1</sup>) was asked from 2.1.1 (Broad Sense and Narrow Sense of Grammar), the second (O<sup>8</sup>) was asked from 2.2 (*Notions of Grammatical*

Unit, Arrangement, Rank, Structure, Class, System, Function, *Category*, Process, Tool/Device, Realization), and the third (O<sup>3</sup>) was asked from 2.5 (Types of Morph and Morpheme). Then, in case of subjective long questions, one long question was asked from 2.7 (Nida's Principle for the Identification of Morpheme) and the other one was asked from 2.12 (Types of Syntactic Construction). Both of them were asked under the same question No. 10 as alternative or optional questions. It means, any one of two should be done.

Then, in 2063, two objective questions and two subjective short questions were asked from this unit. In the case of objective questions, the first (O<sup>13</sup>) was asked from 2.2 (*Notions of Grammatical Unit, Arrangement, Rank, Structure, Class, System, Function, Category, Process, Tool/Device, Realization*) and the next (O<sup>5</sup>) was asked from 2.8 (Word, Stem, Base, Root, Suffix). Then, in case of subjective short questions, the first (S<sup>1</sup>) was asked from 2.6 (Types of Allomorphic Alternation) and the next (S<sup>2</sup>) was asked from 2.11 [Immediate Constituent (IC) Analysis].

### **Repeated Items**

After analyzing the above data, the researcher also found that from unit two, within seven years (2057-2063), only four objective questions and eleven subjective short questions were repeated. In case of objective questions, 2057<sup>O1</sup> and 2062<sup>O8</sup> were exactly same from 2.2 (Notions of Grammatical Category). Similarly, 2057<sup>O2</sup> and 2060<sup>O16</sup> were same except some terminological difference. And both of them were from 2.13 (Types of Syntactic Linkage). Then, in case of subjective short questions, 2059<sup>S2</sup>, 2060<sup>S2</sup> and 2062<sup>L10</sup> were same except some difference in terminology and weightage of the question. And all of them were from the same language item 2.7 (Nida's Principle for the Identification of Morpheme). Similarly, 2057<sup>S2</sup>, 2058<sup>S5</sup>, 2061<sup>S3</sup> and 2063<sup>S2</sup> were same except some terminological



difference. And all of them were from the same language item 2.11 [Immediate Constituent (IC) Analysis]. Then, 2057<sup>S1</sup>, 2058<sup>S4</sup>, 2059<sup>S5</sup>, and 2062<sup>L10</sup> were same except some difference in terminology, focus point and weightage of the question. They all (S<sup>1</sup>, S<sup>4</sup>, S<sup>5</sup>, and L<sup>10</sup>) were from the same language item 2.12 (Types of Syntactic Construction). In addition to this, 2057<sup>S1</sup> and 2059<sup>S3</sup> were also same from 2.13 (Types of Syntactic Linkage) but only in some of the portions of the questions. However, the repetition of the questions in unit two was satisfactory.

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

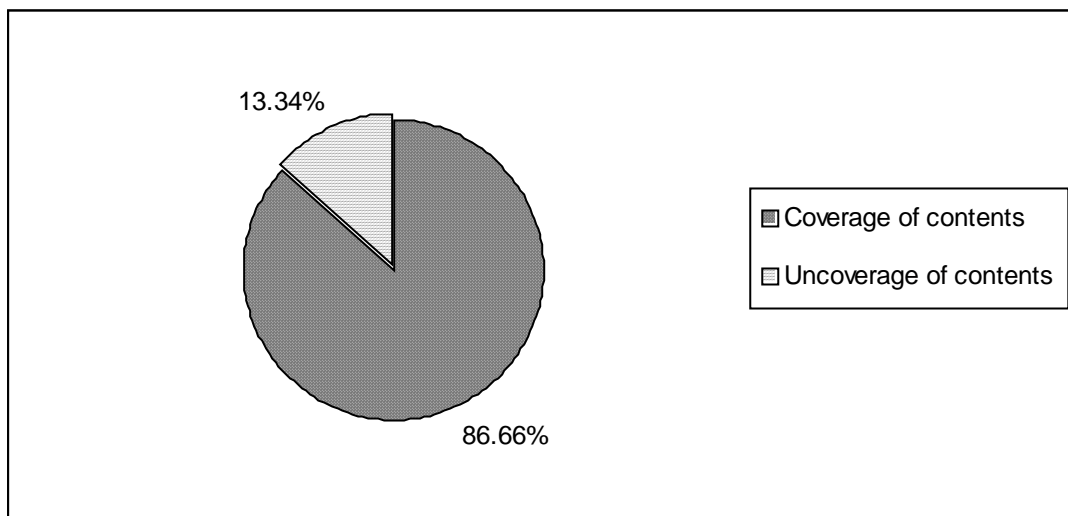
**Figure 2: Content Coverage from Unit Two in Percentage**

Total coverage of contents = 13

Total uncoverage of contents = 2

Total areas of contents = 15

∴ Content coverage in percentage =  $13/15 \times 100 = 86.66\%$



From the above table and description, it is clear that there are 15 language items in unit two from 2.1 to 2.13 according to course contents. Out of them, 13 language items were represented in the test during the seven years (2057-2063) but two language items were discarded while constructing test items. Therefore, the coverage of course contents in

terms of test contents in unit two is 86.66%. 13.34% of the contents was not covered in the test. However, "Grammar: Theory and Practice Tests" have high content validity (here content validity = content coverage) in unit two in a whole since they have covered more than 70% course contents but if we see the content coverage of unit two yearwise, then the content coverage of it in 2058 was very high and in 2060 was very low in terms of test items represented in the tests in these years.

### 3.1.3 Analysis of Content Representativeness (Coverage) in Unit Three

**Table 5: Representation of Test Contents in Terms of Course Contents in Unit Three**

S.N.	Course Contents	Test Contents						
	Language Items	Representation of Test Items						
Unit-3	Generative Grammar	2057	2058	2059	2060	2061	2062	2063
3.1	Basic Assumptions	L <sup>9</sup>						
3.1.1	The Meaning of Generation	L <sup>9</sup>						
3.1.2	Basic Concepts	L <sup>9</sup>						
3.1.3	Competence and Performance							
3.1.4	Levels of Adequacy					S <sup>4</sup>	O <sup>7</sup>	O <sup>15</sup> , S <sup>4</sup>
3.1.5	Markedness and Core Grammar							O <sup>14</sup>
3.2	Versions of Generative Grammar					L <sup>9</sup>	L <sup>9</sup>	
3.2.1	1957 Model	S <sup>4</sup>		S <sup>4</sup>	S <sup>5</sup>	L <sup>9</sup>	L <sup>9</sup> , S <sup>6</sup>	L <sup>9</sup>
-	Phrase Structure Rules	√			O <sup>9</sup>	O <sup>1</sup>		
-	Transformational Rules	√			√		√	
-	Morphophonemic Rules							
3.2.2	1965 Model/Standard Theory		L <sup>9</sup>	L <sup>11</sup> , S <sup>4</sup>		L <sup>9</sup>	L <sup>9</sup>	L <sup>9</sup>

-	Syntactic Component : ▪ Base Categorical Rules and Lexicon ▪ Transformations		√	√				
-	Semantic Component		√					S <sup>7c</sup> (P: Opt.)
-	Phonological Component		√	O <sup>11</sup>				
3.2.3	GB Syntax			L <sup>9</sup> (Opt.)	L <sup>11</sup>	L <sup>9</sup>	L <sup>9</sup>	
-	The Base Component: D-Structures, X-Bar Theory, θ-Criterion, Lexicon				√		√	
-	The Transformational Component: S-Structure, Movement and Bounding Theory			O <sup>8</sup>			√	O <sup>9</sup>
-	The LF Component: Binding Theory						√, O <sup>14</sup> , O <sup>15</sup>	
-	The PP Component: The Empty Category Principle			O <sup>7</sup>	√		√	
-	Case Theory			√			√	
3.3	Argumentation			L <sup>9</sup> (Opt.)	S <sup>3</sup>			
-	Syntactic Argumentation			√	√			S <sup>7a</sup> (P:Opt)
-	Constituency: Movement and Substitution			√				
-	Constituency: Some Additional Tests			√				
	Total No. of Asked Questions in Unit Three	O-0 S-1 L-1	O-0 S-0 L-1	O-3 S-1 L-3	O-1 S-2 L-1	O-1 S-1 L-1	O-3 S-1 L-1	O-3 S-2[1(P:Opt.)] L-1
	Total Unit Marks: 20	18	12	33	25	19	21	24

As shown in the above table, there are altogether 25 language items in unit three stretching from 3.1 to 3.3. Out of them, 24 language items were represented in the test during the seven years (2057-2063) . Looking

at the data diachronically, among these 25 language items, the most represented one was 3.2.1 (1957 Model) which was represented in six years out of seven years. Similarly, the language items 3.2.1 (Phrase Structure Rules, Transformational Rules) and 3.2.2 (1965 Model) were represented in five years; 3.2.3 (GB Syntax) was represented in four years; 3.1.4 (Levels of Adequacy), 3.2.2 [Syntactic Component (Base: Categorical Rules and Lexicon, and Transformations), Semantic Component, and Phonological Component], 3.2.3 (The Transformational Component: S-Structure, Movement and Bounding Theory, and the PP Component: the Empty Category Principle), and 3.3 (Syntactic Argumentation) were represented in three years; 3.2 (Versions of Generative Grammar), 3.2.1 (Morphophonemic rules), 3.2.3 (The Base Component: D-Structure, X-Bar Theory,  $\theta$ -Criterion, Lexicon, and Case Theory), and 3.3 (Argumentation) were represented in two years, and 3.1 (Basic Assumptions), 3.1.1 (The Meaning of Generation) 3.1.2 (Basic Concepts), 3.1.5 (Markedness and Core Grammar), 3.2.3 (The LF Component: Binding Theory), 3.3 (Constituency: Movement and Substitution, and Constituency: Some Additional Tests) were represented only in one year out of seven years. But the language item 3.1.3 (Competence and Performance) was not represented in any year.

Looking at the data synchronically, one subjective short and one long questions were asked from this unit in 2057. In case of subjective short and long questions, one short question ( $S^4$ ) was asked from 3.2.1 (Phrase Structure Rules and Transformational Rules), and one long question ( $L^9$ ) was collectively asked from 3.1 (Basic Assumptions), 3.1.1 (The Meaning of Generation), 3.1.2 (Basic Concepts). Moreover, there was no representation of objective question in this year.

In 2058, there was no representation of both objective question and subjective short question, and only one long question (L<sup>9</sup>) was asked from 3.2.2 (1965 Model).

In 2059, three objective questions, one subjective short, and three subjective long questions were asked from this unit. Regarding objective questions, the first (O<sup>11</sup>) was asked from 3.2.2 (Phonological Component), the second (O<sup>8</sup>), and the third (O<sup>7</sup>) were asked from 3.2.3 (The Transformational Component: S-Structure, *Movement* and Bounding Theory, and the PP Component: *The Empty Category Principle* Respectively). Then, regarding subjective short question, only one question (S<sup>4</sup>) was integratively asked from 3.2.1 (1957 Model), and 3.2.2 (1965 Model). Then after, regarding subjective long questions, the first (L<sup>1</sup>) was asked from 3.2.2 (Syntactic Component: Transformations); both the second and the third were asked from 3.2.3 (Case Theory) and 3.3 (Argumentation) respectively. But both the second and third were asked under the same question No. 9 as alternative or optional questions. It means any one of two should be attempted.

In 2060, one objective question, two subjective short questions and one long question were asked from this unit. Regarding objective question, only one question (O<sup>9</sup>) was asked from 3.2.1 (Phrase Structure Rules). Then, regarding subjective short questions, the first (S<sup>5</sup>) was asked from 3.2.1 (Transformational Rules) and the next (S<sup>3</sup>) was asked from 3.3 (Syntactic Construction). Then after, regarding long question, only one long question (L<sup>11</sup>) was asked from 3.2.3 (GB Syntax Emphasizing the Base Component: D-Structure, X-Bar Theory, „-*Criterion*, Lexicon, and the PP Component: *The Empty Category Principle*).

