

# CHAPTER – I

## INTRODUCTION

### 1.1. Background of the Study

Examination is needed in order to test the achievement of the learners in formal education system. Specifically in the field of language education, examination is needed to test whether the learners have got mastery over language skills and related notions and aspects such as listening, speaking, reading, writing, spelling, pronunciation and vocabulary etc. Language education covers teaching, learning and testing to measure whether students have learnt the content taught. Therefore, we can say that in the education system, examination has great contribution and plays a very influential role.

Taking of the development and aim of examination system Mayazaki (1963.p.13) writes

*‘Examination was originated from China for the first time for clinical use where it was employed in the selection of public officials. It was not employed in the schools at that time. Schools were meant for providing educations for students but there was not any sorts of examination prescribed to evaluate the effectiveness of the teaching learning process.’*

It clarifies the fact that the examination system used in other system for the first time rather than in the education system. Oral examination was taken as the system of testing proficiency of any person. The Boston School Committee in England used printed examination only in 1845 A.D. for the first time (Encyclopedia American 1991, p.715-16). According to Oxford Advanced Learners Dictionary (7<sup>th</sup> ed.), ‘An examination is a formal test of somebody’s knowledge or ability in a particular subject, especially by means of written questions or practical exercises.’

Likewise, Anderson (1969,p.364) says ‘ As many subjects what things he is supposed to get out of a course and he will answer in terms of the test his teachers’ sets for him.’ An examination is to measure a child’s ability and it must be a measure of the present and potential future ability as well as measure of the memory of the past. The child must be capable of responding to a situation at any time in any circumstances. If only trained response is tested, it is useless and meaningless. Examination must not demand that the students come like overfilled tumbler and aborted memorized facts, having so much that they do not know where or in what questions to overflow’ ( Underwood:1968,p.1-2).

Hornby (6<sup>th</sup> edition, p. 431-432) has also defined, ‘examination- noun- a formal written, spoken, or practical test, specially at school, to see how much you know about a subject or what you can do’. ‘Exam is the usual for the written, spoken or practical test at school or college, especially an important one that you need to do in order to get a qualification, examination is very important. A test is something that students might be given in addition to, or something instead of, regular exam, to see how much they have learnt.

Examination is the sole of testing nevertheless it has been highly criticized by some eminent philosophers like Russell (1916), Krishnamurti ( 1953,p.14), Spencer ( 1988,p.6-7) questioning the perpetuation of the examination itself. Krishnamurti ( 1953.p.14) says ‘ we have made examinations and degrees that avoid current issues. Similarly Spencer ( 1988,p.6-7) says ‘I don’t think that testing is necessarily useful or even excusable. At the best, testing does more harm good at the ones,’ Narayan (Times of India, Apr 17, 1998 -1) says ‘in a civilized world, the examination should have no place because it is the culmination of all sadistic impulses

In conclusion, we can say that examinations are the only authentic tools for measuring students’ achievements in some certain academic period. Examinations in the past set artificial standard even for the native speakers and have often

demanding skills similar to those acquired by the great English essayist and critics and it also makes the teacher able to ascertain which parts of the language programme have been found difficult by the class. In this way, examination is also important for a teacher to evaluate the effectiveness of the syllabus as well as the methods and materials he/she is using.

### **1.1.2 Language Testing**

Hornby (6<sup>th</sup> edition, p.432), A test is something that students might be given in addition to, or something instead of, regular exam, to see how much they have learned. There is some confusion regarding the terminology used to denote the different types of language tests in use. Most test specialists, however, agree on the following broad divisions: achievement/attainment tests, proficiency tests, aptitude tests and diagnostic tests. This group can be further subdivided into class progress tests and (standardized) achievement tests.

### **1.1.2 Class progress Test**

This book has been concerned chiefly with class progress tests, since these are the most widely used types of tests. Most teachers are, at some time or other, required to construct such tests. Each progress test situation is unique and can only be evaluated fully by the class teacher in the light of his or her knowledge of the students, the programme which they have been following and the class teacher's own particular aims and goals. It is illogical to expect general purpose tests and books of tests to function as effectively as a test constructed specially for a particular situation: hence the purpose of this book in encouraging teachers to construct their own tests.

The progress test is designed to measure the extent to which the students have mastered the material taught in the classroom. It is based on the language programme which the class has been following and is just as important as an assessment of the teacher's own work as the student's own learning. Results

obtained from the progress test enable the teacher to become more familiar with the work of each student and with the progress of the class in general. The class progress test is a teaching device, its backwash effect on teaching and motivation being important futures. A good progress test should encourage the students to perform well in the target language and to gain additional confidence. Its aim is to stimulate learning and to reinforce what has been taught. Good performances may act as a means of encouraging the students, and although poor performances may act as an incentive to more work, the progress test is chiefly concerned with allowing the students to show what they have mastered. Scores on it should thus be high ( provide, of course, that progress has indeed been made ). Whereas in standardized achievement and proficiency tests, a wide range of performance should be indicated, the progress test should show a cluster of scores around the top of the scale.

### **1.1.3 Achievement Test**

Achievement (or attainment) tests, though similar in a number of ways to progress tests, are far more formal tests and are intended to measure achievement on a larger scale. Most annual school examinations take the form of achievement tests; all public tests which are intended to show mastery of a particular syllabus are also achievement tests. These tests are based on what the students are presumed to have learnt – not necessarily on what they have actually learnt nor on what has actually been taught. Achievement tests frequently take the form of secondary school entrance tests and school certificate examination; many are based on a published syllabus and exert a strong influence on the teaching and schools. Constructor of such tests rarely teach any of the students being tested (often an advantaged provided that the test constructors are very familiar with the teaching and learning problems of the testees). Indeed, this is often a prerequisite before anyone can be appointed to any position of responsibility in connection with this type of test, though this principle obviously cannot always be applied to school examinations.

Several achievement tests are standardized: they are pre-tested, each item is analyzed and revised where necessary, norms are established and comparisons made between performances of different students and different schools. Since such tests are administered year after year, it is possible to compare performances of students one year with those of students taking the test another year.

If students have followed a structural approach to language learning, it is clearly unfair to administer to a communicative achievement test at the end of their course. It is equally unfair to administer a structural-based test to those students who have followed a communicative approach to learning the target language. A good achievement test should reflect the particular approach to learning that has previously been adopted.

#### **1.1.4 Proficiency Test**

Whereas an achievement test looks back on what should have been learnt, the proficiency test looks forward, defining a student's language proficiency with reference to a particular task which he or she will be required to perform. Proficiency tests are in no way related to any syllabus or teaching programme, indeed, many proficiency tests are indeed for students from several different schools, countries and even language backgrounds. The proficiency test is concerned simply with measuring a student's control of the language in the light of what he or she will be expected to do with it in the future performance of a particular task. Does the student know enough in English for example to follow a certain university or college course given in the medium of English? Does the student now know enough in English in order to function efficiently in a particular type of employment? The proficiency test is thus concerned with measuring not general attainment but specific skills in the light of the language demands made later on the student by a future process of study or job.

### **1.1.5 Aptitude Test**

A language aptitude test (or thought test) is designed to measure the students' probable performance in a foreign language which he or she has not started to learn: i.e. it assesses aptitude for learning a language. Language learning aptitude is a complex matter, consisting of such factors as intelligence, age, motivation, memory, phonological sensitivity and sensitivity to grammatical patterning. The relative weighting given to these elements must depend on many factors and thus vary considerably from one individual to another. Some specialists in this field maintain that it is neither possible nor desirable to take an overall measurement of language aptitude: consequently aptitude is sometimes divided into various aspects according to the specific task for which a person is being trained: example listening interpreting, translating aptitude tests generally seek to predict the student's probable strengths and weaknesses in learning a foreign language by measuring performance in an artificial language. The ability to learn new phonemic distinctions and also to use language patterns in an unfamiliar but systematic manner is tested by means of the artificial language. Since few teachers are conversed with the complex field of aptitude testing, it is not necessary to go into further details here.

### **1.1.6 Diagnostic Test**

Although the term diagnostic test is widely used, few tests are constructed solely as diagnostic tests. Achievement and proficiency tests, however, are frequently used for diagnostic purposes: areas of difficulty are diagnosed in such tests so that appropriate remedial action can be taken later. Sections of tests which lend themselves particularly well to diagnostic purposes are phoneme discrimination tests, grammar and usage tests, and certain controlled writing tests. Clearly, weaknesses indicated in a test of vocabulary are not highly significant in themselves and can only be regarded as indicating general weaknesses. Similarly, many existing tests of reading comprehension are not very suitable for diagnostic purpose. Tests of writing and oral production can be used diagnostically provided

that there is an appreciation of the limits to which such tests can be put. Since diagnosing strengths and of teaching, the teacher should always be alert to every facet of achievement revealed in a class progress test.

Note that diagnostic testing is frequently carried out on groups of students rather than for individuals. If only one or two students make a particular error, the teacher will not pay too much attention. However, if several students in the group make a certain error, the teacher will note the error and plan appropriate remedial teaching.

### **1.1.7 Qualities of Good Test**

A test is considered a good when it has good validity and reliability. Validity is arguably the most important criteria for the quality of a test. The term validity refers to whether or not the test measures what it claims to measure. On a test with high validity the items will be closely linked to the test's intended focus. For many certification and licensure tests this means that the items will be highly related to a specific job or occupation. If a test has poor validity then it does not measure the job-related content and competencies it ought to. When this is the case, there is no justification for using the test results for their intended purpose. There are several ways to estimate the validity of a test including content validity, concurrent validity, and predictive validity. The face validity of a test is sometimes also mentioned.

#### **i. Content Validity**

While there are several types of validity, the most important type for most certification and licensure programs is probably that of content validity. Content validity is a logical process where connections between the test items and the job-related tasks are established. If a thorough test development process was followed, a job analysis was properly conducted, an appropriate set of test specifications were developed, and item writing guidelines were carefully

followed, then the content validity of the test is likely to be very high. Content validity is typically estimated by gathering a group of subject matter experts (SMEs) together to review the test items. Specifically, these SMEs are given the list of content areas specified in the test blueprint, along with the test items intended to be based on each content area. The SMEs are then asked to indicate whether or not they agree that each item is appropriately matched to the content area indicated. Any item that the SMEs identify as being inadequately matched to the test blueprint or flawed in any other way, are either revised or dropped from the test.

## **ii. Concurrent Validity**

Another important method for investigating the validity of a test is concurrent validity. Concurrent validity is a statistical method using correlation, rather than a logical method. Examinees who are known to be either masters or non-masters on the content measured by the test are identified, and the test is administered to them under realistic exam conditions. Once the tests have been scored, the relationship is estimated between the examinees' known status as either masters or non-masters and their classification as masters or non-masters (i.e., pass or fail) based on the test. This type of validity provides evidence that the test is classifying examinees correctly. The stronger the correlation is, the greater the concurrent validity of the test is.

## **iii. Predictive Validity**

Another statistical approach to validity is predictive validity. This approach is similar to concurrent validity, in that it measures the relationship between examinees' performances on the test and their actual status as masters or non-masters. However, with predictive validity, it is the relationship of test scores to an examinee's future performance as a master or non-master that is estimated. In other words, predictive validity considers the question, "How well does the test predict examinees' future status as masters or non-masters?" For this type of

validity, the correlation that is computed is between the examinees' classifications as master or non-master based on the test and their later performance, perhaps on the job. This type of validity is especially useful for test purposes such as selection or admissions.

#### **iv. Face Validity**

One additional type of validity that you may hear mentioned is face validity. Like content validity, face validity is determined by a review of the items and not through the use of statistical analyses. Unlike content validity, face validity is not investigated through formal procedures and is not determined by subject matter experts. Instead, anyone who looks over the test, including examinees and other stakeholders, may develop an informal opinion as to whether or not the test is measuring what it is supposed to measure. While it is clearly of some value to have the test appear to be valid, face validity alone is insufficient for establishing that the test is measuring what it claims to measure. A well developed exam program will include formal studies into other, more substantive types of validity.

### **1.1.8 Reliability**

According to Colin Phelan and Julie Wren (2005-06), Reliability is the degree to which an assessment tool produces stable and consistent results. According to him, here are its' types.

#### **i. Test-retest Reliability**

Test-retest reliability is a measure of reliability obtained by administering the same test twice over a period of time to a group of individuals. The scores from Time 1 and Time 2 can then be correlated in order to evaluate the test for stability over time.

## **ii. Parallel forms Reliability**

Parallel forms reliability is a measure of reliability obtained by administering different versions of an assessment tool (both versions must contain items that probe the same construct, skill, knowledge base, etc.) to the same group of individuals. The scores from the two versions can then be correlated in order to evaluate the consistency of results across alternate versions.

## **iii. Inter-rater Reliability**

Inter-rater reliability is a measure of reliability used to assess the degree to which different judges or raters agree in their assessment decisions. Inter-rater reliability is useful because human observers will not necessarily interpret answers the same way; raters may disagree as to how well certain responses or material demonstrate knowledge. It is used to assess the degree to which different raters/observers give consistent estimates of the same phenomenon. It is an agreement between the scores assigned by two raters (calculated as a percentage of agreement between the two or a correlation between the two).

## **iv. Internal Consistency Reliability**

Internal consistency reliability is a measure of reliability used to evaluate the degree to which different test items that probe the same construct produce similar results.

### **a. Average Inter-item Correlation**

Average inter-item correlation is a subtype of internal consistency reliability. It is obtained by taking all of the items on a test that probe the same construct (e.g., reading comprehension), determining the correlation coefficient for each pair of items, and finally taking the average of all of these correlation coefficients. This final step yields the average inter-item correlation.

### **b. Split-half Reliability**

Split-half reliability is another subtype of internal consistency reliability. The process of obtaining split-half reliability is begun by “splitting in half” all items of a test that are intended to probe the same area of knowledge (e.g., World War II) in order to form two “sets” of items. The entire test is administered to a group of individuals, the total score for each “set” is computed, and finally the split-half reliability is obtained by determining the correlation between the two totals “set” scores.

