

**EFFECTIVENESS OF TEST-RETEST METHOD TO MEASURE
THE RELIABILITY OF THE TEST ITEMS**

**A Thesis Submitted to the Department of English Language
Education, University Campus, Kirtipur in Partial Fulfillment for
Master's Degree in Education
(Specialization in English Language Education)**

**By
Sher Bahadur Adhikari**

**Faculty of Education
Tribhuvan University
Kirtipur, Kathmandu, Nepal
2007**

**EFFECTIVENESS OF TEST-RETEST METHOD TO MEASURE
THE RELIABILITY OF THE TEST ITEMS**

**A Thesis Submitted to the Department of English Language
Education, University Campus, Kirtipur in Partial Fulfillment for
Master's Degree in Education
(Specialization in English Language Education)**

By

Sher Bahadur Adhikari

M.Ed. (Second Year)

Faculty of Education

Tribhuvan University

Kirtipur, Kathmandu, Nepal

2007

T.U. Regd. No.: 258-81

Campus Roll No.:

Exam Roll No. (2nd year): 28281/059

Date of Approval of the Thesis:

Proposal: 063/12/26

Date of Submission: 2064-02-06

RECOMMENDATION FOR ACCEPTANCE

This is to certify that **Mr. Sher Bahadur Adhikari** has prepared this dissertation entitled "**Effectiveness of Test-Retest Method to Measure the Reliability of the Test Items**" under my guidance and supervision.

I recommend this dissertation for acceptance.

Date:

Dr. Tirth Raj Khaniya

(Guide)

Professor

Department of English Language Education

Faculty of Education

University Campus, T.U.

Kirtipur, Kathmandu, Nepal

RECOMMENDATION FOR EVALUATION

This dissertation has been recommended for evaluation by the following research guidance committee:

Dr. Chandreshwar Mishra

Reader and Head,
Department of English Language Education
T.U., Kirtipur, Kathmandu, Nepal

Chairperson

Dr. Tirth Raj Khaniya (Guide)

Professor
Department of English Language Education
T.U., Kirtipur, Kathmandu, Nepal

Member

Dr. Bal Mukunda Bhandari

Reader,
Department of English Language Education
T.U., Kirtipur, Kathmandu, Nepal

Member

Date:

EVALUATION AND APPROVAL

This dissertation has been evaluated and approved by the following
Thesis Evaluation Committee:

Dr. Chandreshwar Mishra

Reader and Head,
Department of English Language Education
Tribhuvan University
Kirtipur, Kathmandu, Nepal

Chairperson

Dr. Jai Raj Awasthi

Professor,
Department of English Language Education
Tribhuvan University
Kirtipur, Kathmandu, Nepal

Member

Dr. Tirth Raj Khaniya (Guide)

Professor,
Department of English Language Education
Tribhuvan University, Kirtipur
Kathmandu, Nepal

Member

Date:

I dedicate this dissertation to my parents and my wife who helped me a lot to introduce me as a person what I deserve these days.

ACKNOWLEDGEMENTS

I would like to extend my sincere gratitude to my research supervisor Prof. Dr. Tirth Raj Khaniya, Department of English Education, University Campus, T.U., Kirtipur for his invaluable help, encouragement, suggestions and right guidance. This research work would remain incomplete without his support and constant encouragement. In fact, I feel it an opportunity to work on my thesis under his guidance.

My equal gratitude falls upon Prof. Dr. Shishir Kumar Sthapit, Professor at the Department of English Education, Prof. Dr. Shanti Basnyat, Chairperson, English and other Foreign Language Education Subject Committee and Prof. Dr. Jai Raj Awasthi, Professor at the Department of English Education, for their kind cooperation and suggestions.

Likewise, I would like to extend my heartfelt gratitude to Reader, Dr. Chandreshwar Mishra, Reader, Head, Department of English Education, University Campus, T.U., Kirtipur for his creative suggestions and encouragement for carrying out this study.