Similarly, in 2061, one objective, one subjective short and one long questions were asked from this unit. Regarding these questions, only one objective question (O<sup>1</sup>) was asked from 3.2.1 (Phrase Structure Rules); only one subjective short question (S<sup>4</sup>) was asked from 3.1.4 (Levels of Adequacy), and only one long question (L<sup>9</sup>) was collectively asked under 3.2 (Versions of Generative Grammar) from 3.2.1 (1957 was collectively asked under 3.2 (Versions of Generative Grammar), from 3.2.1 (1957 Model) 3.2.2 (1965 Model), and 3.2.3 (GB Syntax).

Then, in 2062, three objective questions, one subjective short and one long questions were asked from this unit. Regarding objective questions, the first (O<sup>7</sup>) was asked from 3.1.4 (Levels of Adequacy), and both the second (O<sup>14</sup>) and the third (O<sup>15</sup>) were asked from the same language item 3.2.3 (The LF Component: Binding Theory). Then, regarding subjective short and long questions, only one short question (S<sup>6</sup>) was asked from 3.2.1 (Transformational Rules), and only one long question (L<sup>9</sup>) was asked from 3.2 (Versions of Generative Grammar) with special reference to GB Theory.

Then after, in 2063, three objective, two subjective short and one long questions were asked from this unit. Regarding objective questions, the first (O<sup>15</sup>) was asked from 3.1.4 (Levels of Adequacy), the second (O<sup>14</sup>), 3.1.5 (Markedness and Core Grammar), and the third (O<sup>9</sup>) from 3.2.3 (The Transformational Component: S-Structure, Movement and *Bounding Theory*). Then, regarding subjective short questions, the first (S<sup>4</sup>) was asked from 3.1.4 (Levels of Adequacy), and the next (S<sup>7</sup>) was collectively asked as, 'S<sup>7a</sup>' from 3.3 (Syntactic Argumentation) and 'S<sup>7c</sup>' from 3.2.2 (Semantic Component). And both 'S<sup>7a</sup>' and 'S<sup>7c</sup>' were asked under the same question (S<sup>7</sup>) as both partial and optional questions.

Moreover, regarding long question, only one long question was collectively asked from 3.2.1 (1957 Model) and 3.2.2 (1965 Model).

### **Repeated Items**

After analyzing the above data, the researcher also found that from unit three, within seven years (2057-2063), five subjective short questions and three long questions were repeated. Looking at these questions integratively, 2061<sup>S4</sup> and 2063<sup>S4</sup> were same from 3.1.4 (Levels of Adequacy) except some terminological difference. Similarly, 2061<sup>L9</sup> and 2062<sup>L9</sup> were same except some difference in terminology and focus point of the questions. Both of them were from the same language item 3.2 (Versions of Generative Grammar). Then, 2060<sup>S5</sup> and 2062<sup>S6</sup> were same except some difference in terminology and focus point in question. Both of them were from 3.2.1 (Transformational Rules). Then after, 2059<sup>S4</sup> and 2063<sup>L9</sup> were similar except some difference in terminology, focus (Point) and weightage of the questions. Both of them were integratively asked from the language items 3.2.1 (1957 Model) and 3.2.2 (1965 Model). In addition to this, 2060<sup>S3</sup> and 2063<sup>S7a</sup> were also same from 3.3 (Syntactic Argumentation) but only in some of the portions in question. Reviewing the above table, some of the language items were represented almost every year, for example, 3.2.1 (1957 Model) but one of them was not represented in any year, for example, 3.1.3 (Competence and Performance). It doesn't seem good so it would be appreciable to include the neglected language item instead of repeating the same because 'one getting too much and the other one being starved is not good'. However, the repetition of the test items in unit three is satisfactory.

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

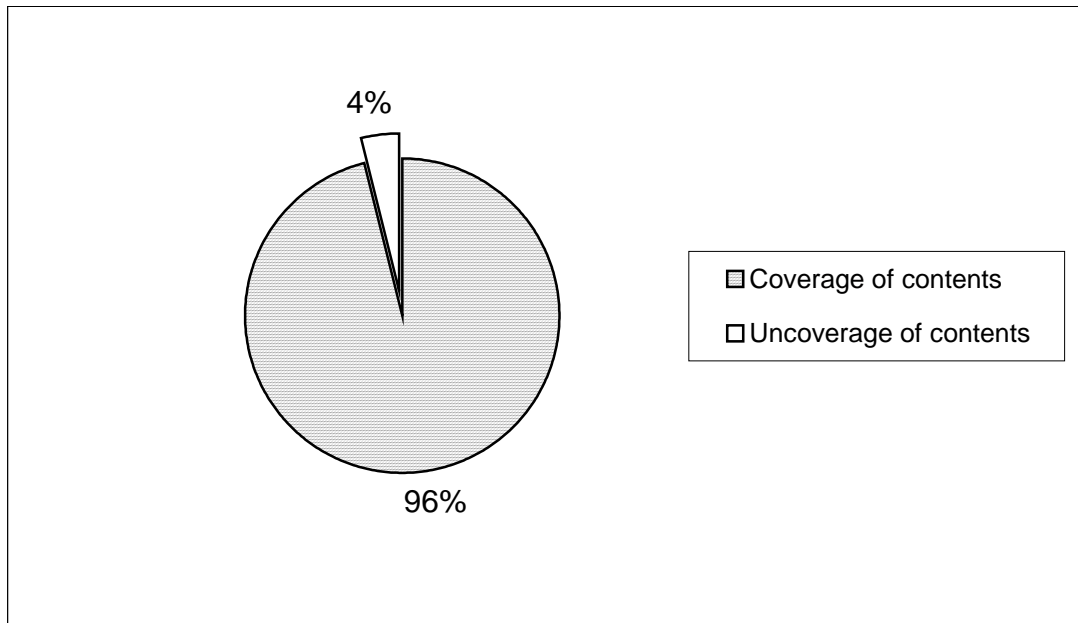
### Figure 3: Content Coverage from Unit Three in Percentage

Total coverage of contents – 24

Total uncoverage of contents – 1

Total areas of contents – 25

$$\therefore \text{Content coverage in percentage} = \frac{24}{25} \times 100 = 96\%$$



From the above table and description, it is clear that there are 25 language items in unit three from 3.1 to 3.3 according to course contents. Out of them, 24 language items were represented in the test during the seven years (2057-2063) but one language item was neglected while constructing test items. Therefore, the coverage of course contents in terms of test contents in unit three is 96%. 4% of the contents was not covered in the test. However, "Grammar: Theory and Practice Tests" have high content validity (Content Coverage) in unit three since they have covered more than 70% course contents.



### 3.1.4 Analysis of Content Representativeness (Coverage) in Unit Four

**Table 6: Representation of Test Contents in Terms of Course**

#### **Contents in Unit Four**

S.N.	Course Contents	Test Contents						
	Language Items	Representation of Test Items						
Unit-4	Other Models of Grammar	2057	2058	2059	2060	2061	2062	2063
4.1	Generalized Phrase Structure Grammar (GPSG)			O <sup>16</sup>				
4.1.1	Framework							
4.2	Lexical Functional Grammar (LFG)			L <sup>10</sup>				L <sup>10</sup> (Opt.)
4.2.1	Framework			L <sup>10</sup>				L <sup>10</sup> (Opt.)
4.3	Relational Grammar (RG)				O <sup>11</sup>		S <sup>3</sup>	
4.3.1	Framework						S <sup>3</sup>	
4.4	Case Grammar	O <sup>15</sup> , L <sup>10</sup> (Opt.)	O <sup>7</sup> , O <sup>15</sup> , L <sup>10</sup> (Opt.)	O <sup>6</sup>	O <sup>7</sup> , O <sup>10</sup> L <sup>9</sup> (Opt.)	L <sup>10</sup> (Opt.)	O <sup>11</sup>	O <sup>3</sup> , O <sup>7</sup>
4.5	Systematic Grammar	L <sup>10</sup> (Opt.)	L <sup>10</sup> (Opt.)		L <sup>9</sup> (Opt.)			
4.6	Tagmemic Grammar (Tagmemics)		O <sup>1</sup> , O <sup>14</sup>		O <sup>5</sup>	L <sup>10</sup> (Opt.)		O <sup>10</sup> L <sup>10</sup> (Opt.)
4.7	Stratificational Grammar	O <sup>13</sup>				L <sup>10</sup> (Opt.)		O <sup>1</sup>
4.8	Functional Grammar						S <sup>2</sup>	
4.9	Communicative Grammar				O <sup>2</sup>	O <sup>12</sup>		
	Total No. of Asked Questions In Unit Four	O-2 S-0 L-2	O-4 S-0 L-2	O-2 S-0 L-1	O-5 S-0 L-2	O-1 S-0 L2	O-1 S-2 L-0	O-4 S-0 L-1
	Total Unit Marks : 15	14	16	14	17	13	13	16

The above table shows that there are altogether 12 language items in unit four stretching from 4.1 to 4.9. Within seven years (2057-2063), out of 12 language items, 11 language items were represented in the test.

If we see diachronically, among these twelve language items, the most represented one was 4.4 (Case Grammar) which was represented in each and every year's examination from 2057 to 2063. Similarly, the language item 4.6 (Tagmemic Grammar), was represented in four years; 4.7 (Stratificational Grammar) was represented in three years; 4.2 (Lexical Functional Grammar), 4.2.1 (Framework of LFG), 4.3 (Relational Grammar), and 4.9 (Communicative Grammar) were represented only in two years; then the language items 4.1 (Generalized Phrase Structure) and 4.8 (Functional Grammar) were represented only in one year out of seven years but the language item 4.1.1 (Framework of GPSG) was not represented in any year during seven years. If we observe the above table and description, it shows that the language item 4.4 (Case Grammar) is the most dominant one since it has coverage in all the seven years. However, this is said to be dominant, it would be better not to repeat the same item every year but to cover the untouched language items instead alternatively. However, all the language items except 4.1.1 (Framework of GPSG) were represented so this unit is said to have content validity in terms of coverage.

Synchronically speaking, two objective questions and two subjective long questions were asked from this unit in 2057. In case of objective questions, the first (O<sup>15</sup>) was asked from 4.4 (Case Grammar) and the next (O<sup>13</sup>) was asked from 4.7 (Stratificational Grammar). Then, in case of subjective long questions, one long question was asked from 4.4 (Case Grammar) and the other one was asked from 4.5 (Systematic Grammar). Both of them were asked under the same question No. 10 as alternative questions or optional questions. It means, one of the two questions should be done. Moreover, no subjective short question was asked from this unit during seven years (2057-2063) except in 2062.

In 2058, four objective questions and two subjective long questions were asked from this unit. In case of objective questions, the first (O<sup>7</sup>) and the second (O<sup>15</sup>) were asked from the language item 4.4 (Case Grammar), and the third (O<sup>1</sup>) and the fourth (O<sup>14</sup>) were asked from 4.5 (Tagmemic Grammar). Then, in case of subjective long questions, both the two long questions were asked exactly in the same way as in 2057 from the same language items but they had just reversed their position to each other.

In 2059, two objective questions and one subjective long question were asked from this unit. In case of objective questions, the first (O<sup>16</sup>) was asked from 4.1 (Generalized Phrase Structure Grammar) and the next (O<sup>6</sup>) was asked from 4.4 (Case Grammar). Then, in case of subjective long question, only one long question (L<sup>10</sup>) was collectively asked from 4.2 (Lexical Functional Grammar) and 4.2.1 (Framework of LFG) as a compulsory question.