### **v. Inter-judge Reliability**

According to CET, ( 2003 ), If we have judges making decisions about a group of people, we have to be concerned with inter-judge reliability, the extent to which the judges agree with each other. With two judges and several examinees, we can calculate the correlation between the two judges. If we have two judges who are categorizing a number of people, we can study how often they are in agreement or summarize this into a statistic by calculating Cohen’s kappa coefficient.

## **1.1.9 Structure of School Education**

According to National Curriculum Framework for School Education in Nepal, In view of the national policy, international practice, and present educational context and concerns, Nepali formal school education will be of 12 years with two tiers- primary and secondary. The primary level will be from grade 1 to grade 8, and the secondary level will be from grade 9 to grade 12. As a preparatory stage for primary education, there will be two years of pre-primary education.

At the lower level secondary level 16.4 percent of the total time is allocated to English language teaching. Three year English curriculum of grade 6, 7, and 8 has two purposes. Firstly to enable student to exchange ideas with people of any nationality who speak or write English and secondly, to them to the vast treasures of the knowledge and pleasure available in written and spoken English language

learning is regarded as a skill. So the lower secondary level English curriculum emphasizes the development of the four skills of languages. A person who has not developed listening, speaking, reading and writing skills cannot be considered good users of language. The curriculum includes information on instructional objectives and skills, scopes and sequence, number of periods allocated for each unit, the principles of teaching English and the methodology to follow while teaching English in classroom. There are varieties of methods, techniques and activities included. Among them teachers are advised to select the teaching method which suits their local context.

### **i. Level wise Objectives of School Education**

The level wise objectives of school education are presented below according to National Curriculum Framework For School Education (Pre-primary-12) In Nepal ( 2005) , Ministry of Education and Sports and Curriculum Development Centre, Sanothimi, Bhaktapur

#### **a. Objectives of Pre-Primary Education**

The ultimate goal of pre-primary education is to bring about holistic development of children and to facilitate children to have smooth transition to primary education. More specifically, preprimary education will have the following objectives:

- i. To provide an opportunity to develop physical, emotional, social, mental, moral and creative aspects.
- ii. To develop habit for personal sanitation.
- iii. To develop respect for others.
- iv. To develop positive behaviour and attitude towards school.
- v. To prepare for primary schooling.

**b. Objectives of Primary Education (Grades 1-8)**

The main aim of primary education is to develop the innate ability of each child through child centered education. Its ultimate aim is to produce citizens who believe in the nation and in democracy and are aware of their responsibility towards the social and natural environment. Students are expected to be competent in communicating ideas, are independent and hard working, health conscious and ethical. More specifically, the objectives of primary education are the following:

- i. To develop basic knowledge of democratic values and norms and a positive spirit for Nepali nationality.
- ii. To develop basic knowledge of Nepali history, society, and multiculturalism.
- iii. To develop basic knowledge and skills on at least two languages (Mother tongue, Nepali, English) necessary for self expression and communication with others.
- iv. To develop basic knowledge in science, mathematics, environment, health, information technology, and life skills.
- vi. To develop creative and expressive skills.
- vii. To make children aware of child rights and the importance of respecting everyone's human rights.
- viii. To contribute to children's overall development—i.e. physical, mental, emotional, and social.
- ix. To develop life skills such as problem solving, creative thinking, inductive thinking, decision-making, safe health practices, and teamwork.
- x. To develop civic consciousness.
- xi. To develop respect for cultural differences and an eagerness to learn about different cultural beliefs, practices and customs.
- xii. To develop a strong sense of non-discrimination towards others despite their caste, ethnicity, religion, language, gender, class, and disability.

xiii. To develop a positive outlook towards work and respect for labor.

The allocation marks in the DLE examination in English is as indicated in the grid below:-

Language skills	Listening	Speaking	Reading	Writing	Total Marks
Marks	20	20	30	30	100

A set of questions are prepared by CDC for submission to Office of Controller Examination to guide the questions setters and examiners and for dissemination to Lower Secondary Schools according to CDC (1993). *Lower secondary education curriculum*. Bhaktapur: Curriculum Development Centre, HMG/MOES.

## **1.2 Statement of the Problem**

Reliability is inevitable part of every examination question set. Scoring reliability of any question set among many scorers may be varied. The study has attempted to find out scoring reliability of a DLE English question set among many scorers.

## **1.3 Objectives of the Study**

The present study has the following objectives:

- i. To find out inter-rater scoring reliability.
- ii. To suggest pedagogical implications for the improvement of test construction, effective classroom activities i.e. teaching and learning.
- iii. To provide feedback to test examiners.

## 1.4 Significance of the Study

The study will be beneficial to the concerned institute as it tries to find out the reliability of basic level governmental final examination. Moreover, the study is expected to be significant to all those who are in schools and are involved in the programme of teaching learning; especially teachers, examiners, guardians, question makers, syllabus designers and textbook writers of school basic level education.

## 1.5 Delimitations of the Study

My study covered two schools of Koshi-Haraicha Municipality, which are Haraincha Model English School and Lovely Angels English School. The District Level Examination question set belongs to DLE English question Set – 2069.

## 1.6 Definition of the Terms Used

**Validity-** Inherent quality of a good test that reminds a test measure what it intends to measure.

**Correlation Coefficient-** It is statistical term which finds the degree of relationship between two or more than two variables.

**Variables-** The factors that affect the study.

**Test-** test is something that students might be given in addition to, or something instead of, regular exam, to see how much they have learned

**G theory-** Generalizability theory

**CTT-** Classical Test Theory

**SMEs -** Subject Matter Experts

**% -** Mathematical Term to express in 100.

**Reliability-** Reliability is the degree to which an assessment tool produces stable and consistent results

**PTI-** Professional Testing Inc.

**SLC-** School Leaving Certificate

**DLE-** District Level Examination

**Examination-** a formal test of somebody's knowledge or ability in a particular subject especially by means of written questions or practical exercises.

**Co-relation-** a statistical tool with the help of which, we can determine whether or not two or more variables are co-related and if they are co-related, what is the degree and direction of co-relation

**Co-relation co-efficient-** degree of linear relationship between two or more variables

**Test-retest-** A method for finding reliability on which a test is administered two times within a short time gap. The correlation of the performances of the students on the two tests is interrupted as the correlation co-efficient of the test. The higher the correlation is, the higher the reliability of the test.

## **CHAPTER- II**

### **REVIEW OF RELATED LITERATURE AND THEORITICAL FRAMEWORK OF THE STUDY**

#### **2.1 Review of Theoretical Literature**

There is the dearth of literature in this particular field of study. However, a few works and researches have been carried out in connection with examination and in some way to this topic.

Pradipt (2008) carried out a research entitled 'The Reliability of SLC English Question Papers' and concluded the consistency of the score was higher in public schools than that of private schools.

Dhakal (2000) carried out a research on the 'Effectiveness of Discrete test and Integrative Test as a Measure of English Proficiency'. He found that both of these test could be the effective measures of language proficiency test.

Batala (2004) carried out a research entitled 'Validation of the SLC English Examination' in which he found the predictive validity of the SLC examination is very low; where as its content validity is found to be quite satisfactory.

Khanal (1999) studied on the 'Effectiveness of the Cloze Test over Conventional Objective Test in testing Reading Comprehension'. He found that the cloze test was equally good as the objective tests. Moreover, the cloze test was easier to construct and administer. Because of these additional advantages it was recommended in School Level English Language learning situation in Nepal.

Kshetree (2001) carried out a study entitled 'A Study on the Effect of the SLC Examination' A comparative study of he result between Rautahat and Makawanpur districts. He found that the proficiency of boys is little bit higher than those of the girls proficiency in comprehending written text and seen text is

higher than that in unseen texts. There is a wide gap between and among the proficiency levels of various schools research, students of Rautahat and comprehend written text better than Makawanpur.

Sharma (1999) has carried out a research entitled 'A Study in Making an English Answer Book'. He compared various making techniques used by examiners to check SLC compulsory English answer books. He found that there was a variation in making of SLC English answer book with respect to examiner's qualification, training, and their attitude towards a particular language item etc.

Singh (1996) carried out a comparative study between Grade 8 of public and private sectors on the achievement of listening comprehension test. He found that students of private schools are far better than their counter part of public school.

Adhikari (2004) conducted a research on the "Use of Multiple Sets of Question Papers in the SLC Examination" and concluded that question papers are more or less similar but they did not have adequate information for the examination.

Kandel (2006) carried out a research entitled 'A study on the Co-relationship between Sent-up test and SLC Examination Results'. And found out that the predictive validity of send-up test examination English question paper is very high. Send-up test English question paper is helpful and relevant to SLC in terms of course standard and question format factors influencing the English Examination are: Efforts of influencing the SLC examination, standard of question, unfair checking of answer sheets, exam centre, practicing unfair means in the proper guidance, supervision, and learning environment and availability sufficient resources and materials to practice for the exam.

Pokhrel (2006) carried out a research entitled on "Errors in the SLC Question Papers:A case of Grammatical Item" and found that a large number of question

papers were erroneous and students committed mistakes due to overgeneralization or being confused.

## **2.2 Implications of the Review**

The previous carried theses are of great importance for the emerging research. Every country has its' own education system and examination system. In context to Nepal, validity and reliability of every school level examination should be addressed. The addressing is made by such researches. District Level Examination is end of basic level examination in Nepal. Reliability of examination of this level and scoring reliability among many scorers of each question set are not analysed to make this examination and scoring of students more reliable. This is an attempt to raise question about scoring reliability among many scorers. Being not analysed of examination review because of giving less importance to Basic Level and other difficulty but the unpublished thesis is of great importance. It keeps open to all for further criticize, analyse and evaluate. This is a window through view of examination can be peeped.

## **2.3 Theoretical/ Conceptual Framework of the Study**

My study is based on two theories i.e. Classical Test Theory and Generalizability Theory.

### **2.3.1 Classical Test Theory**

According to Classical Test Theory, It is a trite observation that all human endeavors are replete with error. And human endeavor called science is no exception. There are many possible ways to error in measurement. In other words, there are many sources of errors. These sources may vary dependent upon the particular branch of science involved. The question now is to tackle the problem of errors of measurement. The answer to this question appears to be simple: develop a theory of errors, or some would say, set up an error model. Indeed, in psychological and educational measurements, this is an approach that has been

followed since more than a century. And the earliest theory around is classical test theory.

Classical test theory (CTT) has been the foundation for measurement theory for over 80 years. The conceptual foundations, assumptions, and extensions of the basic premises of CTT have allowed for the development of some excellent psychometrically sound scales. Because total test scores are most frequently used to make decisions or relate to other variables of interest, sifting through item-level statistics may seem tedious and boring. However, the total score is only as good as the sum of its parts, and that means its items. Several analyses are available to assess item characteristics.

By defining true score an explicit, abstract formulation of measurement error is given. These are relevant for reliability also, and for considering various types of equivalence or comparability of test forms.

### **2.3.2 Generalizability Theory**

According to Generalizability Theory, A particular behavioral measurement is conceived of as a sample from a universe of admissible observations, or a domain of suitable or appropriate observations. The universe or domain is characterized by one or more sources of error variation, called facets. In a study where students' performance is rated by judges on performance criteria, judges and Test Theory performance criteria are the facets of a measurement, and each facet consists of a set of conditions. Usually the facets are assumed to be indefinitely large. The universe, then, is defined as all possible conditions of the facets. Ideally, a person's universe score is his or her average score over all conditions. In a measurement situation, however, error is at stake, and this calls for the estimation of variance components. These estimates of variance components are highly informative, whereas the so-called generalizability coefficients - the G theory counterparts of classical reliability coefficients - are straightforward ratios of appropriate universe-score variances to total-score variances. In addition to

variance components standard errors are considered as appropriate indicators of uncertainty of, e.g., performance assessments of student achievement or school effectiveness. The purpose of G theory is to generalize from an observation at hand, i.e. the observed score, to the appropriate universe of observations. This domain or universe is defined by all possible conditions of the facets of the study. It should be noticed that the object of measurement (e.g., the persons to whom a test is presented) is not a facet.

Likewise, an observed score obtained with a psychological or other behavioral measurement instrument is just one of many possible scores that could have been obtained, since an alternative context, other measurement conditions and varying circumstances may lead to other observed scores.

In other words, an observed score is usually obtained for a particular test form. Another, equivalent test form, however, may have been as appropriate for our measurement purpose, but might have led to a different observed test score. Consequently, if one wants to model observed scores, one has to take into account many sources of variation (including error variation). This also applies if one considers the reliability of scores obtained from a measurement instrument. Classical reliability provides a decomposition of the observed score into a true score and only one type of error. Theoretically this error is undifferentiated. The several reliability estimation procedures lead to specific conceptualizations of error. Parallel test forms reliability, e.g., considers the lack of equivalence between the forms as the source of error, test-retest reliability the time of testing, and internal consistency reliability the variability in test items. Generalizability theory or G theory, in contrast to classical test theory, provides a decomposition of an observed score taking into account more sources of variation, dependent upon the specific measurement situation. So G theory also recognizes that multiple sources of error may operate in a measurement implying that there is no unitary definition of reliability. This is, basically, a consequence of the view that a specific test score or any other particular behavioral measurement for that matter

(e.g. a job performance score, or an expert's performance assessment of student achievement) is conceived of as a sample from a universe of admissible or suitable observations. Such a universe is characterized by one or more sources of variation, the so-called facets.

Likewise, according to G Theory, A particular behavioral measurement is conceived of as a sample from a universe of admissible observations, or a domain of suitable or appropriate observations. The universe or domain is characterized by one or more sources of error variation, called facets. In a study where students' performance is rated by judges on performance criteria, judges and Test Theory performance criteria are the facets of a measurement, and each facet consists of a set of conditions. Usually the facets are assumed to be indefinitely large. The universe, then, is defined as all possible conditions of the facets. Ideally, a person's universe score is his or her average score over all conditions. In a measurement situation, however, error is at stake, and this calls for the estimation of variance components. These estimates of variance components are highly informative, whereas the so-called generalizability coefficients - the G theory counterparts of classical reliability coefficients - are straightforward ratios of appropriate universe-score variances to total-score variances. In addition to variance components standard errors are considered as appropriate indicators of uncertainty of, e.g., performance assessments of student achievement or school effectiveness. The purpose of G theory is to generalize from an observation at hand, i.e. the observed score, to the appropriate universe of observations. This domain or universe is defined by all possible conditions of the facets of the study. It should be noticed that the object of measurement (e.g., the persons to whom a test is presented) is not a facet.