Similarly, my thanks go to Dr. Vishnu Singh Rai, Prof. Dr. Yogendra Prasad Yadav, Prof. Dr. Govinda Raj Bhattarai, Dr. Bal Mukunda Bhandari, Mr. Ram Ekwel Singh, Dr. Anjana Bhattarai, Dr. Tapasi Bhattacharya, Mr. Padma Lal Bishwakarma and Mrs. Madavi Khanal, for their help.

I would like to extend my heartfelt thanks to all students of grade 10 of Shree Arunodaya H.S. School, Pharping, Taudaha Rastriya Secondary School, Kirtipur, Bageshwaree H.S. School, Malekhu, Nila

Kantha H.S. School, Dhading Bensi, Dhading who were directly involved in this research work.

I am indebted to the Head teachers and English teachers who cordially helped me to conduct this research work.

I would like to extend my thanks to my father Shanti Bahadur Adhikari and mother Aishwarya Adhikari who made me what I'm today. My thanks go to my elder brother Harka Bd. Adhikari who constantly encouraged me and supported me to earn my degree.

My warmth thank goes to my wife Mrs. Apsara Adhikari who helped, encouraged, supported and urged me to gain my higher study as well as to conduct this research work. Without her sparking support this research would have remained in complete.

At last, I would like to extend my thanks to the students of Bageshwori H.S. School. I express my thanks to my friend Mr. Rudra Prasad Adhikari for his cordial help. I would like to express my gratitude to Mr. Deepak and Narayan Basnet of Resunga Computer Service, Kirtipur for the computer work of this thesis.

Sher Bahadur Adhikari

ABSTRACT

The ultimate goal of this study is to determine the effectiveness of test retest method to measure the reliability of the test items. The analysis and findings of this study is based on the performance of the students of the grade ten marks secured by the students in the both first time and second time administered test and researcher himself on the basis of the various journals, articles, books and self preparing devices.

The major findings of the study are; subjective test items lacks the reliability but objective test items are much more reliable. The competence and performance level of the students varies from school to school. Due to the unclear instruction students could not do better in exam. Students are interested and easily solve the textual test items but they lack the consistency in non-textual test items.

It was found that the question paper is designed beyond the limited course so while preparing test items, the limitation should be with in the designed course. It can be concluded that piloting or sampling test is compulsory before administering the real test. So that all level, interest, competence and performance of the students can be addressed. To make the test items more reliable TM should be focused rather than NTM. In order to make the test items more reliable other methods should be applied like: alternative, split half and internal consistency. So that pure test items can be designed. While designing the test items purpose of the test should be clear so that a test can be measured what it is expected to test. It can be concluded that gender should be equally address so that all boys and girls can express their performance freely. It means test items should not be paradoxical and gender bias. Special trainings and

workshops are required for the teacher so that test items design can be uniformity.

The existing study consists of four chapters.

The first chapter deals with the introduction of the study. It includes the general background, ELT in Nepal, teaching and testing, teaching and testing English in Nepal, validity and reliability, test retest method review of related literature, objectives of the study and significance of the study.

The second chapter deals with methodology applied to carry out this research by the researcher.

The third chapter is an analysis and interpretation of the marks secured by the students of the grade ten to find out the effectiveness of the test retest method. Mostly this chapter is divided into three sections. The first section deals with the analysis of the student's performance and marks secured in the first and second time administered test. The second chapter deals with comparative analysis of the researcher. The end section deals with the researcher's own analysis.

The fourth chapter is related with the findings derived from the analysis and interpretation of the data. Mainly objective questions are more reliable than subjective questions.

Finally, some recommendations are made for further improving the test items. Some insights are also give to improve test items as well as to design the reliable test items.