In 2060, five objective questions and two subjective long questions were asked from this unit. In case of objective questions, the first (O<sup>11</sup>) was asked from 4.3 (Relational Grammar), both the second (O<sup>7</sup>) and the third (O<sup>10</sup>) were asked from 4.4 (Case Grammar), the fourth (O<sup>5</sup>) was asked from 4.6 (Tagmemic Grammar), and the fifth (O<sup>2</sup>) was asked from 4.9 (Communicative Grammar). Then, in case of subjective long questions, one long question was asked from 4.5 (Systematic Grammar) and the other one was asked from 4.4 (Case Grammar). Both of them were asked under the same question No. 9 as alternative questions or optional questions. It means, any one of two should be done

In 2061, only one objective question and two subjective long questions were asked from this unit. In case of objective question, only

one question ( $O^{12}$ ) was asked from 4.9 (Communicative Grammar). Then, in case of subjective long questions, one long question was asked from 4.4 (Case Grammar) and the other one was collectively asked from the language items 4.6 (Tagmemic Grammar) and 4.7 (Stratificational Grammar). Both of them were asked in the same question No. 10 as alternative questions or optional questions. It means, anyone of two should be done.

Similarly, in 2062, only one objective question and two subjective short questions were asked from this unit. In case of objective question, only one question ( $O^{11}$ ) was asked from 4.4 (Case Grammar). Then, in case of subjective short questions, the first question ( $S^3$ ) was collectively asked from 4.3 (Relational Grammar) and 4.3.1 (Framework of RG), and the second one ( $S^2$ ) was asked from 4.8 (Functional Grammar). Moreover, there is no representation of long question from this unit in 2062.

Likewise, in 2063, four objective questions and one subjective long question were asked from this unit. In case of objective questions, the first question ( $O^3$ ) and the second question ( $O^7$ ) were asked from 4.4 (Case Grammar), the third question ( $O^{10}$ ) was asked from 4.6 (Tagmemic Grammar), and the fourth one ( $O^1$ ) was asked from 4.7 (Stratificational Grammar). Then, in case of subjective long question, only one long question ( $L^{10}$ ) was collectively asked from 4.2 (Lexical Functional Grammar) and 4.2.1 (Framework of LFG), and also added a question from the next language item 4.6 (Tagmemic Grammar) in the same question (i.e.  $L^{10}$ ) as a choice or optional question so that testees can answer any one part of the question either the first part or the second part.

## Repeated Items

After analyzing the above data, the researcher also found that from unit four, within seven years (2057-2063), six objective questions and nine subjective long questions were repeated. In case of objective questions, 2057<sup>O15</sup> and 2063<sup>O7</sup> were same except some terminological difference. Both were from the same language item 4.4 (Case Grammar). Similarly, 2059<sup>O6</sup> and 2060<sup>O7</sup> were also same except some terminological difference from the same language item 4.4 (Case Grammar). Likewise, 2057<sup>O13</sup> was repeated in place of 2063<sup>O1</sup> except some terminological difference. This was from the language item 4.7 (Stratificational Grammar). Then, in case of subjective long questions, 2059<sup>L10</sup>, collectively asked from 4.2 (LFG) and 4.2.1 (Framework of LFG) was exactly repeated in place of 2063<sup>L10</sup> but a question from the language item 4.6 (Tagmemic Grammar) was also added at the latter part of the same repeated question (L<sup>10</sup>) of 2063 as a choice or optional question. In the same way, subjective long questions: 2057<sup>L10</sup>, 2058<sup>L10</sup>, 2060<sup>L9</sup> and 2061<sup>L10</sup> from the same language item 4.4 (Case Grammar) were same except some terminological difference. Likewise, subjective long questions, 2057<sup>L10</sup>, 2058<sup>L10</sup> and 2060<sup>L9</sup> from the same language item 4.5 (Systematic Grammar) were also same except some terminological difference. Moreover, no subjective short question was repeated in any year. Therefore, it would be better to reduce the frequency of repetition of the same questions and increase the percentage of proper coverage of the language items.

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

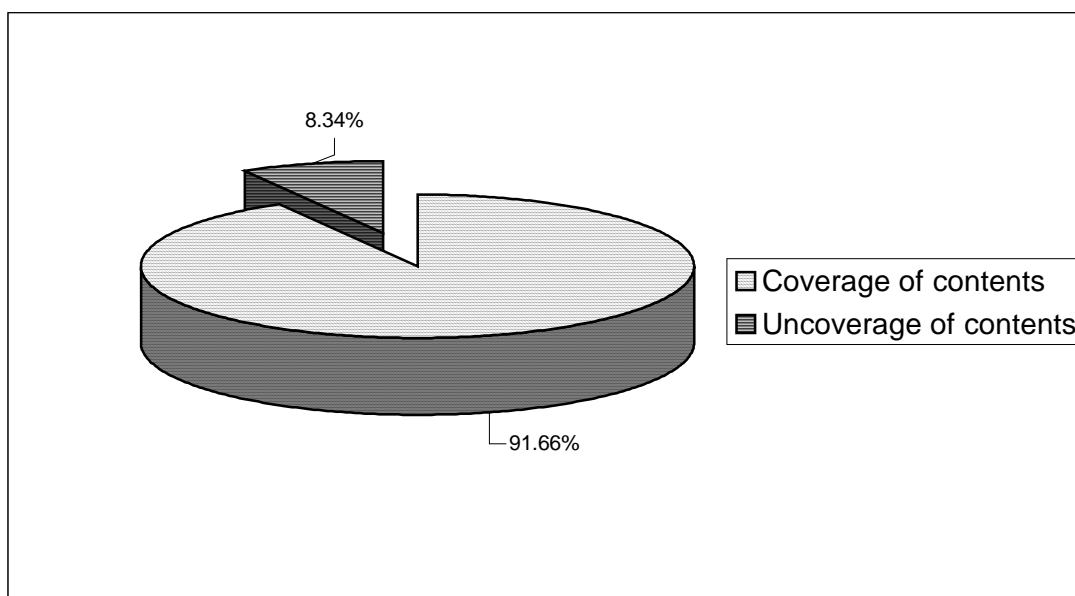
#### Figure 4: Content Coverage from Unit Four in Percentage

Total coverage of contents = 11

Total uncoverage of contents = 1

Total areas of contents = 12

∴ Content coverage in percentage =  $11/12 \times 100 = 91.66\%$



From the above table and description, it is clear that there are 12 language items in unit four from 4.1 to 4.9 according to course contents. Out of them, eleven language items were represented in the test during the seven years (2057-2063) but one language item was discarded or missed while constructing test items. Therefore, the coverage of course contents in terms of test contents in unit four is 91.66%. 8.34% of contents was not covered in the test papers. However, the test papers of "Grammar: Theory and Practice" have high content validity in unit four since they have covered more than 70% course contents.

### 3.1.5 Analysis of Content Representativeness (Coverage) in Unit Five

**Table 7: Representation of Test Contents in Terms of Course**

#### **Contents in Unit Five**

S.N.	Course Contents	Test Contents						
	Language Items	Representation of Test Items						
Unit-1	Pedagogic Grammar	2057	2058	2059	2060	2061	2062	2063
5.1	Theoretical Grammar and Pedagogic Grammar	S <sup>3</sup>			S <sup>4</sup>	S <sup>5</sup>		
5.2	Characteristic Features of Pedagogic Grammar		S <sup>2</sup>	S <sup>3</sup>			S <sup>7</sup>	
5.3	Writing a Pedagogic Grammar							S <sup>5</sup>
5.4	Using a Pedagogic Grammar							
	Total No. of Asked Questions in Unit: 5	O-0 S-1 L-0	O-0 S-1 L-0	O-0 S-1 L-0	O-0 S-1 L-0	O-0 S-1 L-0	O-0 S-1 L-0	O-0 S-1 L-0
	Total Unit Marks: 5	6	6	6	6	6	6	6

The above table shows that there are altogether 4 language items in unit five stretching from 5.1 to 5.4. Within seven years (2057-2063), out of four language items, three language items were represented in the test. Looking diachronically, among these four language items, the most represented ones were 5.1 (Theoretical Grammar and Pedagogical Grammar) and 5.2 (Characteristic Features of Pedagogic Grammar) which were represented in three years out of seven years. 5.3 (Writing a Pedagogic Grammar) was represented only in one year out of seven years. But the language item 5.4 (Using a Pedagogic Grammar) was not represented in any year. If we observe the above representative table, it seems that 5.1 (Theoretical and Pedagogical Grammar) and 5.2

(Characteristic Features of Pedagogic Grammar) were the most dominant language items and the rest ones were minor or less important. There are few language items in unit five so it would be better if the test items covered the whole language items maintaining proper weightage of them as far as possible than the repetition of the same language items time and again. Moreover, no subjective long question and objective question were asked during seven years (2057-2063) from unit five. However, all the language items except 5.4 (Using a Pedagogic Grammar) were represented so this unit is said to have content validity in terms of coverage.

Synchronically speaking, no long and objective questions were asked in any year (i.e. 2057-2063) from unit 5 as I already mentioned above. Hence, only one subjective question was asked from the language items 5.1 (Theoretical Grammar and Pedagogic Grammar), 5.2 (Characteristic Features of Pedagogic Grammar), and 5.3 (Writing a Pedagogic Grammar) in 2057, 2058, 2059, 2060, 2061, 2062 and 2063 (i.e. each seven years) respectively. But the language item 5.4 was not represented in any seven years.

### **Repeated Items**

After analyzing the above data, the researcher also found that from unit five, within seven years (2057-2063), only six subjective short questions were repeated or same. Regarding them, 2057<sup>S3</sup>, 2060<sup>S4</sup> and 2061<sup>S5</sup> were exactly same. They all (S<sup>3</sup>, S<sup>4</sup> and S<sup>5</sup>) from the language item 5.1 (Theoretical Grammar and Pedagogic Grammar). Similarly, 2058<sup>S2</sup>, 2059<sup>S3</sup> and 2062<sup>S7</sup> were same or repeated except some terminological difference. All of them were asked from the same language item 5.2 (Characteristic Features of Pedagogic Grammar). In case of subjective long and objective questions, there was no chance of repetition of the



long and objective questions since there were no such questions during the period. Therefore, it would be better to reduce the frequency of repetition of the questions and increase the percentage of coverage of the language items.

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

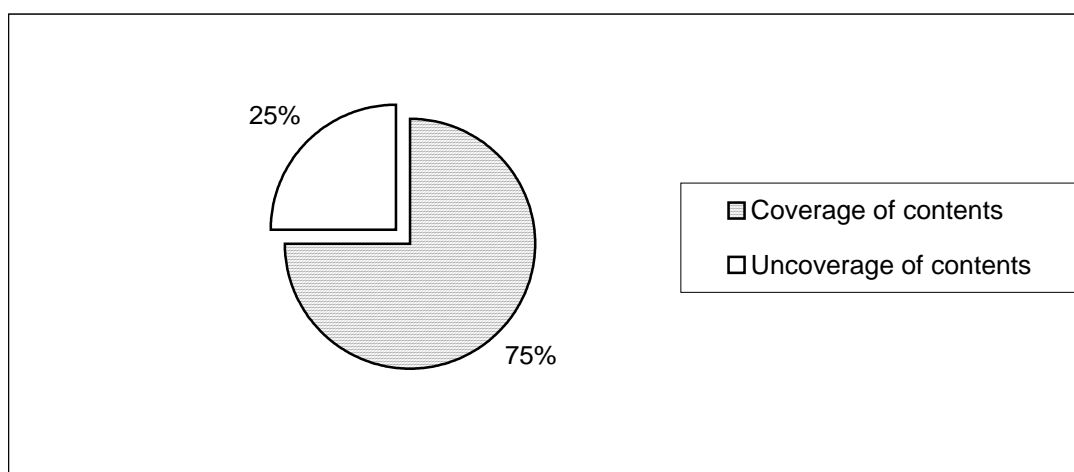
**Figure 5: Content Coverage from Unit Five in Percentage**

Total coverage of contents = 3

Total uncoverage of contents = 1

Total areas of contents = 4

∴ Content coverage in percentage =  $\frac{3}{4} \times 100 = 75\%$



From the above table and description, it is clear that, there are 4 language items in unit five from 5.1 to 5.4 according to course contents. Out of them, only three language items were represented in the test during the seven years (2057-2063). One language item was neglected while constructing test items. Therefore, the coverage of course contents in terms of test contents in unit five is 75%. 25% of contents was not covered in the question papers. However, "Grammar: Theory and Practice Tests" have high content validity in unit 5 since they have covered more than 70% course contents.