## **CHAPTAR- III**

### **METOHOD AND PROCEDURE OF THE STUDY**

#### **3.1 Methods and Procedures of the Study**

The methodology to carry out the research was as follows:

##### **3.1.1 Design and Method of Study**

I wanted to do the research in scoring reliability. I wanted to find out scoring reliability of District Level Examination Question Set among many scorers. I made random smplng to choose two schools and students. I consulted with principals of both school and took permission. I informed the date of examination to the students have just passed class-8 in the last academic year-2069 BS then I prepared the question set of 2069 BS. I administered an exam to the students. I did it to 45 students of private schools out of 75 students. After that I marked code number to all answer sheets and photocopied two sets of each answer sheet. Then I gave each answer sheet of all students to three different scorers. After getting the data from them, I analyzed, compared and calculated the data and found out reliability of scoring DLE English question set.

#### **3.2 Population, Sample and Sampling Strategy**

##### **3.2.1 Population**

The populations of this study were students of grade eight of Koshi-Haraincha Municipality Morang.

##### **3.2.2 Sample**

To collect the data I used both qualitative and quantitative method. I also used the purposive and randm sampling method. I took 45 students of two schools out of 75 students.

### **3.2.3 Sampling Method**

I went to Koshi-Haraincha Municipality Ward No 13 and 8 at Haraicha Model English School and Lovely Angels English School. I used qualitative and quantitative method. I also used purposive sample method. The informants were both male and female, and from different castes.

### **3.3 Study Area/Field**

The subject area of my thesis is scoring reliability of DLE English question set. I compared, calculated and analyzed the following terminologies.

Inter-scorer reliability of a question set.

Reliability among many scorers of a question set

Co-relation co-efficient

### **3.4 Data Collection Procedure**

I arranged a question set of DLE and took test to the students. Then I photocopied the answer sheets and gave three subject teachers to check and score the answer sheets. Two scorers belong to public schools and one belongs to private school.

### **3.5 Data Analysis and Interpretation Procedure**

After collecting the data, I presented the data and analyzed statistically and descriptively. I presented data in graph first and calculated co-relation co-efficient.

#### **3.5.1 Co relational Analysis**

The main concern of co relational analysis is to find out the existence of relationship between two or more variables. The focus is on determining the nature pattern, magnitude or the degree of relationship. If a relationship exists between two variables, it means that scores within a certain range on one measure (variable) are associated with scores within a certain range on another variable. There is a relationship between intelligence test and academic achievement; i.e.

person whose score is high on intelligence tests tends to have higher grade points and persons who score is high on intelligence tests tend to have lower grade points. Thus, if high scores on one variable are associated with high scores on the other variables or low scores on one variable are low with low scores on the other variable, there is a positive relationship between these two variables. Contrary, to this, if high scores on one variable are associated with low scores on another variables there exists a negative relationship, between those two variables. Lastly, if there is not found any systematic pattern of relationship between two variables. There is no relationship between those variables; this is known as zero relationship. The nature or pattern of relationship, i.e. positive and negative, and zero can be graphically presented in scattered plot or scattered diagram. The data in this research all those diagrams and are presented in the following pages of analysis.

There are two major weaknesses in showing correlation in a scatter plot/scatter gram. Firstly, it is time consuming because correlation research is a group study and to prepare a scattered plot for each student takes a lot of time. Secondly, scattered plot cannot display the quantitative measure of degree of relationship. So, analysis in correlation research is displayed through correlation coefficient or coefficient of co relation.

A corelation is a number, quantitative measure that represents the degree of relationship between two variables. The measurement of co relation between two variables results in a value that ranges from both 0.00 and +1 or/and 00 and -1. This value is called the co relation coefficient. A co relation coefficient is a decimal number between 00 and +1 or 00 and -1, which includes the degree to which two variables are related.

For the purpose of interpreting co relation coefficient quantitatively, Carret (1970) has given the following established criterion:

- a)  $r$ 's from 0.00 to  $\pm 0.20$  = very low, negligible.
- b)  $r$ 's from  $\pm 0.21$  to  $\pm 0.40$  = low, present, but a slight.

- c) r's from  $\pm 0.41$  to  $\pm 0.70$  = substantial or marked.
- d) r's from  $\pm 0.71$  to  $1.00$  = high or very high.

Similarly, Sthapit et al. (2004) have also given some relevant criteria for the interpretations of co-relation co-efficient are given below;

Criterion for Interpretation of Co relation Coefficient:

Degree	Direction (Positive)	Direction (Negative)
Perfect	+1.00	-1.00
Very high	+0.75	-0.75 to -1.00
High	+0.50 to 0.75	-0.50 to -0.50
Low	+0.25 to + 0.50	-0.25 to -0.50
Very low	0.00 to 0.025	0.00 to -0.025
Absent	0.00	0.00

### 3.5.2 Characteristics of Correlation Study:

The characteristics of correlation study are as follows:

- 1) Co relational research is objective based, whether relationship studies or prediction studies because all other processes depend on the objective of the research. It can be explained with comparing to the experimental study.
- 2) I can deal with multiple variables at the same time, so it is appropriate to conduct in the situation where it is difficult to conduct experimental study. The researcher first conducted the co relational study and found out the effective relationship, as positive negative or zero, if the relationship was negative then he eliminates these variables and he only further he carried out the experimental research. Data were collected in natural setting.
- 3) The researcher found out degree of relationship rather than all or nothing question.

- 4) In co relational study qualitative data was collected and analyzed, i.e. statistical approach was used for analysis, co relation co-efficient was found out.
- 5) It was a group study unlike case study and ethnographic research, a group was used as the population of study, it provides information about group of students or other entities. But on the other hand, it failed to tell position of the individual like in case study.

### 3.5.3 Method of Correlation Co-efficient Calculation

There are many ways or technique of calculating or finding out the co relation co-efficient statistical process. Here is given one of the methods given by Karl Pearson (1988) is as follows:

The raw scores obtained from the data operations directly used. There are certain steps to carry out the process. They are given below:

- Step 1:** List the scores for each student in parallel column according to raters, where 'x' stands for scores on one variable 'y' stands for score on next variables.
- Step 2:** Square each and enter the numbers in the  $x^2$  and  $y^2$  column.
- Step 3:** Multiply the 'x' and 'y' enter into the 'xy' column.
- Step 4:** Add up each column, and
- Step 5:** Apply the following formula.
- Step 5:** Apply the following formula.

$$\text{Correlation co-efficient } r_{xy} = \frac{N\sum xy - \sum x \sum y}{\sqrt{N\sum x^2 - (\sum x)^2} \sqrt{N\sum y^2 - (\sum y)^2}}$$

According to Martyn Penscombe (1999) researcher generally regard any co relation co-efficient which is positive or negative 0.3 below as weak relation and 0.7 below as strong relation but for James Dean Brown (1999) weak equals to 0.40 and below and strong equals to 0.80 and above.

## CHAPTER–IV

### ANALYSIS AND INTERPRETATION

#### 4.1 Introduction

Under this chapter, the data have been analyzed, interpreted and presented systematically. The first objective of this study was to find out the scoring reliability of DLE question papers on the basis of inter-rater scoring method. To accomplish the objective, forty five students were selected from each of the two private schools of Koshi Haraincha Municipality (henceforth school 'A' and school 'B') respectively. The scoring reliability of inter-rater DLE question papers (2069) i.e. consistency of score has been presented and described in this chapter and the chapter is followed by the influencing factors of the result.

This chapter has two main sections. The first section deals with the analysis of the scoring reliability of inter-rater of question paper and the second section deals with the analysis by the researcher himself. For the purpose of examining the scoring reliability of the rater of question papers of DLE, the researcher at first, compared the score (performance) of the students received in the test examination with the performance (score) they achieved in their DLE English examination. Then, the correlation analysis of the three sets of scores (performance) has been shown. In support of the finding of the scoring reliability of DLE question papers, the data has been further analyzed.

In order to analyze and present the result, statistical as well as descriptive tools have been adopted. The researcher has analyzed and interpreted the data descriptively by using simple statistical tools such as percentage, mean correlation, co-relation co-efficient presented through tables, graphs and bar diagrams. Each of the tables presented in this study is followed by graph for the purpose of clarifying ideas. In fact, the tables and graphs are complementary to each other. Clarity and convenience, rather than economy in space was guiding principle behind this kind of presentation.

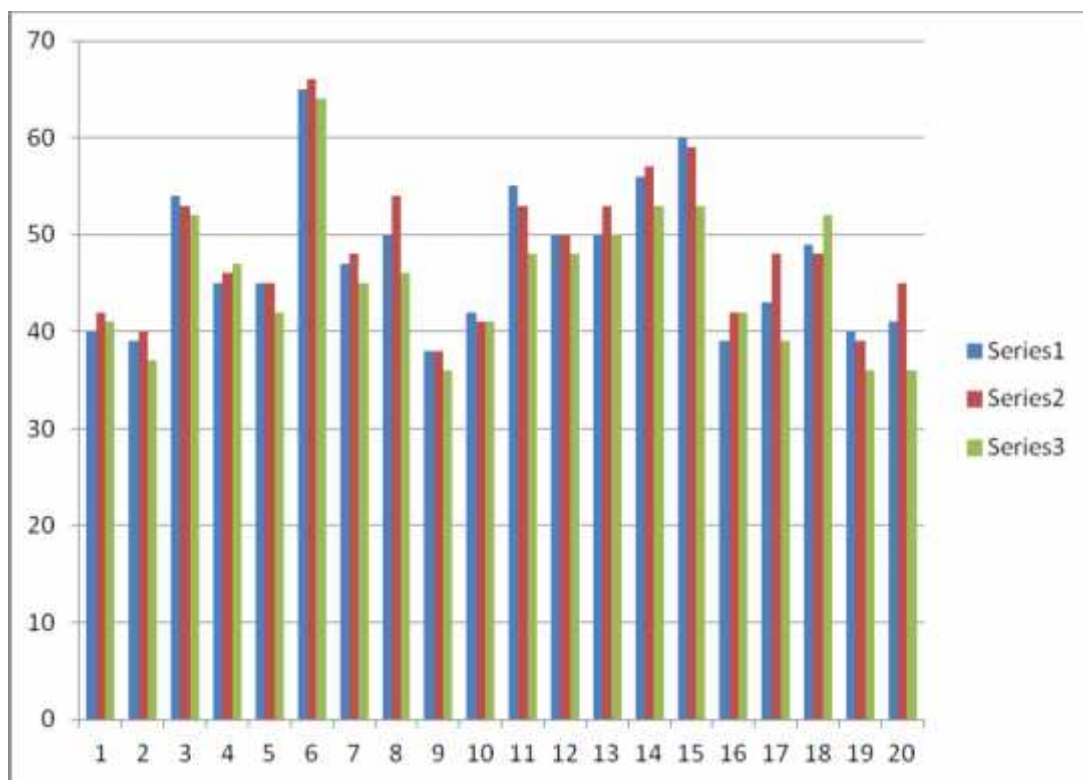
## 4.2 Reliability of the DLE Set in terms of the rater

Based on the rating done by three teachers (raters), the reliability of the DLE test sets have been marked in this study. In doing so, the scores of students were compared in a pair scorer's respectively to find out the scoring reliability of DLE question papers. The reliability has been identified first by means of the comparison of the the performance of students of two schools as follows:

## 4.3 The School-Wise scoring Comparison

I presented school wise figure and table below.

### 4.3.1 Comparison among Three Sets of Scores of the Students of School A.



**Figure: 1** Comparison among Three Sets of Scores of the Students of School A.

*Table – 1* Comparison among Three Sets of Scores of the Students of School ‘A’

Raters	No. of students			
	Number of students obtaining marks below 40	Number of students obtaining marks between 40 to 49	Number of students obtaining marks between 50 to 59	Number of students obtaining marks above 60
A	3	9	6	2
B	4	8	7	1
C	5	9	5	1

The figure and table reveal that all the twenty students obtained 51% above in the inter-rating of scoring of DLE English examination test. For this I calculated percentage where the lowest score was 36 out of full marks of 70. There was in rating of A rater nine students scored 40 and 40 above, six students scored 50 and 50 above and two students scored 60 and 60 above. In rating of B rater eight students scored 40 and 40 above, seven students scored 50 and 50 above and one student scored 60 and 60 above. In rating of C rater nine students scored 40 and 40 above, five students scored 50 and 50 above and one student scored 60 and 60 above. So scores were less varied raters to raters, therefore inter rater scoring reliability of School A was reliable.