TABLE OF CONTENTS

	<u>Page No.</u>
Recommendation for Acceptance	i
Recommendation for Evaluation	ii
Evaluation and approval	iii
Dedication	iv
Acknowledgements	v
Abstract	vii
Table of Contents	ix
List of Tables	xi
Abbreviations	xii

CHAPTER I

INTRODUCTION

1.1 General Background	15
1.1.1 ELT in Nepal	15
1.2 Importance of English	2
1.2 Teaching and Testing	18
1.3 Teaching and Testing English in Nepal	20
1.4 Validity and Reliability	21
1.5 Test-Retest Method	26
1.6 Review of Related Literature	27
1.7 Objectives of the Study	31
1.8 Significance of the Study	31

CHAPTER II

METHODOLOGY

2.1 Sources of Data	32
2.1.1 Primary Sources	32

2.1.2 Secondary Sources	32
2.2 Population of the Study	32
2.3 Sample Population	32
2.4 Tools for Data Collection	33
2.5 Process of Data Collection	33
2.6 Limitations of the Study	34

CHAPTER III

ANALYSIS AND INTERPRETATION

3.1 Introduction	35
3.2 Analysis of the First and Second Time Marks Secured by the Students of the Arunodaya Higher Secondary School Pharping, Kathmandu	35
3.3 Analysis of the First and Second Time Marks Secured by the Students of the Tauadaha Rastriya Secondary School, Kirtipur	37
3.4 Analysis of the First and Second Time Marks Secured y the Students of the Bageshwori Higher Secondary School, Richoktar, Malekhu, Dhading	38
3.5 Analysis of the First and Second Time Marks Secured by the Students of the Nila Kantha Higher Secondary School, Dhading Bensi, Dhading	39
3.6 Comparative Analysis of the Researcher	40
3.7 Researcher's Own Analysis	46

CHAPTER IV

FINDINGS AND RECOMMENDATIONS

4.1 Findings	49
4.2 Recommendations	37
References	40
Appendices	41

LIST OF TABLES

	Page No.
Table No. 1: Marks Secured by the Students of Arunodaya Higher Secondary School, Pharping	36
Table No. 2: Marks Secured by the Students of Tauadha Rastriya Secondary School, Kirtipur	37
Table No. 3: Marks Secured by the Students of the Bageshwori Higher Secondary School, Richoktar, Malekhu, Dhading	38
Table No. 4: Marks Secured by the Students of the Nila Kantha Higher Secondary School, Dhading Bensi, Dhading	39
Table No. 5: First Time Marks Secured by the Students of Arunodaya H.S. School, Pharping	41
Table No. 6: Second Time Marks Secured by the Students of Arunodaya H.S. School, Pharping	41
Table No. 7: First Time Marks Secured by the Students of the Taudaha Rastriya S. School, Kirtipur	42
Table No. 8: Second Time Marks Secured by the Students of the Taudaha Rastriya S. School, Kirtipur	43
Table No. 9: First Time Marks Secured by the Students of the Bageshwari H.S. School, Malekhu, Dhading	44
Table No. 10: Second Time Marks Secured by the Students of the Bageshwari H.S. School, Malekhu, Dhading	44
Table No. 11: First Time Marks Secured by the Students of the Nil Kantha H.S. School, Dhading Bensi, Dhading	45
Table No. 12: Second Time Marks Secured by the Students of the Nila Kantha H.S. School, Dhading Bensi, Dhading	46

ABBREVIATIONS

CDC	Curriculum Development Center
ELT	English Language Teaching
f	Frequency
FSM	First Secured Marks
M	Mean
N	Number of Items
NG	Nepal Government
No.	Number
NTM	Non-Textual Materials
SSM	Second Secured Marks
TM	Textual Materials
Σ	Summation

CHAPTER I

INTRODUCTION

1.1 General Background

1.1.1 ELT in Nepal

We know that English language is widely used language in the world. English language is the largest of the UNO. English is the language of world politics, science and technology, medicine, economics, marketing, trade and so no. The university education of the world is provided in English language. Most of the international books are written in English language. English language is taught and learned in Nepal as compulsory subject from class one to twelve. Since the foundation of the Darbar school in 1910B.S the How of English language is increasing in Nepal. Since then a lot of changes were made in the field of education, different commissions purposed many curriculum for educational changes.