### 3.1.6 Analysis of Content Representativeness (Coverage) in Unit Six

**Table 8: Representation of Test Contents in terms of Course Contents in Unit Six**

S.N.	Course Contents	Test Contents						
	Language Items	Representation of Test Items						
Unit-VI	Grammatical Practice	2057	2058	2059	2060	2061	2062	2063
6.1	Introduction	O <sup>16</sup>						
6.2	Grammatical Metalanguage						O <sup>2</sup>	
6.3	The Lexicon			O <sup>1</sup>				
6.4	The Copula and Sub-Verb Agreement	O <sup>3</sup>			S <sup>6</sup>	O <sup>4</sup> , S <sup>7f</sup>	S <sup>4d</sup>	
6.5	Introduction to Phrase Structure					O <sup>3</sup>	S <sup>4e</sup>	O <sup>2</sup> , O <sup>4</sup> , O <sup>6</sup>
6.6	More Phrase Structure Rules		O <sup>3</sup>				S <sup>4a</sup> L <sup>11a</sup> (opt.)	O <sup>8</sup>
6.7	The Tense and Aspect System							O <sup>16</sup>
6.8	Modal Auxiliaries and Related Phrasal Forms	O <sup>8</sup> , S <sup>6a</sup>	O <sup>11</sup>		S <sup>8</sup>		O <sup>9</sup>	
6.9	The Tense-Aspect Modality System In Discourse				O <sup>3</sup>			
6.10	Negation	S <sup>7b</sup>	L <sup>11a</sup> (opt.)			O <sup>9</sup>	S <sup>4c</sup>	S <sup>3f</sup>
6.11	Yes/No Questions	L <sup>11a</sup> (Opt.)			O <sup>4</sup>		L <sup>11d</sup> (Opt.)	
6.12	Imparatives	L <sup>11c</sup> (Opt.)	O <sup>4</sup> , S <sup>6a</sup> , S <sup>8d</sup> L <sup>11b</sup> (Opt.)		O <sup>8</sup>	O <sup>7</sup> , O <sup>10</sup> , L <sup>11a</sup> (Opt.)	S <sup>5e</sup>	

6.13	Wh-Questions	O <sup>4</sup> , S <sup>6b</sup> , S <sup>7c</sup> , L <sup>11b</sup> (Opt.)	O <sup>5</sup>		L <sup>10c</sup>	O <sup>5</sup> , S <sup>8</sup>	S <sup>8</sup> , L <sup>11b</sup> (Opt.)	
6.14	Other Structures that Look Like Questions	O <sup>12</sup>	S <sup>6b</sup> , L <sup>11c</sup> (Opt)		O <sup>6</sup>	L <sup>11b</sup> (Opt.)		S <sup>3e</sup> , S <sup>6</sup>
6.15	Articles	O <sup>5</sup> , S <sup>6c</sup>		S <sup>7</sup>		O <sup>8</sup>	O <sup>5</sup> , O <sup>13</sup> , S <sup>5f</sup>	
6.16	Reference and Possession							
6.17	Partitives, Collectives and Quantifiers			O <sup>14</sup>	O <sup>12</sup> , O <sup>15</sup>			
6.18	The Passive Voice	O <sup>7</sup> , S <sup>6d</sup>	S <sup>7</sup>			L <sup>11c</sup> (Opt.)	S <sup>5a,b,c,d</sup>	S <sup>3d</sup> , L <sup>11a</sup> (Opt.)
6.19	Sentences With Indirect Objects	S <sup>8d</sup>	S <sup>6d</sup> , S <sup>83</sup> , L <sup>11d</sup> (Opt.)			S <sup>6d</sup> , S <sup>7c</sup>	S <sup>4b</sup>	
6.20	Adjectives			O <sup>13</sup>				
6.21	Prepositions	O <sup>11</sup> , S <sup>8f</sup>	O <sup>10</sup> , S <sup>63</sup> , S <sup>8c</sup>			S <sup>7b</sup>	O <sup>6</sup>	O <sup>12</sup>
6.22	Phrasal Verbs	O <sup>6</sup> , S <sup>6e</sup> , S <sup>8c</sup>		O <sup>2</sup>	O <sup>13</sup>			S <sup>3c</sup> , S <sup>8</sup>
6.23	Non-Referential It and There as Subjects	O <sup>10</sup> , S <sup>8b</sup>	O <sup>12</sup>		S <sup>7</sup>	S <sup>6f</sup>	O <sup>4</sup> O <sup>16</sup>	
6.24	Coordinating Conjunctions	L <sup>11d</sup> (Opt.)	O <sup>9</sup> , S <sup>6f</sup> , S <sup>8a,b</sup>		L <sup>10a</sup>	S <sup>6b,c,e</sup> S <sup>7d,e</sup> , L <sup>11d</sup> (Opt.)		
6.25	Adverbials	S <sup>7a</sup> , S <sup>8e</sup>	O <sup>6</sup> , S <sup>6c</sup>	O <sup>4</sup>				
6.26	Logical Connectors		O <sup>8</sup>	S <sup>6</sup>		S <sup>6a</sup>	O <sup>10</sup>	O <sup>11</sup>
6.27	Conditional Sentences	S <sup>8a</sup>				S <sup>7a</sup>		L <sup>11b</sup> (Opt.)

6.28	Introduction to Relative Clauses	O <sup>9</sup> S <sup>6f</sup>	S <sup>8f</sup>		L <sup>10b</sup>		O <sup>12</sup> L <sup>11c</sup> (Opt.)	
6.29	More on Relative Clauses							
6.30	Focus and Emphasis				S <sup>7</sup>	O <sup>6</sup>	S <sup>4f</sup>	S <sup>3a,b</sup>
6.31	Complementation			O <sup>3</sup>				
6.32	Other Aspects of Complementation and Embedded Clauses							
6.33	Reported Speech and Writing			O <sup>10</sup>				S <sup>11c</sup> (Opt.)
6.34	Degree-Comparatives and Equatives			O <sup>9</sup>				
6.35	Degree-Complements And Superlatives							
6.36	Conclusion							
	Total No. of Asked Questions in Unit Six	O-11 S-3 L-1	O-9 S-3 L-1	O-8 S-2 L-0	O-7 S-3 L-1	O-8 S-3 L-1	O-9 S-3 L-1	O-7 S-3 L-1
	Total Unit Marks: 40	41	39	20	37	38	39	37

As displayed in the above table, there are altogether 36 language items in unit six stretching from 6.1 to 6.36. Within seven years (2057-2063), out of 36 language items, 31 language items were represented in the test. Diachronically looking, among these 36 language items, the most represented ones were 6.10 (Negation), 6.12 (Imperative), 6.13 (Wh-Questions), 6.14 (Other Structures that Look Like Questions), 6.18 (the Passive Voice), 6.21 (Prepositions), 6.23 (Non-referential It and There as Subjects), and 6.26 (Logical Connectors) which were represented in five years out of seven years. Similarly, the language items 6.4 (The Copula

and Subject-Verb Agreement), 6.8 (Modal Auxiliaries and Related Phrasal Forms), 6.15 (Articles), 6.19 (Sentences with Indirect Objects), 6.22 (Phrasal Verbs), 6.24 (Coordinating Conjunctions), 6.28 (Introduction to Relative Clauses), and 6.30 (Focus and Emphasis) were represented in four years; then, the language items 6.5 (Introduction to Phrase Structure), 6.6 (More Phrase Structure Rules), 6.11 (Yes/No Questions), 6.25 (Adverbials), and 6.27 (Conditional Sentences) were represented in three years; then after, the language items 6.17 (Partitives, Collectives and Quantifiers), and 6.33 (Reported Speech and Writing) were represented in two years; and then, the language items 6.1 (Introduction), 6.2 (Grammatical Metalanguage), 6.3 (The Lexicon), 6.7 (The Tense and Aspect System), 6.9 (The Tense-Aspect Modality System in Discourse), 6.20 (Adjectives), 6.31 (Complementation), and 6.34 (Degree-Comparatives and Equatives) were represented only in one year out of seven years. After all, the other remaining language items which were not represented in any year are 6.16 (Reference and Possession), 6.29 (More on Relative Clauses), 6.32 (Other Aspects of Complementation and Embedded Clauses), 6.35 (Degree-Complements and Superlatives), and 6.36 (Conclusion).

Looking at the data synchronically, eleven objective questions, three subjective short questions and one subjective long question were asked from this unit in 2057. In case of objective questions, the first (O<sup>16</sup>) from the language item 6.1 (Introduction), the second (O<sup>3</sup>) from 6.4 (The Copula and Subject-Verb Agreement), the third (O<sup>8</sup>) from 6.8 (Modal Auxiliaries and Related Phrasal Forms), the fourth (O<sup>4</sup>) from 6.13 (Wh-Questions), the fifth (O<sup>12</sup>) from 6.14 (Other Structures that Look Like Questions), the sixth (O<sup>5</sup>) from 6.15 (Articles), the seventh (O<sup>7</sup>) from 6.18 (The Passive Voice), the eighth (O<sup>11</sup>) from 6.21 (Prepositions),

the ninth (O<sup>6</sup>) from 6.22 (Phrasal Verbs), the tenth (O<sup>10</sup>) from 6.23 (Non-Referential It and There as Subjects) and the eleventh (O<sup>9</sup>) from 6.28 (Introduction to Relative Clauses) were asked. Then, in case of subjective short questions, the first question (S<sup>6</sup>) was collectively asked as, 'S<sup>6a</sup>' from 6.8 (Modal Auxiliaries and Related Phrasal Forms), 'S<sup>6b</sup>' from 6.13 (Wh-Questions), 'S<sup>6c</sup>' from 6.14 (Articles), 'S<sup>6d</sup>' from 6.18 (The Passive Voice), 'S<sup>6e</sup>' from 6.22 (Phrasal Verbs), and 'S<sup>6f</sup>' from 6.28 (Introduction to Relative Clauses); the second question 'S<sup>7</sup>' (i.e. S<sup>7a,b</sup> and <sup>c</sup>) was collectively asked from the language items 6.25 (Adverbials), 6.10 (Negation), and 6.13 (Wh-Questions) respectively, and the third question 'S<sup>8</sup>' (i.e. S<sup>8a,b,c,d, e</sup> and <sup>f</sup>) was collectively asked from the language items 6.27 (Conditional Sentences), 6.23 (Non-referential It and There as Subjects), 6.22 (Phrasal Verbs), 6.19 (Sentences with Indirect Objects), 6.25 (Adverbials), and 6.21 (Prepositions) respectively. Then, in case of subjective long question, only one long question (L<sup>11</sup>) was integratively asked as, 'L<sup>11a</sup>' from 6.11 (Yes/No Questions), 'L<sup>11b</sup>' from 6.13 (Wh-Questions), 'L<sup>11c</sup>' from 6.12 (Imperatives), and 'L<sup>11d</sup>' from 6.24 (Coordinating Conjunctions). And all these sub-questions (i.e. L<sup>11a</sup>, L<sup>11b</sup>, L<sup>11c</sup>, and L<sup>11d</sup>) under the main question (L<sup>11</sup>) were asked as optional questions. It means, any two of these four sub-questions were asked to be attempted.