### 4.3.2 Comparison among Three Sets of Scores of the Students of School 'B'.

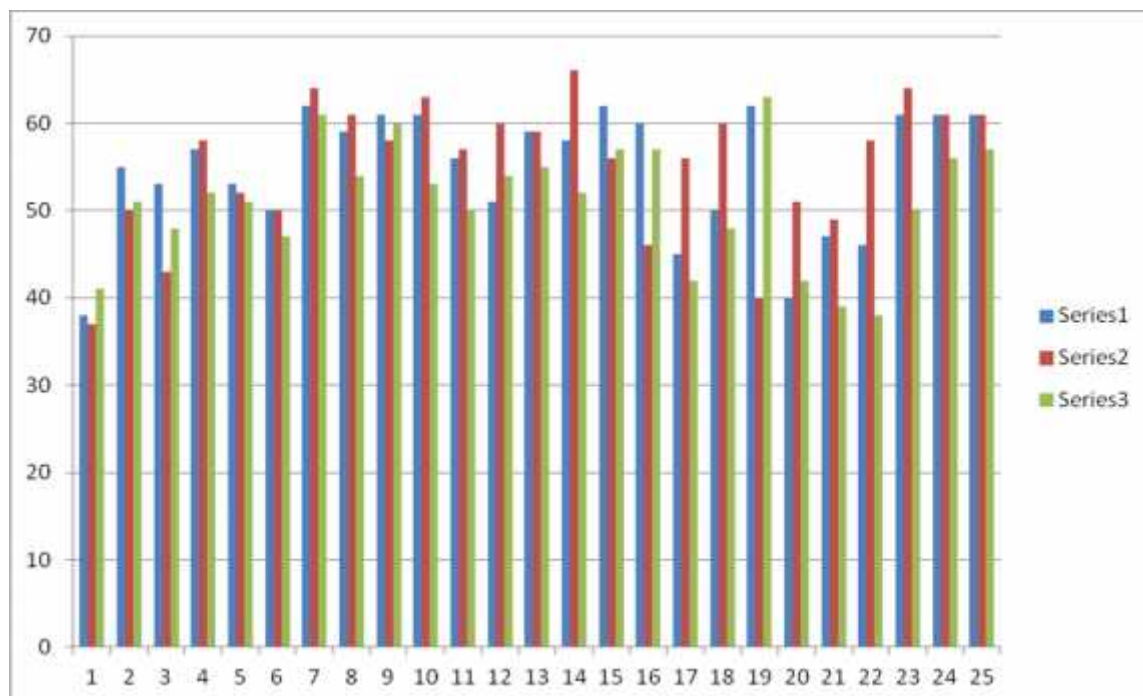


Figure: 2 Comparison among Three Sets of scores of the Students of School 'B'

	No. of students			
Raters	Number of students obtaining marks below 40	Number of students obtaining marks between 40 to 49	Number of students obtaining marks between 50 to 59	Number of students obtaining marks above 60
A	1	4	11	9
B	1	4	12	8
C	2	6	14	3

Table – 2 Comparison among Three Sets of scores of the Students of School 'B'

The figure and table show that all the twenty students obtained 52% above in the inter-rating of scoring of DLE English examination test. For this I calculated percentage of score where lowest score was 37 out of full mark of 70. There was, in rating of A rater four students scored 40 and 40 above, eleven students scored 50 and 50 above and nine students scored 60 and 60 above. In rating of B rater four students scored 40 and 40 above, twelve students scored 50 and 50 above and 2eight student scored 60 and 60 above. In rating of C rater six students scored 40 and 40 above, fourteen students scored 50 and 50 above and three student scored 60 and 60 above. So scores were similar raters to raters, therefore inter rater scoring reliability of School B was more reliable.

#### 4.4 The Sex-Wise scoring Comparison

I presented sex wise figure and table below.

##### 4.4.1 Comparison among Three Sets of Scores of the Girl Students of School 'A' and 'B'.

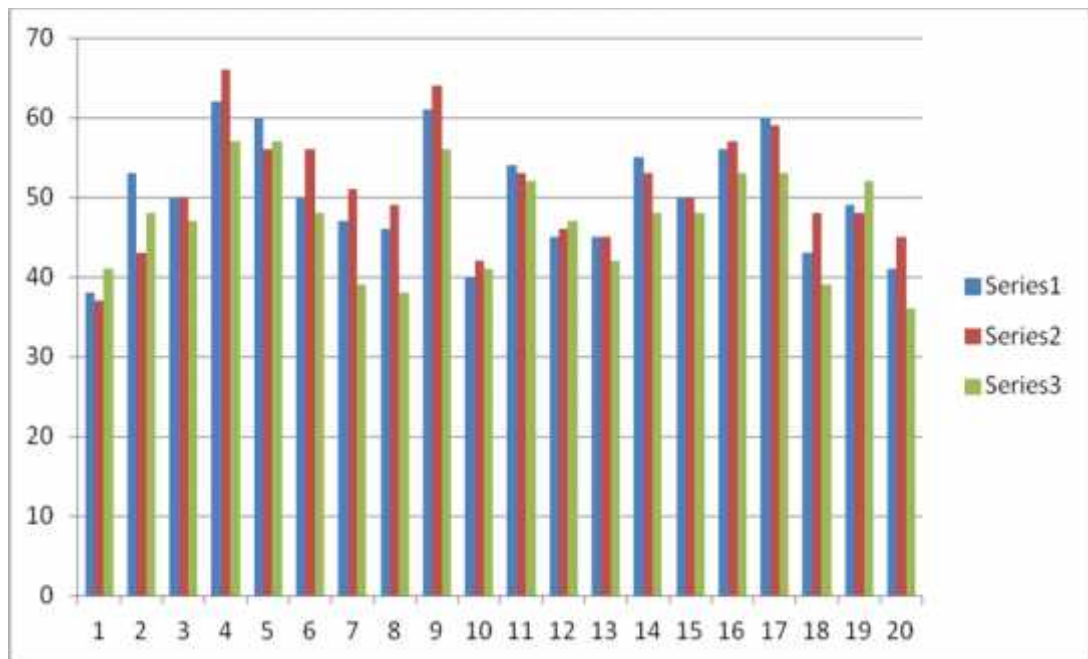


Figure: 3 Comparison among Three Sets of Scores of the Girls Students of School 'A' and 'B'

*Table – 3 Comparison among Three Sets of Scores of the Girls Studetnts of School 'A' and 'B'*

Raters	No. of Girl students			
	Number of students obtaining marks below 40	Number of students obtaining marks between 40 to 49	Number of students obtaining marks between 50 to 59	Number of students obtaining marks above 60
A	1	8	7	4
B	1	8	9	2
C	4	9	7	0

The figure and table that all the twenty students obtained 51% above in the inter-rating of scoring of DLE English examination test. For this I calculated percentage of score where lowest score was 36 out of full mark of 70. There was in rating of “A” rater eight girl students scored 40 and 40 above, seven girl students scored 50 and 50 above and four students scored 60 and 60 above. In rating of “B” rater eight students scored 40 and 40 above, nine girl students scored 50 and 50 above and two student scored 60 and 60 above. In rating of “C” rater nine girl students scored 40 and 40 above, seven girl students scored 50 and 50 above and no one student scored 60 and 60 above. So scores were similar from raters to raters, therefore inter rater scoring reliability of girls was more reliable.

#### 4.4.2 Comparison among Three Sets of Scores of the Boys Students of School 'A' and 'B'.

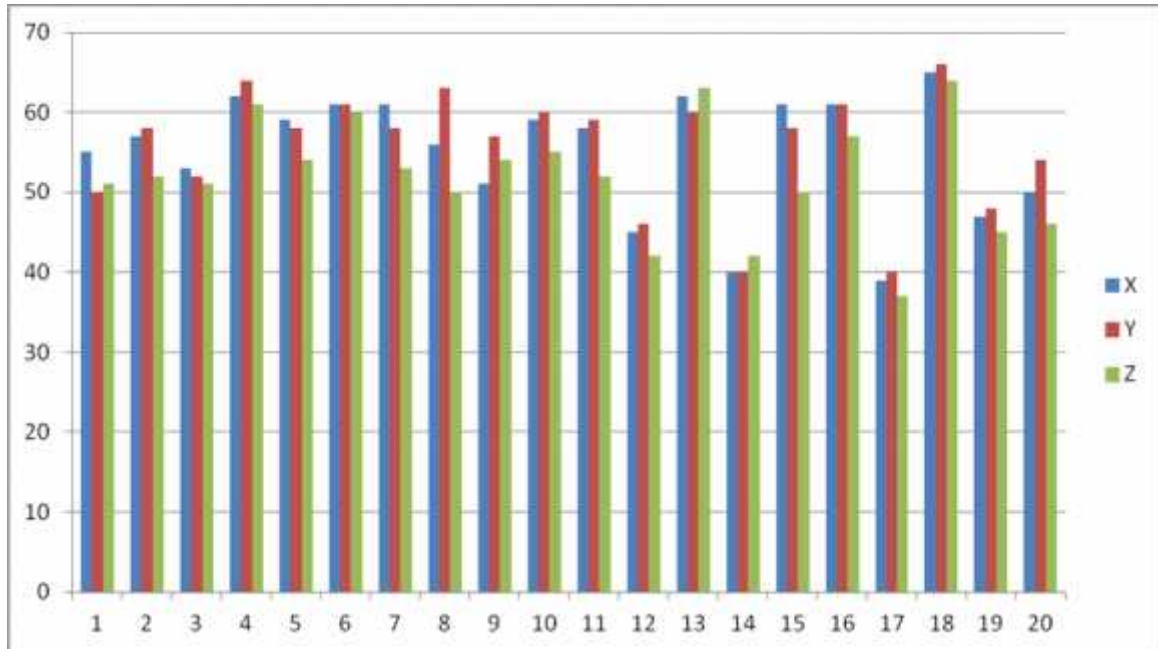


Figure: 4 Comparison among Three Sets of Scores of the Boys Students of School 'A' and 'B'.

The figure reveals that all the twenty students obtained 52% above in the inter-rating of scoring of DLE English examination test. For this I calculated percentage of score where lowest score was 36 out of full mark 70. There was in rating of A rater, five boys students scored 40 and 40 above, ten boys students scored 50 and 50 above and seven boys students scored 60 and 60 above. In rating of B, rater four boy students scored 40 and 40 above, ten boys students scored 50 and 50 above and seven students scored 60 and 60 above. In rating of C rater six boys students scored 40 and 40 above, twelve boys girl students scored 50 and 50 above and four boys girl student scored 60 and 60 above. So scores were not so varied raters to raters, therefore inter-rater scoring reliability of boys was reliable.

## 4.5 The Scorer-Wise Scoring Comparison

I presented scorer wise figure and table below.

### 4.5.1 Comparison among Three Sets of Scores of the all Students of both Schools.

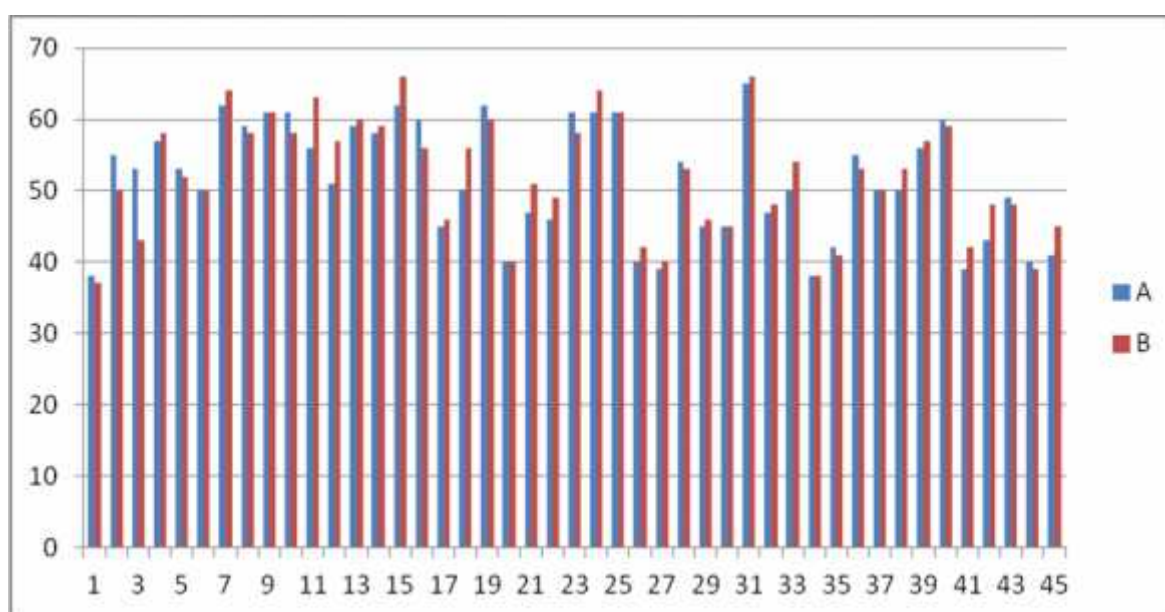


Figure: 5 Comparison among three sets of scores of the students of both schools

Table – 4 Comparison among three sets of scores of the students of both schools

Raters	No. of students			
	Number of students obtaining marks below 40	Number of students obtaining marks between 40 to 49	Number of students obtaining marks between 50 to 59	Number of students obtaining marks above 60
A	4	13	17	11
B	5	12	19	9
C	7	15	19	4

The figure and table reveal that all the forty five students obtained 51% above in the inter-rating of scoring of DLE English examination test. For this I calculated percentage of score where lowest mark was 36 out of full marks of 70. There was in rating of A rater, four students scored below 40 marks, thirteen students scored 40 and 40 marks above, seventeen students scored 50 and 50 above and eleven students scored 60 and 60 marks above. In rating of B rater, five students scored below 40 marks, twelve students scored 40 and 40 above, nineteen students scored 50 and 50 above and nine students scored 60 and 60 marks above. In rating of C rater, seven students scored below 40 marks, fifteen students scored 40 and 40 marks above, nineteen students scored 50 and 50 marks above and four student scored 60 and 60 marks above. So scores were less varied raters to raters, therefore inter-rater scoring reliability of three raters 'A', 'B' and 'C' was reliable.