The regular updating and improvement in education system in Nepal. English language curriculum has undergone through several ongoing processes of revision. So the curriculum of Nepal is changing form time to time. In the process of teaching learning process the curriculum plays significant role. The curriculum is the over all plan through which teaching learning activities are conducted in educational institutions, for the purpose of teaching and learning English language: a scientific English curriculum was developed in 2028 B.S. The secondary English curriculum has been changing from time to time. This existing new English curriculum in grade nine and ten was implemented from July 1999 (Shrawan 2056). This curriculum is based on communicative

approach. So it emphasizes on communicative skills, appropriateness fluency, accuracy and for developing the four language skills in the students.

Many subjects were taught and learned in those early days. Literature used to be the focus of study. Literary texts were perused, analyzed and critically studied. Mainly reading and writing were the language skills given priority to others. Of course, the study of the classics and their literary evaluation had less to foster the linguistic competence among the learners. English (Literature) was thought to be an end for its own sake rather than being a means of communication. Almost all the textbooks were in English and the medium of instruction of examination was also English.

1.2 Importance of English

There are many reasons, which have placed English in a high position. Western civilization has made a profound influence in the eastern countries. Especially, the reign of British in India has left the legacy of English in India and Nepal could not be an exception to this.

In the above line from the history of English in Nepal, we learned that for some decades Nepalese schools followed the English language aimlessly. During the same period the question of medium of instruction was raised in worldwide basis. UNESCO constituted a committee of experts in 1951 to consider the question of medium of instruction, on a world wide basis. The committee gave priority to the mother tongue, the language that children can effectively use, to be the medium of instruction at primary level and also recommended that the use of mother tongue be extended to as late stage in education as possible. "From all aspects, psychological, sociological and educational, the committee considered the mother tongue of children to be the best medium of

education (UNESCO 1953:11)". Finally, after lengthy arguments UNESCO suggested that a better way to introduce a second language might be to teach it first as a subject, using the mother tongue as a medium.

English was given the position of one of the UN languages in the secondary school curriculum. Theoretically, schools were allowed to choose from among the five UN languages (English, Chinese, French, Russian and German) but practically English was uniquely chosen for teaching. It was a compulsorily optional subject. Thus, 'it is one of the five official languages of the UN and as such plays a vital role in international communication' (Pathykrum, 2030:61).

English in the Nepalese context is regarded as a foreign language (out of 6 UN languages). The importance of teaching and learning of English can be viewed from several prospective. The need and demand of this language is increasing day by day. English is taught in Nepal as one of the ten modern languages, it can not be substituted for any other modern language. It is because English has become indispensable vehicle to the transmission of modern civilization into the nation. We can not neglect all this influence and force of English upon our national culture. We cordially accept it as a foreign language for unknown years to come because 'nothing in the foreseeable future is likely to change its present status as, the first foreign language' (Malla. 1977:23). So several reasons have placed English in a high position. Some of the reasons are discussed below:

Firstly, the practical reasons, i.e. the English language has an immense role in national and international communication. In the words of Kansakar (1979:15) the teaching of English in Nepal can broadly be divided for two different purposes. "It can be taught for international

purposes, that is, for carrying as international relations on a world scale and also for national purpose as for example, for using it as a lingua franca. In Nepal English is taught especially for the first purpose. It is through the English language that we seek to communicate with the outside world". Thus, English has become an important language. To some countries, it has served in national communication and to others it has been serving in the international communication in the world.

Secondly, English is important for academic reasons. It is widely used in academic places for teaching and learning or studying and training to approach the world body of knowledge. In our context, an SLC student rarely meets opportunities for communicating in English. But he may feel frequent urge to do so as he proceeds to university education. Here, textbooks and reference materials he is supposed to consult are in English.

Fiscal reason is the third reason. It is clear that English opens the doors of better employment. If someone intends to choose a particular job, most probably requires in him the functional knowledge of English.