In 2058, nine objective questions, three subjective short questions and one subjective long question were asked from this unit. In case of objective questions, the first (O<sup>3</sup>) from 6.6 (More Phrase Structure Rules) the second (O<sup>11</sup>) from 6.8 (Modal Auxiliaries and Related Phrasal Forms), the third (O<sup>4</sup>) from 6.12 (Imperatives), the fourth (O<sup>5</sup>) from 6.13 (Wh-Questions), the fifth (O<sup>10</sup>) from 6.21 (Prepositions), the sixth (O<sup>12</sup>) from 6.23 (Non-referential It and There as Subjects), the seventh (O<sup>9</sup>)

from 6.24 (Coordinating Conjunction), the eighth (O<sup>6</sup>) from 6.25 (Adverbials), and the ninth (O<sup>8</sup>) from 6.26 (Logical Connectors) were asked. Then, in case of subjective short questions, the first question (S<sup>6</sup>) was collectively asked as, 'S<sup>6a</sup>' from 6.12 (Imperatives), 'S<sup>6b</sup>' from 6.14 (Other Structures that Look Like Questions), 'S<sup>6c</sup>' from 6.25 (Adverbials), 'S<sup>6d</sup>' from 6.19 (Sentences with Indirect Objects), 'S<sup>6e</sup>' from 6.21 (Prepositions) and 'S<sup>6f</sup>' from 6.24 (Coordinating Conjunctions); the second question (S<sup>7</sup>) was asked from 6.18 (The Passive Voice), and the third one 'S<sup>8</sup>' (i.e. S<sup>8a</sup> and b, c, d, e and f) was collectively asked from the language items 6.24 (Coordinating Conjunctions), 6.21 (Prepositions), 6.12 (Imperatives), 6.19 (Sentences with Indirect Objects), and 6.28 (Introduction to Relative Clauses) respectively. Then, in case of subjective long question, only one long question (L<sup>11</sup>) was collectively asked as, 'L<sup>11a</sup>' from 6.10 (Negation), 'L<sup>11b</sup>' from 6.12 (Imperatives), 'L<sup>11c</sup>' from 6.14 (Other Structures that Look Like Questions), and 'L<sup>11d</sup>' from 6.19 (Sentences with Indirect Objects). And all these sub-questions (i.e. L<sup>11a</sup>, L<sup>11b</sup>, L<sup>11c</sup>, and L<sup>11d</sup>) under the main question (L<sup>11</sup>) were asked as optional questions. It means, any two of these four sub-questions were asked to be answered.

In 2059, eight objective questions and two subjective short questions were asked from this unit. In case of objective questions, the first question (O<sup>1</sup>) from 6.3 (The Lexicon), the second (O<sup>14</sup>) from 6.17 (Partitives, Collectives and Quantifiers), the third (O<sup>13</sup>) from 6.20 (Adjectives), the fourth (O<sup>2</sup>) from 6.22 (Phrasal Verbs), the fifth (O<sup>4</sup>) from 6.25 (Adverbials), the sixth (O<sup>3</sup>) from 6.31 (Complementation), the seventh (O<sup>10</sup>) from 6.33 (Reported Speech and Writing), and the eighth (O<sup>9</sup>) from 6.34 (Degree-Comparatives and Equatives) were asked. Then, in the case of subjective short questions, the first question (S<sup>6</sup>) was asked

from 6.26 (Logical Connectors), and the next question (S<sup>7</sup>) was asked from 6.15 (Articles). Then, there was no representation of long question from this unit in 2059.

Similarly, in 2060, seven objective questions, three subjective short questions, and one long question were asked from this unit. Regarding objective questions, the first (O<sup>3</sup>) from 6.9 (The Tense-Aspect Modality System in Discourse), the second (O<sup>4</sup>) from 6.11 (Yes/No Questions), the third (O<sup>8</sup>) from 6.12 (Imperatives), the fourth (O<sup>6</sup>) from 6.14 (Other Structures that look like questions), both the fifth (O<sup>12</sup>) and the sixth (O<sup>15</sup>) from the same language item 6.17 (Partitives, Collectives, and Quantifiers), and the seventh (O<sup>13</sup>) from 6.22 (Phrasal Verbs) were asked. Then, regarding subjective short questions, the first question (S<sup>6</sup>) was asked from 6.4 (The Copula and Subject-Verb Agreement) the second one (S<sup>7</sup>) was integratively asked from 6.23 (Non-referential It and There as Subjects) and 6.30 (Focus and Emphasis), and the third one (S<sup>8</sup>) was asked from 6.8 (Modal Auxiliaries and Related Phrasal Forms). Then after, regarding long question, only one long question (L<sup>10</sup>) was integratively asked as, 'L<sup>10a</sup>' from 6.24 (Coordinating Conjunctions), 'L<sup>10b</sup>' from 6.28 (Introduction to Relative Clauses), and 'L<sup>10c</sup>' from 6.13 (Wh-Questions).

In 2061, eight objective questions, three subjective short questions and one long question were asked from this unit. In case of objective questions, the first question (O<sup>4</sup>) from 6.4 (The Copula and Subject-Verb Agreement), the second one (O<sup>3</sup>) from 6.5 (Introduction to Phrase Structure), the third (O<sup>9</sup>) from 6.10 (Negation), both the fourth (O<sup>7</sup>) and the fifth (O<sup>10</sup>) from the same language item 6.12 (Imperatives), the sixth (O<sup>5</sup>) from 6.13 (Wh-questions), the seventh (O<sup>8</sup>) from 6.15 (Articles), and the eighth (O<sup>6</sup>) from 6.30 (Focus and Emphasis) were asked. Then, in case of subjective short questions, the first question (S<sup>6</sup>) was collectively



asked as, 'S<sup>6a</sup>' was asked from 6.26 (Logical Connectors), S<sup>'6b,c</sup> and <sup>e</sup>' were asked from the same language item 6.24 (Coordinating Conjunctions), 'S<sup>6d</sup>' was asked from 6.19 (Sentences with Indirect Objects), and 'S<sup>6f</sup>' was asked from 6.23 (Non-referential It and There as Subjects); second question (S<sup>7</sup>) was integratively asked as, 'S<sup>7a</sup>' from 6.27 (Conditional sentences), 'S<sup>7b</sup>' from 6.21 (Prepositions), 'S<sup>7c</sup>' from 6.19 (Sentences with Indirect Objects), 'S<sup>7d</sup>' and <sup>e</sup>' both from 6.24 (Coordinating Conjunctions), and 'S<sup>7f</sup>' from 6.4 (The Copula and Subject-verb Agreement), and the third one (S<sup>8</sup>) was asked from 6.13 (Wh-Questions). Then, in case of subjective long question, only one long question (L<sup>11</sup>) was collectively asked as, 'L<sup>11a</sup>' from 6.12 (Imperatives), 'L<sup>11b</sup>' from 6.14 (Other Structures that Look Like Questions), 'L<sup>11c</sup>' from 6.18 (The Passive Voice), and 'L<sup>11d</sup>' from 6.24 (Coordinating Conjunctions). And all these sub-questions (L<sup>11a</sup>, L<sup>11b</sup>, L<sup>11c</sup> and L<sup>11d</sup>) under the main question (L<sup>11</sup>) were asked as optional questions. It means, any three out of these four sub-questions were asked to be answered.

Likewise, in 2062, nine objective questions, three subjective short questions, and only one subjective long question were asked from this unit. In case of objective questions, the first (O<sup>2</sup>) from 6.2 (Grammatical Metalanguage), the second (O<sup>9</sup>) from 6.8 (Modal Auxiliaries and Related Phrasal Forms), both the third (O<sup>5</sup>) and the fourth (O<sup>13</sup>) from the same language item 6.15 (Articles), the fifth (O<sup>6</sup>) from 6.21 (Prepositions), both the sixth (O<sup>4</sup>) and the seventh (O<sup>16</sup>) from the same language item 6.23 (Non-referential It and There as Subjects), the eighth (O<sup>10</sup>) from 6.26 (Logical Connectors), and the ninth (O<sup>12</sup>) from 6.28 (Introduction to Relative Clauses) were asked. Then, in case of subjective short questions, the first (S<sup>4</sup>) was collectively asked as, 'S<sup>4a</sup>' from 6.6 (More Phrase Structure Rules), 'S<sup>4b</sup>' from 6.19 (Sentences with Indirect Objects), 'S<sup>4c</sup>' from 6.10 (Negation), 'S<sup>4d</sup>' from 6.4 (The Copula and Subject-Verb Agreement), 'S<sup>4e</sup>' from 6.5 (Introduction to Phrase structure), and 'S<sup>4f</sup>'

from 6.30 (Focus and Emphasis); second ( $S^5$ ) was collectively asked as, ' $S^{5a, b, c}$  and  $d$ ' were asked from the same language item 6.18 (The Passive Voice), ' $S^{5e}$ ' was asked from 6.12 (Imperatives), and ' $S^{5f}$ ' was asked from 6.15 (Articles), and the third ( $S^8$ ) was asked from 6.13 (Wh-questions). Then, in case of subjective long question, only one long question ( $L^{11}$ ) was collectively asked as, ' $L^{11a}$ ' from 6.6 (More Phrase Structure Rules), ' $L^{11b}$ ' from 6.13 (Wh-Questions), ' $L^{11c}$ ' from 6.28 (Introduction to Relative Clauses), and ' $L^{11d}$ ' from 6.11 (Yes/No Questions). And all these sub-questions ( $L^{11a}$ ,  $L^{11b}$ ,  $L^{11c}$ , and  $L^{11d}$ ) under the main question ( $L^{11}$ ) were asked as optional questions. It means, any three out of these four sub-questions were asked to be answered.

Then, in 2063, seven objective questions, three subjective short questions and only one subjective long question were asked from this unit. In case of objective questions, all the three first ( $O^2$ ), second ( $O^4$ ) and third ( $O^6$ ) from the same language item 6.5 (Introduction to phrase structure), the fourth ( $O^8$ ) from 6.6 (More Phrase Structure Rules), the fifth ( $O^{16}$ ) from 6.7 (The Tense and Aspect System), the Sixth ( $O^{12}$ ) from 6.21 (Prepositions), and the seventh ( $O^{11}$ ) from 6.26 (Logical Connectors). Then, in case of subjective short questions, the first ( $S^3$ ) was integratively asked as, both ' $S^{3a}$  and  $b$ ' from 6.30 (Focus and Emphasis), ' $S^{3c}$ ' from 6.22 (Phrasal Verbs), ' $S^{3d}$ ' from 6.18 (The Passive Voice), ' $S^{3e}$ ' from 6.14 (Other Structures that Look Like Questions), and ' $S^{3f}$ ' from 6.10 (Negation); Second ( $S^6$ ) was asked from 6.14 (Other Structures that Look Like Questions), and the third ( $S^8$ ) was asked from 6.22 (Phrasal Verbs). Then, in case of subjective long question, only one long question ( $L^{11}$ ) was collectively asked as, ' $L^{11a}$ ' from 6.18 (The Passive Voice), ' $L^{11b}$ ' from 6.27 (Conditional Sentences), and ' $L^{11c}$ ' from 6.33 (Reported Speech and Writing). And all these sub-questions ( $L^{11a}$ ,  $L^{11b}$ , and  $L^{11c}$ ) under the main compulsory question ( $L^{11}$ ) were asked as optional questions. It

means, any two out of these three sub-questions were asked to the answered.