#### 4.5.2 Comparison between the Two Sets of Scores of the Students of both School Scored by Rater 'A' and 'B'.

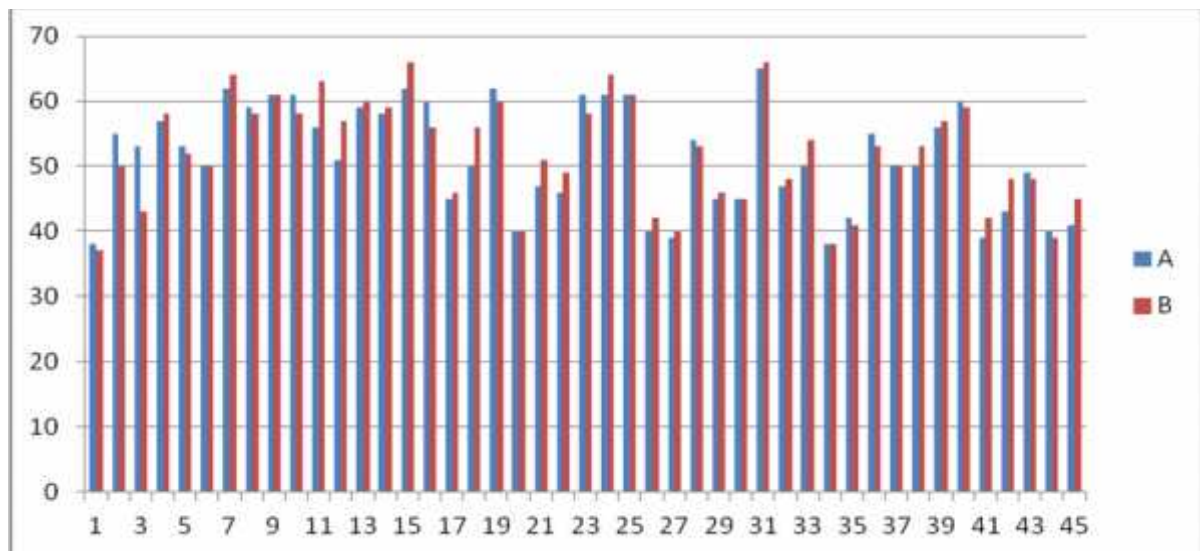


Figure: 6 Comparison between the Two Sets of Scores of the Students of both School Scored by Rater 'A' and 'B'

As seen in the figure, all the forty five students who obtained 54% above marks in the DLE examination in rating of rater “A”, could also obtain 54% above marks in rating of rater “B” because in both’s scoring the lowest mark was 38 out of 70 full marks. The highest mark percentage in rater “A” was 92% and 94% in rater “B” where the highest score was 65 and 66 respectively to both raters. This showed that the DLE question papers had high reliability in DLE English question test in both raters.

#### 4.5.3 Comparison between Two Sets of Scores of the Students of both School Scored by Rater ‘B’ and ‘C’.

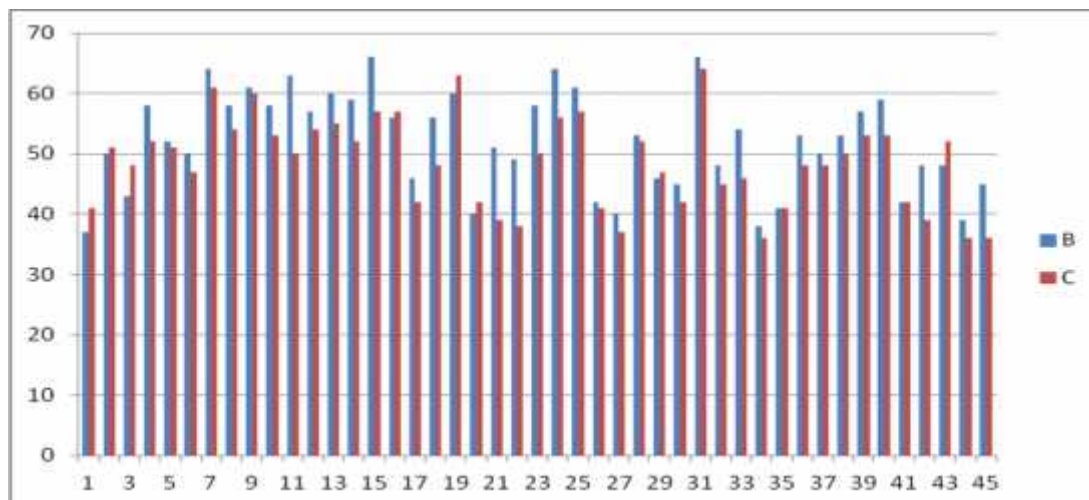


Figure: 7 Comparison between Two Sets of Scores of the Students of both School Scored by Rater ‘B’ and ‘C’

As seen in the figure all the forty five students who obtained 54% above marks in the DLE examination in rating of rater “B”, could also obtain 51% above marks in rating of rater “C” where the lowest mark was 38 and 36 in both’s rating respectively out of 70 full marks. The highest mark percentage in rater “B” was 94% and 91% in rater “C” where the highest mark was 65 and 64 respectively to scorers. This showed that the DLE question papers had high reliability in DLE English question test in both raters.

#### 4.5.4 Comparison between Two Sets of Scores of the Students of both School Scored by Rater ‘C’ and ‘A’.

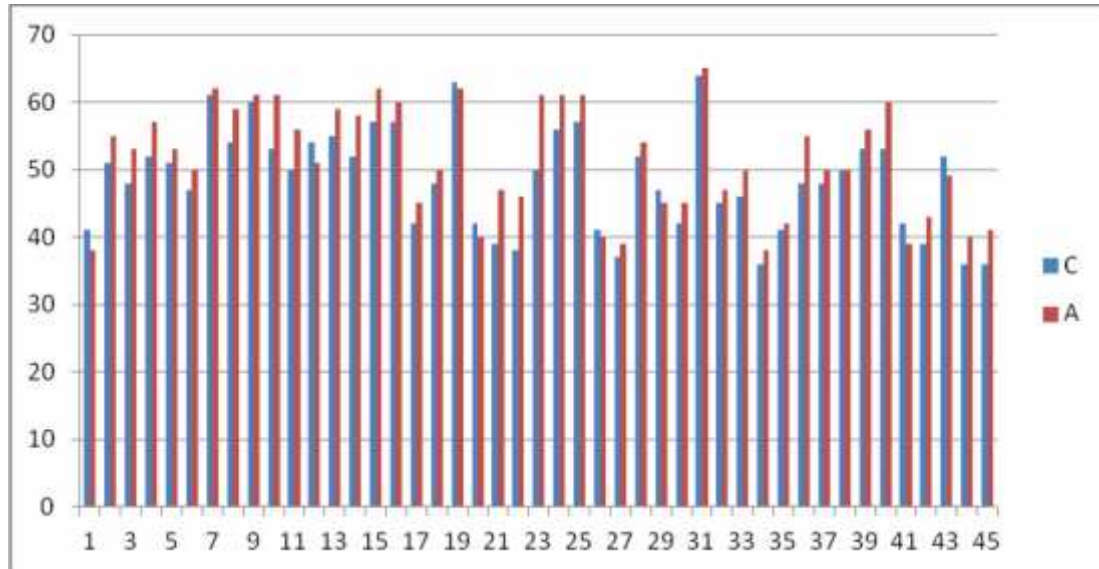


Figure: 8 Comparison between Two Sets of Scores of the Students of both School Scored by Rater ‘C’ and ‘A’

In the figure, all the forty five students who obtained 51% above marks in the DLE examination in rating of rater “C”, could also obtain 54% above marks in rating of rater “A” where the lowest mark was 36 and 38 respectively to scorers out of 70 full mark. The highest mark percentage in rater “C” was 91% and 92% in rater “B” where the highest mark was 64 and 65 respectively to scorers out of full marks 70. This showed that the DLE question papers had high reliability in DLE English question test in both raters.

The raw data used in this research has been presented in the Appendix-II and III in detail. The co relation co-efficient two sets of scores has been analyzed and interpreted under the following headings..

- a) Scorer wise co-relational analysis.
- b) Sex wise co-relational analysis.
- c) School wise co-relation analysis

## 4.6 Scorer-Wise Correlation Analysis

The co-relational analysis of the two sets of scores of the students rated by three raters “A” , “B” and “C” in terms of calculating “A” and “B”, “B’ and “C” and “C” and “A” from two schools has been done using the formula of ‘Co-relation Co-efficient’.

### 4.6.1 Co-relational Analysis of the Two Sets of Scores of Students rated by Rater “A” and “B” from two Schools

*Table - 5 Co-relational Analysis of the Two Sets of Scores of Students rated by Rater ‘A’ and ‘B’ from two Schools.*

X	Y	x=X-A "A"=51	y=Y-A "A"=52	x <sup>2</sup>	y <sup>2</sup>	Xy	r= 0.83
38	37	-13	-15	169	225	195	
55	50	4	-2	16	4	-8	
53	43	1	-9	1	81	-9	
57	58	6	6	36	36	36	
53	52	2	0	4	0	0	
50	50	-1	-2	1	4	2	
62	64	11	12	121	144	132	
589	58	8	6	64	36	48	
61	61	10	9	100	81	90	
61	58	10	6	100	36	60	
56	63	5	11	25	121	55	
51	57	0	5	0	25	0	
59	60	8	8	64	64	64	
58	59	7	7	49	49	49	
62	66	11	14	121	196	154	
60	56	9	4	81	14	36	
45	46	-5	-6	25	36	30	
50	56	-1	4	1	16	-4	

62	60	11	8	121	64	88	
40	40	-11	-12	121	144	132	
478	51	-4	-1	16	1	4	
46	49	-5	-3	25	121	15	
61	58	10	6	100	36	60	
61	64	10	12	100	144	120	
61	61	10	9	100	81	90	
40	42	-11	-10	121	100	110	
39	40	-12	-12	144	144	144	
54	53	3	1	9	1	3	
45	46	-6	-6	36	36	36	
45	45	-6	-7	36	49	42	
65	66	14	14	196	196	196	
47	48	-4	-4	16	16	16	
50	54	-1	2	1	4	-2	
38	38	-13	-14	169	196	182	
42	41	-9	-12	81	144	108	
55	53	4	1	16	1	4	
50	50	-1	-2	1	4	2	
50	53	-1	1	1	1	1	
56	57	5	5	25	25	25	
60	59	9	7	81	49	63	
39	42	-12	-10	144	100	-120	
43	48	-8	-4	64	16	32	
49	48	-2	-4	4	16	8	
40	39	-11	-13	121	169	143	
41	45	-10	-7	100	49	70	
$\Sigma X=2316$	$\Sigma Y=2344$	$\Sigma x=21$	$\Sigma y=3$	$\Sigma x^2=2927$	$\Sigma y^2=3075$	$\Sigma xy=2502$	

From above table shows that the co-efficient between two sets of scores scored by rater “A” and “B” of the the subject English of 45 students two schools is determined to be 0.83. This indicates that the degree of co-relation between two results is found very high and the direction of their relation is positive. The scoring reliability between scorers “A” nd “B” is very reliable.

#### 4.6.2 Co-relational Analysis of the Two Sets of Scores of Students rated by Rater “B” and “C” from two Schools

Table - 6 Co-relational Analysis of the Two Sets of Scores of Students rated by Rater “B” and “C” from two Schools

Y	Z	y=Y-A "A"=52	z=Z-A "A"=49	y <sup>2</sup>	z <sup>2</sup>	yz	r= 0.84
37	41	-15	-8	225	64	120	
50	51	-2	2	4	4	-4	
43	48	-9	-1	81	1	9	
58	52	6	3	36	9	18	
52	51	0	2	0	4	0	
50	47	-2	-2	4	4	4	
64	61	12	12	144	144	144	
58	54	6	5	36	25	30	
61	60	9	11	81	121	99	
58	53	6	4	36	16	24	
63	50	11	1	121	1	11	
57	54	5	5	25	25	25	
60	55	8	6	64	36	48	
59	52	7	3	49	9	21	
66	57	14	8	196	64	112	
56	57	4	8	14	64	32	
46	42	-6	-7	36	49	42	

56	48	4	-1	16	1	-4	
60	63	8	14	64	196	112	
40	42	-128	-7	144	49	84	
51	39	-1	-10	1	100	10	
49	38	-3	-11	121	121	33	
58	50	6	1	36	1	6	
64	56	12	7	144	49	84	
61	57	9	8	81	64	72	
42	41	-10	-8	100	64	80	
40	37	-12	-12	144	144	144	
53	52	1	3	1	9	3	
46	47	-6	-2	36	4	12	
45	42	-7	-7	49	49	49	
66	64	14	15	196	225	210	
48	45	-4	-4	16	16	16	
54	46	2	-3	4	9	-6	
38	36	-14	-13	196	169	182	
41	41	-12	-8	144	64	96	
53	48	1	-1	1	1	-1	
50	48	-2	-1	4	1	2	
53	50	1	1	1	1	1	
57	53	5	4	25	16	20	
59	53	7	4	49	16	28	
42	42	-10	-7	100	49	70	
48	39	-4	-10	16	100	40	
48	52	-4	3	16	9	-12	
39	36	-13	-13	169	169	169	
45	36	-7	-13	49	169	91	
$\Sigma Y=2344$	$\Sigma Z=2186$	$\Sigma y=3$	$\Sigma z=-19$	$\Sigma y^2=3075$	$\Sigma z^2=2505$	$\Sigma yz=2326$	

From above table shows that the co-efficient between two sets of scores scored by rater “B” and “C” of the the subject English of 45 students two schools is determined to be 0.84. This indicates that the degree of co-relation between two results is found very high and the direction of their relation is positive. The scoring reliability between scorers “B” nd “C” is very reliable.

### 4.6.3 Co-relational Analysis of the Two Sets of Scores of Students rated by Rater “C” and “A” from two Schools

*Table - 7 Co-relational Analysis of the two Sets of Scores of Students rated by Rater C and A from two Schools*

<b>Z</b>	<b>X</b>	<b>z=Z-A "A"=49</b>	<b>x=X-A "A"=51</b>	<b>z<sup>2</sup></b>	<b>x<sup>2</sup></b>	<b>zx</b>	<b>r=0.92</b>
41	38	-8	-13	64	169	104	
51	55	2	4	4	16	8	
48	53	-1	1	1	1	-1	
52	57	3	6	9	36	18	
51	53	2	2	4	4	4	
47	50	-2	-1	4	1	2	
61	62	12	11	144	121	132	
54	59	5	8	25	64	40	
60	61	11	10	121	100	110	
53	61	4	10	16	100	40	
50	56	1	5	1	25	5	
54	51	5	0	25	0	0	
55	59	6	8	36	64	48	
52	58	3	7	9	49	21	
57	62	8	11	64	121	88	
57	60	8	9	64	81	72	
42	45	-7	-5	49	25	35	

48	50	-1	-1	1	1	1	
63	62	14	11	196	121	154	
42	40	-7	-11	49	121	77	
39	47	-10	-4	100	16	40	
38	46	-11	-5	121	25	55	
50	61	1	10	1	100	10	
56	61	7	10	49	100	70	
57	61	8	10	64	100	80	
41	40	-8	-11	64	121	88	
37	39	-12	-12	144	144	144	
52	54	3	3	9	9	9	
47	45	-2	-6	4	36	12	
42	45	-7	-6	49	36	42	
64	65	15	14	225	196	210	
45	47	-4	-4	16	16	16	
46	50	-3	-1	9	1	3	
36	38	-13	-13	169	169	169	
41	42	-8	-9	64	81	72	
48	55	-1	4	1	16	-4	
48	50	-1	-1	1	1	1	
50	50	1	-1	1	1	-1	
53	56	4	5	16	25	20	
53	60	4	9	16	81	36	
42	39	-7	-12	49	144	84	
39	43	-10	-8	100	64	80	
52	49	3	-2	9	4	-6	
36	40	-13	-11	169	121	143	
36	41	-13	-10	169	100	130	
$\Sigma Z=2186$	$\Sigma X=2316$	$\Sigma z=-19$	$\Sigma x=21$	$\Sigma z^2=2505$	$\Sigma x^2=2927$	$\Sigma zx=2461$	

The table shows that the co-efficient between two sets of scores scored by rater “C” and “A” of 45 students of two schools is determined to be 0.92. This indicates that the degree of co-relation between two results is found very high and the direction of their relation is positive. Then, the scoring reliability between scorers “C” and “A” is very reliable.