Fourthly, English is widely used in the field of Tourism. Since English is regarded as the world's lingua-franca, people from different parts of the world can visit and share their ideas by means of English.

1.3 Teaching and Testing

Teaching and testing are interrelated each other. Teaching/learning and testing activities are linked to a common goal, that is. the objectives of teaching the learner. Any teaching or learning, which is not directed to the attainment of the set objectives, therefore, regarded as fruitless.

There are a lot of terms in the literature such as testing, assessment, evaluation, examination, measurement etc. Conventionally, it is believed that the above said terms are administered to expose the weaknesses of the learner or not only the means of students fault finding device. They are no longer taken as an end in themselves rather they are taken as a means to an end, language teaching.

Testing as the use of test or the study of theory and practice, their use development and evaluation. Test is any procedure for measuring ability. knowledge or performance" (Richard et al., 1985:291). A test is thus regarded as an attempt to see whether the things taught have been learnt or not.

Ingram (1974:33) states that, "a test is measuring device which will use when we want to compare an individual with other individuals who belong to the same group. If we want to compare for height. We use a yard stick, if we want to compare them in terms of their command of a foreign language we use a language test." Thus, the meaning of the 'tests' has covered a broad area. The tests use for language ability varies with the tests of physical ability and so on. Many attempts have been made over language proficiency test. "For them language test is a device that tries to assess how much has been learned in a foreign language course or some parts of course. For any student whose native language or traditionally thought of as language tests may be primarily tests a language ability" (Oller, 1989:276).

Thus, many attempts have been made for the improvement of tests for the English language of our curriculum. The testing system in all school levels is but in campus level it has its own system. Half-yearly and yearly examinations are administered in school level but in campus level only yearly examination is given. It could be the common comment that

only these sorts of examinations are not enough to assess the learners' further progress in language. On the other hand, it has been found by the nature of tests that some of the language skills has been left or untested in the prevalent examination system in Nepal.

1.4 Teaching and Testing English in Nepal

Education in its modern sense started very late in Nepal. 'The dawn of democracy in 1951. the Nepalese government did feel the necessity of giving education to its people. Till then there were limited schools and colleges and only few upper class people could avail of this opportunity. The masses were not aware of the importance and necessity of education.

Long ago before democracy, education was centered in Gurus' Aashram. People use to send their sons to Gurus for education. The 'Gurus' who lived in a jungle or far from home took full responsibility of the students. The students had to live with their 'Guru' and obey him without question. In Hinduism and Buddhism students had to live in temples and Gumbas for achieving knowledge. It was a sort of residential schooling. The teachers of that time were very authoritative and students were very submissive, teachers should have play active role and students passive. They taught whatever they thought fit for their students.

However since 1951, there has been a rapid growth in the number of schools, colleges, technical, institutes and number of students has also increased tremendously. Nepali and English are taught as compulsory subjects but Nepali as a medium of instruction. All other subjects including English (through translation) are taught in Nepali at school level. Thus the concept of mother tongue education as envisaged by the UNESCO in 1951 nourished in Nepal. However, the implementation of UNESCO ideals in Nepal can still be questioned.

Before 1971, English was taught at all levels of education: primary, lower secondary, secondary, and campus. The implementation of the National Education System Plan (NESP), English language has been introduced from class IV. Thus, English in Nepal is introduced from class IV at primary level with the 9-10 years age group learners. The general objectives of teaching English from primary to secondary levels are same. The behavioral objectives of each grade differ from that of the other. There is a 100 full marks English course in all levels in school. Everything along with situation is changed. The textbooks and syllabuses or courses are innovated taking into account the need of the students, society and nation. Not only this but with the changing time, the old grammar translation method of teaching is changed to direct method and audio-lingual methods to achieve communicative approach. The testing system was based on written form. Higher priority was paid to rote learning and memorization in the past. Traditionally, English was treated as a vehicle for transmitting culture. After many years, literature was replaced by language orientated course.

1.5 Validity and Reliability

i. Validity:

A test is said to be valid if it measures accurately what it is intended to measure. Validity is not the self-correlation of the test, rather it is the correlation of the test with some outside independent criteria.