### **Repeated Items**

After analyzing the above data, the researcher also found that from unit six, within seven years (2057-2063), only four objective questions and ten subjective short questions (Partial Questions) were repeated. In case of objective questions, 2057<sup>O3</sup> and 2061<sup>O4</sup> were same except some terminological difference. And both of them were from the same language item 6.4 (The Copula and Subject-Verb Agreement). Similarly, 2060<sup>O8</sup> and 2061<sup>O10</sup> were also same except some terminological difference. They were from the same language item 6.12 (Imperatives). Then, in case of subjective short questions, no full or whole subjective questions for example S<sup>6</sup>, but some parts of the whole or full question (i.e. partial short questions), for example S<sup>6d</sup>, were repeated as mentioned below. 2062<sup>S4c</sup> and 2063<sup>S3f</sup> were exactly same from 6.10 (Negation) 2059<sup>S7a</sup> and 2062<sup>S5f</sup> were also exactly same from the same language item 6.15 (Articles). Similarly, 2057<sup>S6d</sup> and 2062<sup>S5c</sup> were same from 6.18 (The Passive Voice) except some terminological difference. Then, 2058<sup>S6d</sup> and 2061<sup>S6d</sup> were exactly same from the same language item 6.19 (Sentence with Indirect Objects). Then after, 2062<sup>S4f</sup> and 2063<sup>S3a</sup> were exactly same from 6.30 (Focus and Emphasis). In addition to this, regarding objective questions and subjective short questions (Partial Questions), there was some sort of overlapping or repetition between some of these questions, i.e. 2062<sup>S4d</sup> was same as 2057<sup>O3</sup> and 2061<sup>O4</sup> except some terminological differences. They all (S<sup>4d</sup>, O<sup>3</sup> and O<sup>4</sup>) were from the same language item 6.4 (The Copula and Subject-Verb Agreement). Then, 2058<sup>S6a</sup> was same as 2060<sup>O8</sup> and 2061<sup>O10</sup> except some terminological difference. They all (S<sup>6a</sup>, O<sup>8</sup> and O<sup>10</sup>) were from the same language item 6.12 (Imperatives). Then after, 2058<sup>S6b</sup> and 2060<sup>O6</sup> were same from 6.14 (Other Structures that Look Like Questions) except some terminological difference, and then 2061<sup>O6</sup> was same as 2062<sup>S4f</sup> and 2063<sup>S3a</sup> except some terminological

difference. They all ( $O^6$ ,  $S4^f$  and  $S3a$ ) were from the same language item 6.30 (Focus and Emphasis).

After all, there was no repetition of the same subjective long question during seven years (2057-2063) from this until but it doesn't mean that there was no repetition of the model of long question as well. Putting it in the another way, there was repetition of the models of both subjective long and short questions but not the repetition of the same subjective questions especially long question. However, the above frequency of repetition of the test items in this very large unit is not more than enough. It is O.K. or adequate in terms of repetition because some of the test items should also be repeated otherwise they may be neglected by the testees.

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

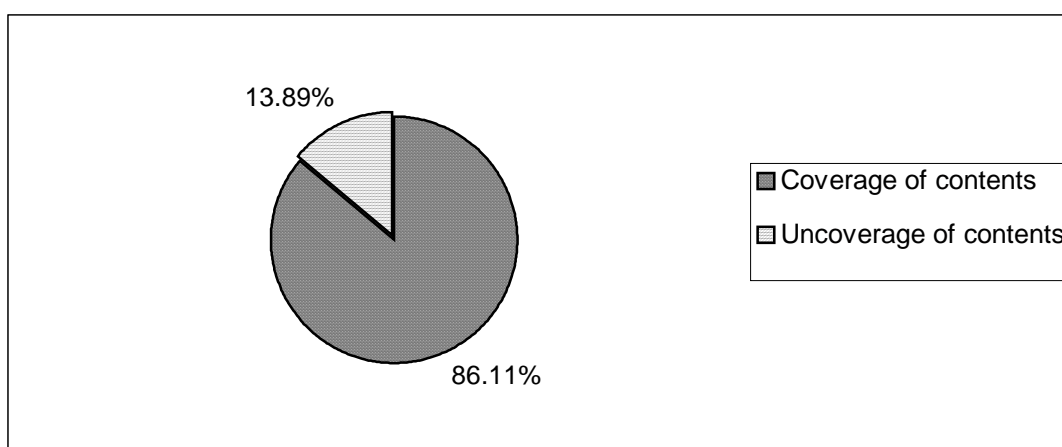
**Figure 6: Content Coverage from Unit Six in Percentage**

Total coverage of contents = 31

Total uncoverage of contents = 5

Total areas of contents = 36

$\therefore$  Content coverage in percentage =  $31/36 \times 100 = 86.11\%$



From the above table and description, it is clear that there are 36 language items in unit six from 6.1 to 6.36 according to course contents.

Out of them, 31 language items were represented in the test during the seven years (2057-2063) but five language items were discarded while constructing test items. Therefore, the coverage of course contents in terms of test contents in unit six is 86.11%. 13.89% of the contents was not covered in the test papers. However, "Grammar: Theory and Practice Tests" have high content validity in unit six as whole since they have covered more than 70% course contents but if we see the content coverage (here content validity) of unit six yearwise, then the content coverage of it in 2059 was very low in terms of test items represented in the test of this year.

### **3.1.7 Analysis of Content Validity of the Test Papers as a Whole In Terms of Content Representativeness (Coverage)**

**Table 9: Representation of Test Contents in terms of Course Contents as a Whole**

S.N.	Units	Course contents	Test contents	Test coverage in percentage
		Total language items	Represented language items	
1	1	3	3	100
2	2	15	13	86.66
3	3	25	24	96.00
4	4	12	11	91.66
5	5	4	3	75.00
6	6	36	31	86.11
Total		95	85	89.47

As shown in the above table, there are altogether 95 language items as a whole within the six units of the "Grammar: Theory and Practice (Eng. Ed. 512)." Out of 95 language items, 85 language items were represented in the test during seven years (2057-2063) and only 10

language items were neglected while constructing the test items. Therefore, the coverage of course contents in terms of represented language items in the test as a whole is 89.47%. It means, the content coverage of "Grammar: Theory and Practice Test Papers" is 89.47% in aggregation. 10.53% of the course contents was not covered in the test papers.

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

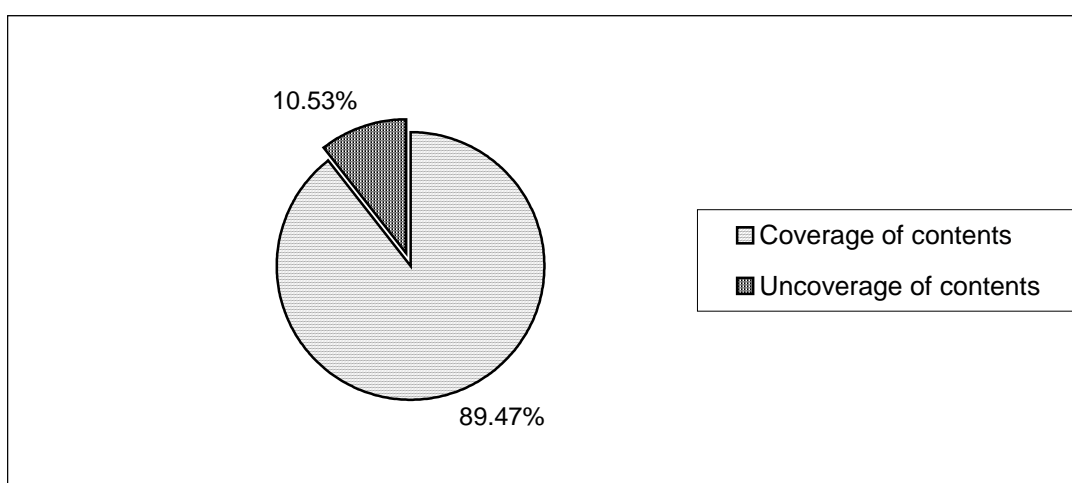
**Figure 7: Content Coverage of Grammar: Theory and Practice Test Papers as a Whole in Percentage**

Total coverage of contents = 85

Total uncoverage of contents = 10

Total areas of contents = 95

∴ Content coverage in percentage =  $85/95 \times 100 = 89.47\%$



After analyzing the above table, the researcher has concluded that regarding content coverage, "Grammar: Theory and Practice Test Papers" as a whole, have very good content validity since they have covered more than 70% course contents (i.e. around 90%). Therefore, "Grammar: Theory and Practice Tests" have really covered what they have been supposed to cover from the course contents. Thus, "Grammar: Theory

and Practice Test Papers" have high content validity in terms of coverage during seven years (2057-2063).

### **3.2 Analysis of Content Relevance**

This is the second part of this chapter which deals with content relevance. As we know, in addition to content coverage, content relevance is another guideline for examining content validity of the test contents (i.e. test items) to know that whether the test tasks included in an exam are relevant to the language activities that are expected to be exercised under the given course or not. So, content relevance means the test items should be congruent with the course objectives that are supposed to be fulfilled.

As there is not a single fixed reliable definition of content relevance, there is not a single universal method or criteria for the purpose of examining content relevance of the test items in question. Therefore, there are some alternative ways or methods of examining content relevance of the test items to examine content validity of those items, such as,

The first way of examining content relevance of the test items is to analyze the percentage or degree of content coverage since the more the test items are constructed, the more content coverage the test paper has and in turn, the more the content coverage the test paper has the more the content relevance of a test.

The next way of examining content relevance of the test items is to compare 'the specification of the behavioural domain in question and the attendant specification of the tasks or test domain' (Messick 1980: 1017 as quoted in Bachman 1998: 244).

The third or last way of examining content relevance of the test items is to examine 'whether the test tasks included in an exam are relevant to the language activities that are expected to be exercised under the given course or not' (Khaniya 2005: 104). It means, content relevance can be examined by examining whether the test items to tasks are congruent with the course objectives or not since the language activities are based on course objectives. Moreover, 'content relevance also requires the specification of the test method facet' (Bachman 1998: 244) which is integrated with testing the test items and plays vital role to make the test items and course objectives congruent to each other because test method facet emphasizes on 'how aspect', i.e. how the test items were constructed whether the test items require the testees to demonstrate the communicative language ability or linguistic competence. So, it depends on the way or method of asking questions.

Regarding these above three methods, according to first, content relevance is examined by analyzing the percentage or degree of content coverage. In terms of coverage, there is no doubt that the test items have high content relevance as they have high content coverage (i.e. 89.47%). But the researcher didn't take it as a main method for examining content relevance thinking that it was not reliable since the test items included in the test might not be congruent with the course objectives.

Regarding the second, content relevance is examined by comparing the specification of the behavioural domain and the specification of the task or test domain. But there is no provision of making specification in higher levels (i.e. above SLC). So, the researcher didn't take it too as a main method for examining content relevance.

Then, regarding the third or last one, content relevance is examined by examining whether the test items are congruent with the course



objectives or not. For this, both test items and course objectives were available on the one hand and this method could be suitable for examining content relevance logically for the researcher in question on the other hand. So, among these three, the researcher selected it as a main method for examining content relevance of the test items thinking that it was more convenient and reliable for the concerned purpose.

To sum up, for the purpose of examining content validity of "Grammar: Theory and Practice (Eng. Ed. 512) Test Papers" in terms of content relevance in question, the researcher went especially through the Eng. Ed. 512 subjective test tasks of seven years (2057-2063) and examined them in relation to the course objectives. Moreover, while examining the content relevance of the test items or tasks, the researcher neither analyzed it unitwise since there are no unitwise objectives nor he analyzed it yearwise but as a whole thinking that it would be better to analyze as a whole rather than the repetition of the same flavour because the nature of the questions in each year is almost same.

### **Analysis of Content Relevance of the Test Items of Grammar: Theory and Practice (Eng. Ed. 512) in Relation to Course Objectives as a Whole**

As the title suggests, the researcher went through the test items of "Grammar: Theory and Practice (Eng. Ed. 512)" from the year 2057 to 2063 as a whole and examined the content relevance of those test items that whether the test items were congruent with the course objectives or not or he examined whether the test tasks included in the test papers of Eng. Ed. 512 were relevant to the language activities that were expected to be exercised under the same course or not. The researcher used only descriptive method (i.e. orthographic text) since content relevance is a logical or rational property of a test.