#### 4.7 Sex-Wise Correlation Analysis

The co-relational analysis of the two sets of sex (Girl and Boy) students rated by three raters “A”, “B” and “C” in terms of calculating “A” and “B”, “B’ and “C” and “C” and “A” from two schools has been done using the formula of ‘Co-relation Co-efficient’.

##### 4.7.1 Co-relational Analysis of the Two Sets of Scores of Girl

##### Students rated by Rater “A” and “B”

Table - 8 Co-relational Analysis of the Two Sets of Scores of Girl Students rated by Rater “A” and “B”

X	Y	x=X-A "A"=50	y=Y-A "A"=51	x <sup>2</sup>	y <sup>2</sup>	xy	r= <b>0.89</b>
38	37	-12	-14	144	196	168	
53	43	3	-3	9	9	-9	
50	50	0	-4	0	16	0	
62	66	12	6	144	36	72	
60	56	10	6	100	36	60	
50	56	0	-3	0	9	0	
47	51	-3	-12	9	144	36	
46	49	-4	-13	16	169	52	
61	64	11	5	121	25	55	
40	42	-10	-9	100	81	90	
54	53	4	2	16	4	8	
45	46	-5	-5	25	25	25	
45	45	-5	-6	25	36	30	
55	53	5	2	25	4	10	
50	50	0	-3	0	9	0	
56	57	6	2	36	4	12	
60	59	10	2	100	4	20	
43	48	-7	-12	49	144	84	
49	48	-1	1	1	1	-1	
41	45	-5	-6	25	36	30	
ΣX=1005	ΣY=1018	Σx= 9	Σy= -64	Σx <sup>2</sup> =945	Σy <sup>2</sup> =988	Σxy= 742	

The table shows that the co-efficient between two sets of scores scored by rater “A” and “B” of the subject English of 20 girl students two schools is determined to be 0.89. This indicates that the degree of co-relation between two results is found very high and the direction of their relation is positive. The scoring reliability between rater “A” and “B” is very reliable.

#### 4.7.2 Co-relational Analysis of the Two Sets of Scores of Girl Students rated by Rater “B” and “C”

Table - 9 Co-relational Analysis of the Two Sets of Scores of Girl Students rated by Rater “B” and “C”

Y	Z	y=Y-A "A"=51	z=Z-A "A"=47	y <sup>2</sup>	z <sup>2</sup>	yz	r= 0.92
37	41	-14	-6	196	36	84	
43	48	-3	1	9	1	-3	
50	47	-4	0	16	0	0	
66	57	6	10	36	100	60	
56	57	6	10	36	100	60	
56	48	-3	1	9	1	-3	
51	39	-12	-8	144	64	96	
49	38	-13	-9	169	81	117	
64	56	5	9	25	81	45	
42	41	-9	-6	81	36	54	
53	52	2	5	4	25	10	
46	47	-5	0	25	0	0	
45	42	-6	-5	36	25	30	
53	48	2	1	4	1	2	
50	48	-3	1	9	1	-3	
57	53	2	6	4	36	12	
59	53	2	6	4	36	12	
48	39	-12	-8	144	64	96	
48	52	1	5	1	25	5	
45	36	-6	-11	36	121	66	
ΣX=1018	ΣZ=942	Σy=-64	Σz=2	Σy <sup>2</sup> =988	Σz <sup>2</sup> =834	Σyz=740	

The table shows that the co-efficient between two sets of scores scored by rater “B” and “C” of the subject English of 20 girl students of two schools is determined to be 0.92. This indicates that the degree of co-relation between two results is

found very high and the direction of their relation is positive. The scoring reliability between rater “B” and “C” is very reliable.

### 4.7.3 Co-relational Analysis of the Two Sets of Scores of Girl Students rated by Rater “C” and “A”

Table - 10 Co-relational Analysis of the Two Sets of Scores of Girl Students rated by Rater “C” and “A”

Z	X	$z=Z-A$ "A"=47	$x=X-A$ "A"=50	$z^2$	$x^2$	$zx$	<b>r=</b> <b>0.86</b>
41	38	-6	-12	36	144	72	
48	53	1	3	1	9	3	
47	50	0	0	0	0	0	
57	62	10	12	100	144	120	
57	60	10	10	100	100	100	
48	50	1	0	1	0	0	
39	47	-8	-3	64	9	24	
38	46	-9	-4	81	16	36	
56	61	9	11	81	121	99	
41	40	-6	-10	36	100	60	
52	54	5	4	25	16	20	
47	45	0	-5	0	25	0	
42	45	-5	-5	25	25	25	
48	55	1	5	1	25	5	
48	50	1	0	1	0	0	
53	56	6	6	36	36	36	
53	60	6	10	36	100	60	
39	43	-8	-7	64	49	56	
52	49	5	-1	25	1	-5	
36	41	-11	-5	121	25	55	
$\Sigma Z=942$	$\Sigma X=1005$	$\Sigma z=2$	$\Sigma x=9$	$\Sigma z^2=834$	$\Sigma x^2=945$	$\Sigma zx=766$	

The table shows that the co-efficient between two sets of scores scored by rater “C” and “A” of the subject English of 20 girl students two schools is determined to be 0.86. This indicates that the degree of co-relation between two results is found very high and the direction of their relation is positive. The scoring reliability between rater “C” and “A” is very reliable.

#### 4.7.4 Co-relational Analysis of the Two Sets of Scores of Boy Students rated by Rater “A” and “B”

Table - 11 Co-relational Analysis of the Two Sets of Scores of Boy Students rated by Rater “A” and “B”

X	Y	x=X-A "A"=5 2	y=Y-A "A"=53	x <sup>2</sup>	y <sup>2</sup>	xy	r= 0.95
55	50	3	-3	9	9	-9	
57	58	5	5	25	25	25	
53	52	1	-1	1	1	-1	
62	64	10	11	100	121	110	
59	58	7	5	49	25	35	
61	61	9	8	91	64	72	
61	58	9	5	81	25	45	
56	63	4	10	16	100	40	
51	57	-1	4	1	16	-4	
59	60	7	7	49	49	49	
58	59	6	6	36	36	36	
45	46	-7	-7	49	49	49	
62	60	10	7	100	49	70	
40	40	-12	-13	144	169	156	
61	58	9	5	81	25	45	
61	61	9	8	81	64	72	
39	40	-13	-13	169	169	169	
65	66	13	13	169	169	169	
47	48	-5	-5	25	25	25	
50	54	2	1	4	1	2	
38	38	-14	-15	196	225	210	
42	41	-10	-12	100	144	120	
50	53	-2	0	2	0	0	
39	42	-13	-11	169	121	143	
40	39	-12	-14	144	196	168	
$\Sigma X=1311$	$\Sigma Y=1326$	$\Sigma x=15$	$\Sigma y=1$	$\Sigma x^2=1891$	$\Sigma y^2=1877$	$\Sigma xy=1796$	

The table shows that the co-efficient between two sets of scores scored by rater “A” and “B” of the subject English of 25 boys students two schools is determined to be 0.95. This indicates that the degree of co-relation between two results is found very high and the direction of their relation is positive. The scoring reliability between rater “A” and “B” is very reliable.

#### 4.7.5 Co-relational Analysis of the Two Sets of Scores of Boy Students rated by Rater “B” and “C”

Table - 12 Co-relational Analysis of the Two Sets of Scores of Boy Students rated by Rater “B” and “C”

Y	Z	y=Y-A "A"=53	z=Z-A "A"=50	y <sup>2</sup>	z <sup>2</sup>	yz	r= <b>0.80</b>
50	51	-3	1	9	1	-3	
58	52	5	2	25	4	10	
52	51	-1	1	1	1	-1	
64	61	11	11	121	121	11	
58	54	5	4	25	16	20	
61	60	8	10	64	100	80	
58	53	5	3	25	9	15	
63	50	10	0	100	0	0	
57	54	4	4	16	16	16	
60	55	7	5	49	25	35	
59	52	6	2	36	2	12	
46	42	-7	-8	49	64	56	
60	63	7	13	49	169	91	
40	42	-13	-8	169	64	104	
58	50	5	0	25	0	0	
61	57	8	7	64	49	56	
40	37	-13	-13	169	169	69	
66	64	13	14	169	194	182	
48	45	-5	-5	25	25	25	
54	46	1	-4	1	16	-4	
38	36	-15	-14	225	196	210	
41	41	-12	-11	144	121	132	
53	50	0	0	0	0	0	
42	42	-11	-8	121	64	88	
39	36	-14	-14	196	196	196	
ΣY=1326	ΣZ=1244	Σy=1	Σz=-8	Σy <sup>2</sup> =1877	Σz <sup>2</sup> =1622	Σyz=1400	

The table shows that the co-efficient between two sets of scores scored by rater “B” and “C” of the subject English of 25 boy students of two schools is determined to be 0.80. This indicates that the degree of co-relation between two results is found very high and the direction of their relation is positive. The scoring reliability between rater “B” and “C” is very reliable.

#### 4.7.6 Co-relational Analysis of the Two Sets of Scores of Boy Students rated by Rater “C” and “A”

Table - 13 Co-relational Analysis of the Two Sets of Scores of Boy Students rated by Rater “C” and “A”

Z	X	$z=Z-A$ "A"=50	$x=X-A$ "A"=52	$z^2$	$x^2$	zx	r= <b>0.86</b>
51	55	1	3	1	9	3	
52	57	2	5	4	25	10	
51	53	1	1	1	1	1	
61	62	11	10	121	100	110	
54	59	4	7	16	49	28	
60	61	10	9	100	91	90	
53	61	3	9	9	81	27	
50	56	0	4	0	16	0	
54	51	4	-1	16	1	-4	
55	59	5	7	25	49	35	
52	58	2	6	4	36	12	
42	45	-8	-7	64	49	56	
63	62	13	10	169	100	130	
42	40	-8	-12	64	144	96	
50	61	0	9	0	81	0	
57	61	7	9	49	81	63	
37	39	-13	-13	169	169	69	
64	65	14	13	194	169	182	
45	47	-5	-5	25	25	25	
46	50	-4	2	16	4	-8	
36	38	-14	-14	196	196	196	
41	42	-11	-10	121	100	110	
50	50	0	-2	0	4	0	
42	39	-8	-13	64	169	104	
36	40	-14	-12	196	144	168	
$\Sigma Z=1244$	$\Sigma X=1311$	$\Sigma z=-8$	$\Sigma x=15$	$\Sigma z^2=1622$	$\Sigma x^2=1891$	$\Sigma zx=1503$	

The table shows that the co-efficient between two sets of scores scored by rater “C” and “A” of the subject English of 25 boy students two schools is determined to be 0.86. This indicates that the degree of co-relation between two results is found very high and the direction of their relation is positive. The scoring reliability between rater “C” and “A” is very reliable.

## 4.8 School-Wise Correlation Analysis

The co-relational analysis of the students of two school rated by three raters “A” , “B” and “C” in terms of calculating “A” and “B”, “B’ and “C” and “C” and “A” has been done using the formula of ‘Co-relation Co-efficient’.

### 4.8.1 Co-relational Analysis of the Two Sets of Scores of Students of School “A” rated by Rater “A” and “B”

Table - 14 Co-relational Analysis of the Two Sets of Scores of Students of School “A” rated by Rater “A” and “B”

X	Y	x=X-A "A"=55	y=Y-A "A"=55	x <sup>2</sup>	y <sup>2</sup>	xy	r= 0.87
38	37	-17	-18	289	324	306	
55	50	0	-5	0	25	0	
53	43	-2	-12	4	144	24	
57	58	2	3	4	9	6	
53	52	-2	-3	4	9	6	
50	50	-5	-5	25	25	25	
62	64	7	9	49	81	63	
59	58	4	3	16	9	12	
61	61	6	6	36	36	36	
61	58	6	3	36	9	18	
56	63	1	8	1	64	8	
51	57	-4	2	16	0	-8	
59	60	4	5	16	25	20	
58	59	3	4	9	16	12	
62	66	7	11	49	121	77	
60	56	5	1	25	1	5	
45	46	-10	-9	100	81	90	
50	56	-5	1	25	1	-5	
62	60	7	5	49	25	35	
40	40	-15	-15	225	225	225	
47	51	-8	-4	64	16	32	
46	49	-9	-6	81	36	54	
61	58	6	3	36	9	18	
61	64	6	9	36	81	54	
61	61	6	6	36	36	36	
$\Sigma X=1368$	$\Sigma Y= 1377$	$\Sigma x= -7$	$\Sigma y=2$	$\Sigma x^2=1231$	$\Sigma y^2= 1408$	$\Sigma xy= 1149$	

The table shows that the co-efficient between two sets of scores scored by rater “A” and “B” of the subject English of 25 students of school “A” is determined to be 0.87. This indicates that the degree of co-relation between two results is found very high and the direction of their relation is positive. The scoring reliability between rater “A” and “B” is very reliable.

#### 4.8.2 Co-relational Analysis of the Two Sets of Scores of Students of School “A” rated by Rater “B” and “C”

Table - 15 Co-relational Analysis of the Two Sets of Scores of Students of School “A” rated by Rater “B” and “C”

Y	Z	y=Y-A "A"=55	z=Z-A "A"=51	y <sup>2</sup>	z <sup>2</sup>	yz	r= 0.78
37	41	-18	-11	324	121	198	
50	51	-5	0	25	0	0	
43	48	-12	-3	144	9	36	
58	52	3	1	9	1	3	
52	51	-3	0	9	0	0	
50	47	-5	-4	25	16	20	
64	61	9	10	81	100	90	
58	54	3	3	9	9	9	
61	60	6	9	36	81	54	
58	53	3	2	9	4	6	
63	50	8	-1	64	1	-8	
57	54	2	3	0	9	6	
60	55	5	4	25	16	20	
59	52	4	1	16	1	4	
66	57	11	6	121	36	66	
56	57	1	6	1	36	6	
46	42	-9	-9	81	81	81	
56	48	1	-3	1	9	-3	
60	63	5	12	25	144	60	
40	42	-15	-9	225	81	135	
51	39	-4	-12	16	144	48	
49	38	-6	-13	36	169	78	
58	50	3	-1	9	1	-3	
64	56	9	5	81	25	45	
61	57	6	6	36	36	36	
$\Sigma Y=1377$	$\Sigma Z=1278$	$\Sigma y=2$	$\Sigma z=2$	$\Sigma y^2= 1408$	$\Sigma z^2=1130$	$\Sigma yz=987$	

The table shows that the co-efficient between two sets of scores scored by rater “B” and “C” of the subject English of 25 students of school “A” is determined to be 0.78. This indicates that the degree of co-relation between two results is found very high and the direction of their relation is positive. The scoring reliability between rater “B” and “C” is very reliable.

#### 4.8.3 Co-relational Analysis of the Two Sets of Scores of Students of School “A” rated by Rater “C” and “A”

Table - 16 Co-relational Analysis of the Two Sets of Scores of Students of School “A” rated by Rater “C” and “A”

Z	X	z=Z-A "A"=51	x=X-A "A"=55	z <sup>2</sup>	x <sup>2</sup>	zx	r= 0.88
41	38	-11	-17	121	289	187	
51	55	0	0	0	0	0	
48	53	-3	-2	9	4	6	
52	57	1	2	1	4	2	
51	53	0	-2	0	4	0	
47	50	-4	-5	16	25	20	
61	62	10	7	100	49	70	
54	59	3	4	9	16	12	
60	61	9	6	81	36	54	
53	61	2	6	4	36	12	
50	56	-1	1	1	1	-1	
54	51	3	-4	9	16	-12	
55	59	4	4	16	16	16	
52	58	1	3	1	9	3	
57	62	6	7	36	49	42	
57	60	6	5	36	25	30	
42	45	-9	-10	81	100	90	
48	50	-3	-5	9	25	15	
63	62	12	7	144	49	84	
42	40	-9	-15	81	225	135	
39	47	-12	-8	144	64	96	
38	46	-13	-9	169	81	117	
50	61	-1	6	1	36	-6	
56	61	5	6	25	36	30	
57	61	6	6	36	36	36	
ΣZ=1278	ΣX=1368	Σz=2	Σx=-7	Σz <sup>2</sup> =1130	Σx <sup>2</sup> =1231	Σzx=1038	

The table shows that the co-efficient between two sets of scores scored by rater “C” and “A” of the subject English of 25 students of school “A” is determined to be 0.88. This indicates that the degree of co-relation between two results is found very high and the direction of their relation is positive. The scoring reliability between rater “C” and “A” is very reliable.

#### 4.8.4 Co-relational Analysis of the Two Sets of Scores of Students of School “B” rated by Rater “A” and “B”

Table - 17 Co-relational Analysis of the Two Sets of Scores of Students of School “B” rated by Rater “A” and “B”

X	Y	x=X-A "A"=47	y=Y-A "A"=48	x <sup>2</sup>	y <sup>2</sup>	xy	r= 0.98
40	42	-7	-6	49	36	42	
39	40	-8	-8	64	64	64	
54	53	7	5	49	25	35	
45	46	-2	-2	4	4	4	
45	45	-2	-3	4	9	6	
65	66	18	18	324	324	324	
47	48	0	0	0	0	0	
50	54	3	6	9	36	18	
38	38	-9	-10	81	100	90	
42	41	-5	-7	25	49	35	
55	53	8	5	64	25	40	
50	50	3	2	9	4	6	
50	53	3	5	9	25	35	
56	57	9	9	81	81	81	
60	59	13	11	169	121	143	
39	42	-8	-6	64	36	48	
43	48	-4	0	16	0	0	
49	48	2	0	4	0	0	
40	39	-7	-9	49	81	63	
41	45	-6	-3	36	9	18	
ΣX=948	ΣY=967	Σx=8	Σy=7	Σx <sup>2</sup> =1110	Σy <sup>2</sup> =1029	Σxy=1052	

The table shows that the co-efficient between two sets of scores scored by rater “A” and “B” of the subject English of 20 students of school “B” is determined to be 0.98. This indicates that the degree of co-relation between two results is found

very high and the direction of their relation is positive. The scoring reliability between rater “A” and “B” is very reliable.

#### 4.8.5 Co-relational Analysis of the Two Sets of Scores of Students of School “B” rated by Rater “B” and “C”

Table - 18 Co-relational Analysis of the Two Sets of Scores of Students of School “B” rated by Rater “B” and “C”

Y	Z	y=Y-A "A"=48	z=Z-A "A"=45	y <sup>2</sup>	z <sup>2</sup>	yz	r= 0.90
42	41	-6	-4	36	16	24	
40	37	-8	-8	64	64	64	
53	52	5	7	25	49	35	
46	47	-2	2	4	4	-4	
45	42	-3	-3	9	9	9	
66	64	18	19	324	361	342	
48	45	0	0	0	0	0	
54	46	6	1	36	1	6	
38	36	-10	-9	100	81	90	
41	41	-7	-4	49	16	28	
53	48	5	3	25	9	15	
50	48	2	3	4	9	6	
53	50	5	5	25	25	25	
57	53	9	8	81	64	72	
59	53	11	8	121	64	88	
42	42	-6	-3	36	9	18	
48	39	0	-6	0	36	0	
48	52	0	7	0	49	0	
39	36	-9	-9	81	81	81	
45	36	-3	-9	9	81	27	
ΣY=967	ΣZ=908	Σy=7	Σz=8	Σy <sup>2</sup> =1029	Σz <sup>2</sup> =1028	Σyz=926	

The table shows that the co-efficient between two sets of scores scored by rater “B” and “C” of the subject English of 20 students of school “B” is determined to be 0.90. This indicates that the degree of co-relation between two results is found very high and the direction of their relation is positive. The scoring reliability between rater “B” and “C” is very reliable.

#### 4.8.6 Co-relational Analysis of the Two Sets of Scores of Students of School “B” rated by Rater “C” and “A”

Table - 19 Co-relational Analysis of the Two Sets of Scores of Students of School “B” rated by Rater “C” and “A”

Z	X	$z=Z-A$ "A"=45	$x=X-A$ "A"=47	$z^2$	$x^2$	zx	r= 0.92
41	40	-4	-7	16	49	28	
37	39	-8	-8	64	64	64	
52	54	7	7	49	49	49	
47	45	2	-2	4	4	-4	
42	45	-3	-2	9	4	6	
64	65	19	18	361	324	342	
45	47	0	0	0	0	0	
46	50	1	3	1	9	3	
36	38	-9	-9	81	81	81	
41	42	-4	-5	16	25	20	
48	55	3	8	9	64	24	
48	50	3	3	9	9	9	
50	50	5	3	25	9	15	
53	56	8	9	64	81	72	
53	60	8	13	64	169	104	
42	39	-3	-8	9	64	24	
39	43	-6	-4	36	16	24	
52	49	7	2	49	4	14	
36	40	-9	-7	81	49	63	
36	41	-9	-6	81	36	54	
$\Sigma Z=908$	$\Sigma X=948$	$\Sigma z=8$	$\Sigma x=8$	$\Sigma z^2=1028$	$\Sigma x^2=1110$	$\Sigma zx=992$	

The table shows that the co-efficient between two sets of scores scored by rater “C” and “A” of the subject English of 20 students of school “B” is determined to be 92. This indicates that the degree of co-relation between two results is found very high and the direction of their relation is positive. The scoring reliability between rater “C” and “A” is very reliable.

## **CHAPTER- V**

### **FINDINGS AND RECOMMENDATIONS**

#### **5.1 Results and Interpretation**

The following are the results of the study.

1. All forty five students obtained 54% above marks in the DLE examination in rating of rater “A” and “B” but 51% in scoring of rater “C”. The highest mark percentage in rater A was 92% and 94% in rater B but 91% in scoring of rater C”. This showed that the DLE question papers had high reliability in DLE English question test in all three raters.
2. The co-efficient between two sets of scores scored by rater “A” and “B” , “B” and “C” and “C” and “A” of the the subject English of 45 students two schools is determined to be 0.83, 0.84 and 0.92 respectively. This indicates that the degree of co-relation between two results is very high and the direction of their relation is positive. The scoring reliability among scorers is quite reliable.
3. The co-efficient between two sets of scores scored by rater “A” and “B” , “B” and “C” and “C” and “A” of the the subject English of 20 girls students of two schools is determined to be 0.89, 0.92 and 0.86 and 25 boys students of two schools is determined 0.95, 0.80 and 0.86 respectively. This indicates that the degree of co-relation between two results is found very high and the direction of their relation is positive. The scoring reliability among scorers is very reliable.
4. In comparison, it was found that scoring reliability of DLE English question test on more raters is very high on both girl and boy students on sex wise.
5. In terms of raters, co-relation between two raters was found very high.

6. No one students of both school were failed in DLE English examination. All passed the exam securing more than 51% scores.
7. The co-efficient between two sets of scores scored by rater “A” and “B” , “B” and “C” and “C” and “A” of the the subject English of a school having 25 students is determined to be 0.87, 0.78 and 0.88 and having 20 students is determined 0.98, 0.90 and 0.92 respectively. This indicates that the degree of co-relation between two results is found very high and the direction of their relation is positive. The scoring reliability among scorers is very reliable.
8. In comparison, rater C has given less mean scores than rater A and B.
9. Factors influencing the scoring of DLE test are: lack of trainings of scoring on raters, lack of marking scheme to scorers, negligence of scorers, lack of universal tools of scoring subjective answers, halo effects, practicing of unfair means and poor availability of resources.

## **5.2 Summery and Conclusion**

On the basis of research, scoring reliability of DLE English question test depends on the no number of raters or scorers. Both less scorers and more scorers have very high scoring reliability.

## **5.3 Implications/ Recommendations**

1. In order to create consistency in the scores, the students should be taught through problem solving methods and rote learning should be discouraged.
2. Marking scheme should be prepared along the preparation of question paper, and should be given to every DLE answer sheet scorer with orientation class about scoring method and marking scheme.
3. Question paper and curriculum should be reviewed from time to time.

4. It is not all research about scoring reliability of DLE English test; any researcher can make further research in terms of scoring more sets of question, score re-score of same test etc.
5. Resources, materials, trained, qualified teachers are lacking in both private and public school. Therefore, there should be easy access to these in the schools.
6. Internal assessment should be introduced by the concerned authority. This could go long way in bringing about scoring reliability in the test.

## REFERENCES

- Bachman, L.F. (1998) *Fundamental consideration in language testing* London; OUP
- Batala, K.B. (2004) *Validation of the SLC examination*. An Unpublished M.Ed. Thesis. Kathmandu: T.U.
- Bhattra, G.R. (2001) *A thematic analysis of research reports*. Kathmandu: Ratna Pustak Bhandar.
- Dhakal K.P (2000) *Effectiveness of discrete test and integrative test as a measure of English proficiency*
- Faculty Education (2003) *Formative research on curriculum and assesment of secondary education*. T.U., Kirtipur
- Hornby, A.S. ( 6<sup>th</sup> Edition) *Oxford advanced learners dictionary*
- Harrison, A. (1991) *A language testing hand book*. ELTS/Longman
- Heaton, J.B. (1998) *Writing English language tests*. Longman
- Hughes, A (1989) *Testing for language teachers*. CUP.
- Kandel, R.P. (2006) *A study on the co-relationship between sent-up and SLC examination results*
- Khanal U. (1999) *Effectiveness of the cloze test over conventional objective test in testing reading comprehension*
- Khaniya, T.R. (Dec. 2003) *Revisit the SLC examination*. An Article in the Kathmandu Post
- Khaniya, T.R. (March, 2004) *Qualities of school educations*. An Article Published in the Kathmandu Post.
- Kshetree , P. (2001) *A study on the effect of the SLC examination*
- Kumar, R. (1996) *Research methodology*. SAGE. London
- Lado, R. (1993) *Language testing London*: Longman

Ministry of Education and Sports,(1999) *a Secondary education curriculum*.  
CDC, Sanothimi Bhaktapur

Ministry of Education and Sports,(1999) *Course book 8*. JEMC, Sanothimi  
Bhaktapur

Ministry of Education and Sports,(1999) *a specification grid based on new  
curriculum*. CDC, Sanothimi, Bhaktapur

National Education System Plan. (NESP). 2028 B.S.

Nepal National Education Planning Commission. 2011B.S.

Pokhrel, K. (2006), *Errors in SLC question papers: A Case of grammatical item*.

Pradipt, T.B. (2008), *Reliability of SLC questions papers*

Sharma, U.N. (1999), *A study on marketing an English answer book*.

Singh, G.B. (1996). *SLC examination in Nepal*. A Critical Study Report.  
Kathmandu: CERID

Sthapit, A.B. et.al (2004) *Statistical methods*. Kathmandu:BAE Pvt. Ltd

Thapa, T.B. (2005) *Predictive validity of grade- IX English annual test*

## **APPENDIX-I**

This question paper has been prepared to draw data information for the research work entitled, "Scoring Reliability of DLE English Question Set" which is carried out under the guidance of lecturer, **Nara Prasad Bhandari**, Department of English language Education, Sukuna Multiple Campus, Indrapur, Morang. The researcher hopes that you all cooperate with him in providing authentic and reliable data/information that will be invaluable contribution to accomplish this research work.