As mentioned in chapter-1, Grammar: Theory and Practice (Eng. Ed. 512) course has only two prescribed course objectives (i.e. to develop insights into formal aspects of language with special emphasis on structural and generative models and to develop insights into English Grammar and its pedagogic application and implications) and six units altogether. The researcher, here, examined all the test items of seven years (2057-2063) from those six units as a whole in relation to the course objectives. Regarding the nature of the course objectives, the first is concerned with the theory part and the next one is concerned with the practice part of the course. Therefore, before analyzing the test tasks, the researcher studied the first objective very deeply and analyzed it from different possible angles to be clear about its thrust, emphasis, demand, coverage, purpose, required skill to be developed, etc., then he started to study all the test items which were related with this first objective (i.e. the questions relevant to the theory part) and analyzed as well as categorized them from the different similar possible angles, such as, their thrust, emphasis, demands, purposes, required skills to be developed, weightage, levels and nature of the questions to examine whether the test tasks were congruent with the objectives or not. Regarding the criteria, if both the objective and test tasks have same purpose, thrust, required skills to be developed, nature in question etc., then they are said to be relevant to each other.

After studying and analyzing the concerned questions, the researcher found out that only about one third (i.e. 22/56) of the questions (2057-2063) had same demands (i.e. nature, skills to be developed, purpose, level of the knowledge, etc.) which the first objective had. For example, the aim or demand of the course objective is to develop insights into formal aspects of language with special emphasis on structural and

generative models. It means, it aims to develop the deep or ground knowledge (i.e. interpretative skill) on the testees to prepare them as competent manpower on theoretical aspects of language especially emphasis with structural and generative models to tackle in question. Similarly, the aim or demand of 2057<sup>L9</sup> (i.e. what is TG Grammar? In what ways is it an improvement on the Structuralists' view of Grammar? Explain.) is to test or examine the deep or ground knowledge on structural and generative models of grammar. As it requires interpretative skill to fulfil the demand of the question, the testees need to interpret what the Structural Grammar is, what the TG Grammar is, what the core points, characteristics or salient features of structural and TG models of grammar are, and how the TG Grammar is an improvement or developed model of Structural Grammar including the testees' cognitive strategies as well. So, regarding their demands, we can claim that the test item (2057<sup>L9</sup>) was congruent with the first course objective. It is because both of them demanded interpretative skill. Moreover, the following test tasks are some of the other examples of content relevance which are congruent with the same objective:

2058<sup>L10</sup> (Opt.) → List and illustrate different types of cases recognized by C.J. Fillmore in "Case Grammar." In what ways does Fillmore's concept of the deep structure differ from that of Chomsky? Illustrate your answer,

2059<sup>L9</sup> (Opt.) → Given a construction "Josh ate the Pizza naked", present your argument with the help of several tests to establish the fact that the Adjectival Phrase (AP) 'naked' in the construction is a part of verb phrase not an immediate constituent of sentence. Argue if the AP is sister of bar level category or zero level,

2061<sup>S3</sup> → Explain the limitations of IC analysis. Draw a tree diagram for the following structure and point out the problem, if any.

"Flying planes can be dangerous."

2062<sup>S3</sup> → How does Relational Grammar treat an unemployed NP? Illustrate, etc.

As mentioned above, the researcher also found that most of the questions or test items from the theory part were not congruent with the concerned objective because of different factors, such as, low quality of education, carelessness of test designer, lack of skilled administrator and exam-controller, weaknesses of the syllabus or course of study itself, lack of sufficient course objectives, and so on. Some of them, for example, are as follows:

2058<sup>S2</sup> → Describe the characteristic features of pedagogic Grammar,

2059<sup>S2</sup> → Describe any two of Nida's principles for the identification of morpheme. Give illustrative examples,

2061<sup>S4</sup> → What are Chomsky's notions of observational, descriptive and explanatory adequacies of grammar? Explain briefly,

2063<sup>S1</sup> → Explain and illustrate various types of allomorphic alternations, etc.

Here, the demand of the concerned objective was not same with the demands of those above test items so those test items were not relevant to the objective. Explicitly speaking, the objective aimed to develop deep or ground level knowledge (i.e. interpretative skill) on the testees to tackle the problems but these questions aimed to test just the surface knowledge of something (i.e. vomiting skill) since surface knowledge of the testees is enough to answer those questions, and this reason may be applicable for examining content relevance of rest of the similar test items. Thus, the

researcher examined the content relevance of the test items under the theory part in relation to the concerned course objective.

Then after, following the same path, the researcher, first of all, studied the second or next objective very deeply and analyzed it from different possible perspectives as similar as the first one. Then, he moved towards the test items which were related with this second objective (i.e. the questions relevant to the practice part) and analysed as well as categorized them from the possible perspectives, such as, their thrust, emphasis, demands, purposes, required skills to be developed, level, and nature of the questions to examine whether the test items were relevant with the objective or not. Regarding the criteria, if both the objective and test tasks share the same purposes, thrust, required skills to be developed, nature in question, etc., then they are said to be relevant to each other. When the researcher studied, analyzed and categorized the concerned questions, the researcher found out that all (i.e. 26/26) the questions (2057-2063) had same demands (i.e. nature, emphasis, skills to be developed, purposes, level of the knowledge, etc.) which the second objective had. For example, the aim or demand of the course objective is to develop insights into English Grammar and its pedagogic applications and implications. It means, it aims to develop grammatical rules or usage or pedagogic applications (i.e. *interpretative skill or grammatical competence*) into English Grammar and also skills in their use, practice or pedagogic implications on the testees to provide them *pedagogic skills or grammatical performance*. Then, the aim or demand of 2057<sup>S8</sup> [i.e. why are the following sentences ungrammatical (or the best awkward) ?]: (a) If she had been there, she did the work (b) There is Santa Claus (c) After two hours the candle has burned off (d) I gave John a book, and a bicycle to Bill (e) John has hardly gone to bed after midnight (f) I served the

Army until 1964] is to test the knowledge of grammatical competence and performance. As it requires *interpretative skill and pedagogic skills* to fulfil the demand of the question, the testees need to interpret why the following sentences are ungrammatical, what and where the wrongs are, what the correct answers are and how the sentences can be corrected, etc. So, regarding their demands in question, we can claim that the test task (2057<sup>S8</sup>) was congruent with its objective. It is because both of them demanded the same skills i.e. *interpretative and pedagogic skills*.

In the above example, the same test item could fulfil all the three demands (i.e. insight, pedagogical applications and implications) of the concerned objective. But it was also found that more than one test items may integratively fulfil the demands of the same objective. It means, out of three demands, one of the demands was fulfilled by one test item and the remaining ones were fulfilled by the next test item. For example, one demand (i.e. insight) was fulfilled by 2062<sup>S4</sup> and the other two (i.e. pedagogical applications and implications) were fulfilled by 2062<sup>S5</sup>. Anyway, all the test items during seven years (2057-2063) from the practice part were congruent with their objective. Some of them, for example, are 2058<sup>S8</sup>, 2058<sup>L11</sup>, 2059<sup>S6</sup>, 2061<sup>S7</sup> and so on. Thus, the researcher examined the content relevance of test items under the practice part too in relation to the concerned course objective.

In conclusion, the researcher, here, studied and analyzed the test items of "Grammar: Theory and Practice (Eng. Ed. 512)" during seven years (2057-2063) in relation to the course objectives to examine the content relevance of these test items. While studying and analyzing them, the researcher studied and analyzed according to the objectives and their respected fields (i.e. parts: theory and practice). It means, he studied and analyzed the subject matter (i.e. test items) partwise since the first

objective and second objective were related with the theory part and practice part respectively, and concluded in this way:

First of all, the researcher studied and analyzed all the test items from the theory part of Eng. Ed. 512 and concluded that the content relevance of those test items was low because only around 39% (i.e. 39.28%) test items of this part were relevant to the objective. So, the test items from theory part were *partially* relevant to the given objective. This was because of different factors, such as, low quality of education, carelessness of test designer, lack of skilled administrator and exam controller, lack of specification grid, weaknesses of the syllabus or course of study itself as both *Generative Grammar and other Models of Grammar* from theory part of this course were not designed especially for pedagogic purposes but for linguistic purposes so it was very difficult to fulfil the objective, lack of sufficient course objectives (i.e. 1 for 5 units), and so on. Therefore, if we improved the above weaknesses in question, the content relevance of the test items would be high.

Then, the researcher studied and analyzed all the test items from the practice part of Eng. Ed. 512 and concluded that the content relevance of the test items was very high because 100% test items of this part were relevant to the objective. So, the test items from practice part were *fully* relevant to the given objective. Moreover, most of the test items from this part have emphasized the same demands (i.e. insight, pedagogic applications and implications) of the objective and they also have the same nature. So, the researcher used single test item (in detail) as model test item.

## **Analysis of Content Validity of Test Items as a Whole in Terms of Content Relevance**

After studying and analyzing the test items of Eng. Ed. 512 partwise considering the nature of the course, its objectives and the sake of the convenient of the researcher as well, the researcher also studied and examined the test items as a whole (i.e. as a single part of Eng. Ed. 512). In this regard, the researcher combined or mixed them together thinking that the purpose or objective of this research was not to examine the content relevance of test items separately (i.e. partwise or objectivewise) but to examine the content validity of the test items of "Grammar: Theory and Practice" as a whole in terms of content relevance and concluded that the content relevance of test items was actually low but logically considerable since just around 60% (i.e. 58.53%) test items of Eng. Ed. 512 were relevant to the objectives. Thus, the test items of "Grammar: Theory and Practice" had actually low but logically nearer to content validity in terms of relevance during the seven years (2057-2063).

### **3.3 An Overview: On the High Content Validity in Terms of Coverage vs. Low Content Validity in Terms of Relevance in the Grammar: Theory and Practice Tests**

It was found that the test papers of "Grammar: Theory and Practice (Eng. Ed. 512)" obtained high content validity in terms of representativeness (coverage) because out of 95 language items, 85 language items were represented in the test, i.e. 89.47%. But the same test papers or test items obtained low content validity in terms of relevance showing only 58.53%. It means, 58.53% test items were relevant to the course objectives and rest of the other were irrelevant. This shows that the test papers having high content validity in terms of coverage do not necessarily have high content validity in terms of relevance and vice-versa.



## CHAPTER IV

### FINDINGS AND RECOMMENDATIONS

This chapter deals with the findings and recommendations of the study. After the analysis of the test papers (test items) from the different angles, the researcher has found the following findings:

#### 4.1 Findings

The general findings of this research are presented below:

Regarding the representative (coverage) principle, the "Grammar: Theory and Practice Tests" obtained high content validity (content coverage) as a whole because out of 95 language items, 85 language items were represented in the test during the seven years (2057-2063), i.e. 89.47%. It was also found that among these 85, the language item 4.4 (Case Grammar) was the most represented by 13 test items in every year's examination from 2057 to 2063.

On the other hand, regarding the relevance principle, the "Grammar: Theory and Practice Tests" obtained actually low but logically nearer to content validity since just around 60% (i.e. 58.53%) test items of Eng. Ed. 512 were relevant to the course objectives during the seven years (2057-2063).

The specific (i.e. unitwise and partwise) findings of this research are as follows:

##### 1. In terms of representativeness (coverage):

- i. In unit one, there were three language items in course contents and all of them were represented in the test during seven years (2057-2063). Therefore, the coverage of course contents in terms of test contents is 100%. That is why, the content validity of the test papers of "Grammar: Theory and Practice" is very

high in unit one since it has covered more than 70% course contents.