Thank you

Researcher

**Lekharaj Shrestha**

Sukuna Multiple Campus,

Morang

**Lower Secondary Education Completion Examination 2069** LSC-3

Subject: English

Full Marks: 70  
Pass Marks: 22.5

200 Minutes

1. Fill in the blanks with correct word given in brackets:

4x3=12

- He is \_\_\_\_\_ university student. (a, an, the)
- We always go to school \_\_\_\_\_ foot. (on, by, in)
- This is the house \_\_\_\_\_ I was born. (who, which, where)
- Sirs, do it \_\_\_\_\_ . (yourself, herself, yourselves)

2. Change the following sentences as indicated in brackets: 6x2=12

- They always go to temple. (into Negative)
- Someone has stolen my bag. (into Passive)
- He killed a mosquito. (into what-question)
- She said to me, "Open the door." (into indirect speech)
- He's done all the work. (into yes/no question)
- English is spoken all over the world. (into Active)

3. Answer any five questions in your own words:

5x2=10

- What does L.W.F. stand for? (The start of the Adventure)
- Who was Ellen? (Ellen's Mountain)
- Where is Gokyo? (Lip to Gokyo)
- What is an encyclopedia? (Encyclopedia)
- How did they start flying? (The story of Icarus)
- Write any two advantages of electricity. (A power station.....)

4. Read the following passage and do the activities that follow.

Film making is one of the world's biggest industries. The largest number of films are feature films which are for entertainment and use actors and sets. Sets are specially built scenery. Documentary films about real places and people are also very popular. A new one, called "Baluon over Nepal" has just been released in Kathmandu.

Filming itself is only a small part of making a film. Films must be edited. The editor chooses short pieces of film and joins them together to make the final film. A sound editor adds sound and music to the film. This is called the soundtrack. Titles and credits are added at the beginning and end and the film is ready.

Answer these questions:

4x2=8

- Why are feature films made?
- Which film is newly released in Kathmandu?
- What does a film editor do?
- What is called the soundtrack?

5. Read the passage carefully and do the activities that follow:

8

Normally the boys or girls below 14 of their age are legally called the children in the context of Nepal. So, to employ them in any sort of physical works is illegal. "Children are in fact the dreams of their parents, the dreamy eyes of the society and further, the future starts of the nation. The future of a nation depends upon the future of its children. Therefore, the nation should ensure the child rights proper care, love and affection, education, freedom and entertainment together with fundamental requirements like housing, clothing, feeding and security.

A. Choose the right answers.

1x4=4

- To employ children in \_\_\_\_\_ is illegal.
  - mental work
  - physical work
  - house work
  - depends upon the future of its children.
- The future of a \_\_\_\_\_ depends upon the future of its children.
  - parents
  - children
  - nation
  - human
- Proper care, love and affection, education etc are examples of \_\_\_\_\_ rights.
  - child
  - woman
  - child
  - economically
- The boys and girls below fourteen are \_\_\_\_\_ called the children.
  - legally
  - socially
  - economically
  - legally

B. Answer the following questions.

1x4=4

- Is physical work of children legal? Why? Why not?
- What should a nation ensure?
- Who are children in fact?
- Make a list of children rights and fundamental requirements.

6. Use any four of the following words or phrases in sensible sentences of your own.

1x4=4

- a lot of, century, put on, frost bite, balloonist, glacier  
Punctuate the following piece of text.  
If its too cold we'll get frost bit said another.  
1x4=4
- Write a paragraph on your best teacher using the given clues below.  
name and age, physical feature, other qualities, your opinion  
6
- Write an essay on any one of the following topics.  
a) A festival you like most  
b) My aim in life  
c) Disadvantages of television  
10
- Write a letter to your friends describing about your village or town.  
10

## APPENDIX-II

In this part of appendix, I had listed name list of the students who appeared in DLE English question test.

S.N.	Name of Students	Name of School	Class	Gender
1	Hima Chaudhary	Lovely Angels Boarding School	8	F
2	Sabin Khawas	Lovely Angels Boarding School	8	M
3	Sunita Gautam	Lovely Angels Boarding School	8	F
4	Anand Shah	Lovely Angels Boarding School	8	M
5	Saroj Luitel	Lovely Angels Boarding School	8	M
6	Ganga Dangal	Lovely Angels Boarding School	8	F
7	Suman Chimariya	Lovely Angels Boarding School	8	M
8	Nikhil Shah	Lovely Angels Boarding School	8	M
9	Saroj Pandey	Lovely Angels Boarding School	8	M
10	Nabin Acharya	Lovely Angels Boarding School	8	M
11	Ashis Chaudhary	Lovely Angels Boarding School	8	M
12	Rubin Gajamer	Lovely Angels Boarding School	8	M
13	Saurav Niroula	Lovely Angels Boarding School	8	M
14	Bibek poudel	Lovely Angels Boarding School	8	M
15	Binju Basnet	Lovely Angels Boarding School	8	F
16	Sabina Khawas	Lovely Angels Boarding School	8	F
17	Manoj Khadka	Lovely Angels Boarding School	8	M
18	Jenisha Limbu	Lovely Angels Boarding School	8	F
19	Sanjaya Diswa	Lovely Angels Boarding School	8	M
20	Diwas Chaudhary	Lovely Angels Boarding School	8	M
21	Gita Dangal	Lovely Angels Boarding School	8	F
22	Punam Cahapagain	Lovely Angels Boarding School	8	F
23	Nirmal Luitel	Lovely Angels Boarding School	8	M
24	Rojina Koirala	Lovely Angels Boarding School	8	F
25	Nischal Regmi	Lovely Angels Boarding School	8	M

26	Prajita Timsina	Haraincha Model English School	8	F
27	Suraj Majhi	Haraincha Model English School	8	M
28	Rakshya Khadka	Haraincha Model English School	8	F
29	Anjila Khadka	Haraincha Model English School	8	F
30	Susmita Khadka	Haraincha Model English School	8	F
31	Sagar Poudel	Haraincha Model English School	8	M
32	Yunesh Dhakal	Haraincha Model English School	8	M
33	Hritik Rimal	Haraincha Model English School	8	M
34	Sagar Koirala	Haraincha Model English School	8	M
35	Rewat Karki	Haraincha Model English School	8	M
36	Ashmi Poudel	Haraincha Model English School	8	F
37	Asmita Mishra	Haraincha Model English School	8	F
38	Bikal Shakya	Haraincha Model English School	8	M
39	Astha Chaudhary	Haraincha Model English School	8	F
40	Anu Shrestha	Haraincha Model English School	8	F
41	Bibek Shrestha	Haraincha Model English School	8	M
42	Khem Subedi	Haraincha Model English School	8	F
43	Sajana Bista	Haraincha Model English School	8	F
44	Anand Agrahari	Haraincha Model English School	8	M
45	Nisha Bhattarai	Haraincha Model English School	8	F

### APPENDIX-III

Name: Sabin Khawas  
class ~~name~~ = Eight (8)  
Sub: English

55

(2)

1. a) He is a university student.  
b) We always go to school by foot.  
c) This is the house where I was born.  
d) Sita, do it yourself.

(3)

2.

- a) They never go to temple.  
b) My bag has been stolen.  
c) What did he kill?

d) She told me ~~that~~ to open the door.

e) Did he done ~~all~~ the work?

f) People speak English all over the world.

(4)

3.

a) Ans) The L.W.F stands for English in action competition  
Living World Film

b) Ans) The height of Mt. Annapurna is 8095m.

c) Ans) Ellen was a blind girl.

e) Ans) Encyclopedia is a such type of book where

different kinds of information are found about the world. (10)

f) Ans ⇒ They start ~~ed~~ flying by flapping their arms

4.

a) Ans ⇒ Feature films are made for entertainment and use actors and sets.

b) Ans ⇒ Balloon over Nepal film ~~has~~ newly released in Kathmandu. (8)

c) Ans ⇒ A film editor chooses short short pieces of film and joins them together to make the final film.

d) Ans ⇒ A sound editor adds sound and music to the film. This is called soundtrack.

5.  
A.

(i) To employ children in physical work is illegal.

(ii) The future of a nation depends upon the future of its children.

(11) Proper care, love and affection, education etc are examples of child rights.

(12) The boys and girls below fourteen are legally called the children.

B.

a. Ans → Yes, physical work of children is legal because in the context of Nepal to employ children in any sort of physical works is illegal.

b. Ans → The nation should ensure the child rights proper care, love and affection, education, freedom and entertainment together with fundamental requirements like housing, clothing, feeding and security.

c. → The boys or girls below 14 of their age are in fact children.

d. → The list of children rights are

- i) Proper care
  - ii) Love and affection
  - iii) Freedom and entertainment and
- the list of fundamental requirements are
- (i) housing
  - (ii) clothing
  - (iii) feeding and security

6.

a lot of :- Yesterday I had bring a lot of goods from the market.

put on :- Nam put on your clothes on the bank of river before you swim.

frost bite :- Generally this disease frost bite is cause on the himalayan region or cold region

balloonist :- Balloonist fly hot air balloon

7. => "If its too cold, we'll get frost bite," said another.

8. => My best teacher.

Those person who teach us, give knowledge and information and show way to success in life is known as teacher. My best teacher name is Manoj Regmi. He lives at mirgouliya. He is 30 years old. He is tall and handsome. He is very cooperative and helpful person. He love us. He teaches us from easy way to make understand. He always ~~and~~ regular come school. He is the principal of our school. He teaches as well all the students passes.

with good marks. I request all the teachers to like him. Then all the students get success.

4

9. A festival of Tihar I like most.

We know that our country is a religious country. Most of the people follow different religions like Hindu, Muslim, Buddhism and Christian etc. Most of the people of our country are Hindu. There are many festivals among them. Tihar is my favorite festival. This festival is celebrated by Hindu people. This festival is celebrated with great joy. People make their houses, shops clean. People put colourful electric lights all over houses. People decorate their houses using different colours and paintings. This festival is celebrated once a year. This festival takes place in the month of Kartik. Every year this festival will be celebrated.

On this festival people wear new clothes and eat delicious food and meat and chutney. The first day is dog festival, people worship dogs on this day. The food is provided to the dog on the fest plate. People put tika on the dog's head and garland. The second day is crow festival, people worship crows on this day. It is taken as our messenger. People provide food, grains to the crows. The third day is Laxmi puja. Laxmi is taken as the goddess of wealth. People worship

cow in the form of Laxmi. The fourth day is ox festival, people worshipped ox in this day. And the last or fifth day is Bhetika. On this day brother go to their sister home and sister worshipped their brother and sister put tilka and garlands on their forehead and put garland. They pray god for and brothers for their long life and healthy life. And brothers also offer to their sisters sweets, money, clothes etc. They provide blessing each other. On this day dandiya is play by the boy and vailo is play by the girl. On this day people play dandiya and vailo. People meet their relatives, friends and family members and exchange love and affection and create relationship. On this festival people celebrate alot.

7

10.

1<sup>st</sup> December, 2014

Dear Ram,

I am well here, I hope you will also in good health there. I am missing you alot. Yesterday I received your letter I became very happy. On your request I am going to write about my village.

My village name is Mirgauliya. It lies in the morning

district, kashizone. My village is not so big it is medium size. Many people live in our village they all have their job and occupation. Many houses are there in my village. Most of the houses are made up of wood and stone. In our village there ~~is~~ facilities of many things like transportation, drinking water, electricity communication facilities etc. In our village there are two schools and 1 health post and hospital. In our village most of the people are farmer. Most of the people utilize their facilities. They get service from village. Our village is full of ~~are~~ up green trees. In our village there is religious place and temple and big garden. The children of our village are studying in the schools. People of our village are cooperative, helpful and hard working. They help each other in the problems and ~~and~~ on the festivals and occasion we celebrate and enjoy together. I request do you come in this village and enjoy with us. @  
Now only this much many many love to you.

8

Your friend,  
Sabin Khawas.

Name: Sabin Khawas  
class ~~name~~ = Eight (8)  
Sub: English

50  
70

2

1. a) He is ~~a~~ university student.  
b) We always go to school by ~~foot~~. (4)  
c) This is the house where ~~I~~ was born.  
d) Sita, do it ~~yourself~~.

2.

- a) They never go to ~~temple~~. (1)  
b) My bag has been ~~stolen~~. (1)  
c) What did he ~~kill~~? (1)  
d) She told me ~~that~~ to open the door.  
e) Did he ~~done~~ all the work?  
f) People speak English all over the world. (1)

3.

- a) Ans => The L.W.F stands for ~~english in action competition~~  
Living World Film. (2)  
b) Ans => The height of Mt. Annapurna is 8075m. (2)  
c) Ans => Ellen was a blind girl. (1.5)  
e) Ans => Encyclopedia is a such type of book where

different kinds of information are found about the world.

1.5

B) Ans ⇒ They start flying by flapping their arms

4.

a) Ans ⇒ Feature films are made for entertainment and use actors and sets

2

b) Ans ⇒ Balloon over Nepal film has newly released in Kathmandu.

2

c) Ans ⇒ A film editor chooses short ~~short~~ pieces of film and joins them together to make the final film.

2

d) Ans ⇒ A sound editor adds sound and music to the film. This is called soundtrack.

2

5

A.

(i) To employ children in physical work is illegal.

(ii) The future of a nation depends upon the future of its children.

4

(ii) Proper care, love and affection, education etc are examples of child rights.

(iv) The boys and girls below fourteen are legally called the children.

B.

2. Ans ⇒ Yes, physical work of children is legal because in the context of Nepal to employ children in any sort of physical work is illegal.

b. Ans ⇒ The nation should ensure the child rights proper care, love and affection, education, freedom and entertainment together with fundamental requirements like housing, clothing, feeding and security.

c. ⇒ The boys or girls below 14 of their age are in fact children.

d. ⇒ The list of children right are

- i) Proper care
- ii) Love and affection
- iii) freedom and entertainment and the list of fundamental requirements are

- (i) housing
- (ii) clothing
- (iii) feeding and security.