- ii. In unit two, out of 15 language items of the course contents, 13 language items were represented in the tests during seven years but 2 language items were discarded while constructing test items. Therefore, the coverage of course contents in terms of test contents is 86.66%. However, "Grammar: Theory and Practice Tests" have high content validity in unit two as a whole since they have covered more than 70% course contents but if we see the content coverage of unit two yearwise, then the content coverage of it in 2058 was very high and in 2060 was very low.
- iii. In unit three, out of 25 language items of the course contents, 24 language items were represented in the tests during seven years but 1 language item was untouched. It means, the coverage of course contents is 96%. So, "Grammar: Theory and Practice Tests" have high content validity in unit three.
- iv. In unit four, there were 12 language items in course contents. Out of them, test items represented eleven language items during seven years. Only one language item was untouched. Therefore, the coverage of course contents is 91.66. That is to say, the content validity of the test papers of "Grammar: Theory and Practice" is high in unit four.
- v. In unit five, out of 4 language items, three language items were represented in the tests during the seven years. One language item was untouched. It means, the coverage of course contents

is 75%. Therefore, "Grammar: Theory and Practice Tests" have high content validity in unit five.

- vi. In unit six, there were 36 language items in the course contents. Out of them, 31 language items were represented in the tests during seven years but five language items were discarded while constructing test items. Therefore, the coverage of course contents in terms of test contents is 86.11%. However, "Grammar: Theory and Practice Tests" have high content validity in unit six as a whole since they have covered more than 70% course contents but if we see the content coverage (here content validity) of unit six yearwise, then the content coverage of it in 2059 was very low in terms of test items represented in the test.

## **2. In terms of relevance:**

- i) The content relevance of the test items from *theory part* of Eng. Ed . 512 was low because only around 39% test items of this part were relevant to the given objective. So, the test items from *theory part* were *partially* relevant to the given objective. This was because of different factors, such as, low equality of education, carelessness of test designer, lack of skilled administrator and exam controller, lack of specification grid, weaknesses of the syllabus or course of study itself as both *Generative Grammar and Other Models of Grammar* from theory part of this course were not designed especially for pedagogic purposes but for linguistic purposes so it was very different to fulfil the objective, lack of sufficient objectives (i.e. 1 for 5 units), and so on.
- ii) The content relevance of the test items from *practice part* of Eng. Ed. 512 was very high since 100% test items of this part were relevant to

the objective. So, the test items from practice part were *fully* relevant to the given objective.

- iii) It has also been found that 'test method facet or how aspect' plays vital role to establish or maintain content relevance because it depends on the way of asking the question, i.e. whether the question has been constructed intending to test the grammatical competence or pedagogic skills of the testees in question.
- 3) It has been found that the test papers of "Grammar: Theory and Practice (Eng. Ed. 512)" have high content validity in terms of content representativeness (coverage) but the same test papers or test items have low content validity in terms of content relevance. This shows that the test papers having high content validity in terms of coverage do not necessarily have high content validity in terms of relevance and vice versa.
- 4) Regarding the coverage of the course contents, "Grammar: Theory and Practice (Eng. Ed. 512) Tests" have really covered what they have been supposed to cover from the course contents since they have covered more than 70 %.
- 5) Regarding the relevance of the test items, "Grammar: Theory and Practice Tests" have not satisfactorily tested what they have been intended to test from the testees according to the course objectives since just around 60% test items were relevant to the given objectives.
- 6) It has been found that the way of asking the questions (i.e. nature and number of the questions) from practice part was same except in 2059 (i.e. 20 marks out of 40) but some variation can be seen at the theory part in question.
- 7) There was also the repetition of the same test items regularly in the tests of Eng. Ed 512. For example, 2057<sup>L10</sup> (Opt.) was repeated

regularly in the years 2058, 2060 and 2061 from the same language item 4.4 (Case Grammar). Likewise, 2057<sup>L10</sup> (Opt.) was repeated in 2058 and 2060 from the same language item 4.5 (Systemic Grammar).

## **4.2 Recommendations**

On the basis of the findings of this present study, some recommendations for further improvement of the test items of "Grammar: Theory and Practice (Eng. Ed 512)" have also been proposed which are as follows:

- 1) Though "Grammar: Theory and Practice Tests" have high content validity in terms of representativeness or coverage as a whole, there is no proper balance in it. For example, if we see the content coverage of unit two yearwise, then the content coverage of it in 2058 was very high and in 2060 was very low. So, there should be proper balance of coverage not only in each units but also in each years.
- 2) Regarding the relevance principle, "Grammar: Theory and Practice Tests" have actually low but logically nearer content validity since just around 60% (i.e. 58.53) test items of Eng. Ed. 512 (Eng. Ed. 512) were relevant to the course objectives as a whole during seven years (2057-2063). So, more than 70% test items should be relevant to the course objectives to have high content validity (here content relevance).
- 3) A test is said to have high content validity only if it follows not only representative or coverage principle but also relevance principle equally. But the people take content validity just as a representation of the course contents and are not aware of the relevance principle and hence such findings (i.e. low or just nearer to content validity in terms of relevance but very high in terms of coverage) are found. Therefore, "Grammar: Theory and Practice

Tests" should equally follow both the principles: the principle of coverage and the principle of relevance to get a test with high content validity.

4) The content relevance of the test items from theory part of Eng. Ed. 512 was low since only around 39% test items of this part were relevant to the given objective. As this was because of different factors, such as, low quality of education, carelessness of test designer, lack of skilled administrator and exam controller, lack of specification grid, weaknesses of the syllabus or course of study itself as both *Generative Grammar and Other Models of Grammar* from theory part of this course were not designed especially for pedagogic purposes but for linguistic purposes so it was very difficult to fulfil the objective, lack of sufficient objectives (i.e. 1 for 5 units), and so on, these all affecting factors should be treated very carefully in order to get high content relevance of those test items as follows:

- i) The quality of education should be standard to make the testees ready to tackle the relevant test items with the objectives. The passion of the teachers towards teaching also plays vital role for this.
- ii) The test designer should be the trained and experienced subject teacher and he should be careful with the course objectives, course contents and specification chart or grid while designing the test items to get a test with high content validity especially in terms of relevance.
- iii) The questions or test items should be geared to the objectives. It means, every test items should test some sort of flavour of the objectives.

- iv) If there is specification grid, there is more possibility of designing the tests that can really test what they have been claimed to test from the testees according to the objectives and hence there is the establishment of content relevance. Therefore, there should be the provision of *specification grid* or any fixed criteria even in higher levels for constructing test items since it provides the test designers all the necessary information related with the test, such as, purpose of the test, type of test, type of the test item, number of the test items, test method facet, skills to be measured or tested, etc.
  - v) The syllabus (i.e. course of study) of "Grammar: Theory and Practice (Eng. Ed. 512)" should be revised including only the language items which are relevant to the course objectives and avoiding the language items which are beyond the level of the students or to be dealt by the linguists only. It means, the syllabus should be revised including only the language items which were especially designed for the pedagogic purposes but not for linguistic purposes so that it would be easy to maintain the content relevance of the test items.
  - vi) Besides general objectives, specific objectives and/or unitwise objectives should be included in the syllabus so that any body can be clear about them while constructing or examining the test items in question.
- 5) It would be better if choices are not given in the questions because it reduces reliability in the name of validity.
- 6) The test items should be constructed in such a way that they should demand or seek the creativity and originality (i.e. interpretative knowledge) of the testees to avoid the trend of cheating or copying

and making the bonafide students foolish. It means, the test items should encourage ownness and avoid the jug and mug theory.

- 7) Repeating the same test items regularly in most of the exam years is not good. Therefore, "Grammar: Theory and Practice Test" should include the test items from the different areas or neglected language items instead of repeating the same so that it increases the representation of the course contents and activates the testees as well.
- 8) The exam controller should take the examination as the life of the examinee so he should make the examination system very effective taking it as the positive spirit and play a vital role in improving the "Grammar: Theory and Practice (Eng. Ed. 512) Tests".

At last, the concerned authority should realize the fact that the test papers should have high content validity but in the name of content validity, other types of validity should not be discarded. For this, the necessary actions (may be conducting some seminars, workshops, conferences, etc.) should be taken. Moreover, the researcher thought that other different researches should be carried out in the related topic for the betterment of the total testing system.



## REFERENCES

- 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 2061*. Unpublished M.Ed. Thesis. Kathmandu: T.U.
- Bachman, L.F. 1998. *Fundamental Considerations in Language Testing*. London: OUP.
- Bachman, L.F. and Palmer, A.S. 1981. The Construct Validation of the FSI Oral Interview. *Language Learning*. 3/1: 67-86.
- Batala, K.B. 2004. *Validation of the SLC English Examination*. Unpublished M.Ed. Thesis. Kathmandu: T.U.
- Bhattarai, A. 2001. *Writing A Research Proposal*. NELTA Vol. 6, No. 1.
- Bhattarai, Jharana. 2005. *The Content Validity of English Textbook of Grade Eight*. Unpublished M.Ed. Thesis. Kathmandu: T.U.
- Brown, F.G. 1976. *Principles of Educational and Psychological Testing*. New York: Holt, Rinehart and Winston.
- Davies, A. 1968. *Language Testing Symposium*. London: OUP.
- Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T., and McNmara, T. 1999. *Dictionary of Language Testing*. Melbourne: University of Melbourne.
- Harmer, J. 1991. *The Practice of English Language Teaching*. London: Longman.
- Harrison, A. 1991. *A Language Testing Handbook*. London: Modern English Publications.
- Heaton, J.B. 1988. *Writing English Language Tests*. London: ELBS.
- Hockett, C.F. 1958. *A Course in Modern Linguistics*. New York: The Macmillan Company.
- Holden, S. 1982. *New ELT Ideas*. Bologna: Modern English Publications.
- Hornby, A.S. 2005. *Oxford Advanced Learner's Dictionary*, Seventh Edition. London: OUP.

- Horrocks, C. 1987. *Generative Grammar*. London: Longman.
- Hughes, A. 1995. *Testing for Language Teachers*. CUP: Cambridge.
- Khaniya, T.R. 1990. *Examinations as Instruments for Educational Changes: Investigating the Washback Effect of the Nepalese English Exams*. Unpublished Ph.D. Thesis. Edinburgh: University of Edinburgh.
- Khaniya, T.R. 2005. *Examination for Enhanced Learning*. Lalitpur: Millennium Publication.
- Kumar, R. 1996. *Research Methodology*. London: Sage Publication.
- Lado, R. 1965. *Language Testing*. London: Longman.
- Larsen-Freeman, D. And Celce-Murcia. 1983. *The Grammar Book for ESL/EFL Teachers*. Rowley: New Delhi House.
- Messick, S.A. 1975. *The Standard Problem: Meaning and Values in Measurement and Evaluation*. American Psychologist 30: 955-66.
- Messick, S.A. 1980. *Test Validity and the Ethics of Assessment*. American Psychologist 35: 1012-27.
- Morrow, K. 1986. *The Evaluation of Tests of Communicative Performance*. In Portal (ed.) 1986: 1-13.
- Neupane, S. 2005. *The Content Validity of English Textbook for Grade Seven*. Unpublished M.Ed. Thesis. Kathmandu: T.U.
- Ojha, D.R. 2006. *Content Validity of ELT Theories and Methods Exam at B.Ed. Level*. Unpublished M.Ed. Thesis. Kathmandu: T.U.
- Paneru, N.R. 2000. *English Sounds and Structure*. Bhotahity: Bidhyarthi Pustak Pasal.
- Portal, M. (ed.) 1986. *Innovations in Language Testing*. London: NFER-NELSON.
- Timilsina, H.L. 2006. *Content Validity of Language Testing Test at M.Ed. Level*. Unpublished M.Ed. Thesis. Kathmandu: T.U.
- Weir, C.J. 1990. *Communicative Language Testing*. London: Prentice Hall.
- Yadav, Y.P. 2004. *Theories of Grammar*. Kathmandu: Students' Book Publishers and Distributors.



## **Appendix-I**

The Syllabus on "Grammar: Theory and Practice  
(Eng. Ed. 512)"

## **Appendix-II**

Question Papers of "Grammar: Theory and Practice  
(Eng. Ed. 512)" Asked in the Final Written Exam  
during the Seven Years (2057-2063)