When a test measures that which it is supposed to measure, and nothing else, it is valid. To the extent that the readings on the measuring device reflect other characteristics—for instance familiarity with the topic when we are looking for general comprehension—or are subject to the kinds of error discussed above under reliability, the validity will be reduced.

The difficulty is that the characteristic or attribute which we wish to measure in education is usually something which is not very obvious, or easy to define. In second language learning, the attribute labelled 'fluency' is conceivably the most elusive of the lot, but 'command of the basic grammatical structures' or 'ability to function at the tourist level', while sounding more concrete, are still in need of a series of operational definitions.

We provoke certain kinds of behaviour in the candidate, in the hope and belief that we can generalize from the way that he behaves on a given occasion on a limited number of tasks, to the way that he will perform on all other relevant occasions. But obviously hoping is not enough. We must have some expression of validity, some way of estimating even approximately that the test is doing the job it is supposed to do. There are three commonly recognized kinds of validity. These are pragmatic validity, content validity, and construct validity. Some writers divide pragmatic validity into two parts, predictive and concurrent (see chapter 3). There is a fourth kind known as face validity. This has to do with the surface credibility or public acceptability of a test, and while it is sometimes important, it can be regarded as a public relations problem rather than a technical one. In this section I shall deal mainly with pragmatic validity.

Estimates of pragmatic validity are obtained by correlating test scores with the scores or ratings obtained from a criterion measure. Criterion measures can be standard examinations, established tests, the ratings obtained from experienced judges on observational samples, By definition, anything which serves as a criterion is taken to possess validity. There are obvious problems here, and it is usual to distinguish between immediate and ultimate validity. If a test correlates highly with the chosen criterion, then the test has immediate or first order validity.

For instance if a test accurately predicts success and failure in a set of examinations, then the test possesses immediate validity. But the criterion here, the examinations, is supposed to be validly related to some kind of scholastic or professional ability. If the examinations are indeed valid predictors of vocational success, and if that is what we wish to predict, then the examinations, and hence the test, possess ultimate validity. In the field of language teaching however, teachers are usually only too aware that the ultimate criterion—ability to use the language—is what matters. The problem is how to arrange the instruction so that the aim is achieved, and how to make sure that the examinations and tests which the learner has to take are good indications of ability to use the target language in real situations.

Not all situations require ultimate validation. One might want to compare the efficiency of two methods of teaching specific grammar points, for instance the use of some and any. The validation for a set of items testing this could be an ongoing count of the number of errors the learners make in their classroom work for a specific period of time, i.e. in this way we would be establishing the criterion by observational samples. Obviously this is not the sort of activity a teacher can routinely engage in, but the point is that for this purpose immediate validity is what is required. If the purpose was to establish how much the ability to distribute some and any contributes to skill in performing in a real communication situation, then of course the ultimate criterion would have to be brought back in.

When a test is essentially a job sample, a particular slice of the kind of activity we wish to assess, we are obviously dealing with a special kind of immediate validity—sometimes referred to as inherent validity. Essay writing is often given as an example: if you want to assess how good a person is at writing essays, you get him to write an essay or

two. Similarly, if you want to know how good a person is at talking in a foreign language you set up a situation in which he does just that—you give him an oral. Further, if essay writing really is what the teaching process is designed to produce, then essay writing is obviously its own criterion. But it could be that essay writing is simply a convenient subskill to test, out of the whole range of skills which goes into writing. In that case the validity of essay writing should in turn be established against an ultimate and composite criterion consisting of writing in all registers judged to be relevant.

However, let us take the situation where the tests really do constitute a job sample, it is usually true that the activity called for is rather complex and therefore the task of judging it is also complex. As has often been pointed out, the judgements of different examiners of the quality of the same essay can vary considerably. In other words the reliability of the assessment of complex, realistic job samples tends to be somewhat low, and this lowers the effective validity. Reliability of essay marking can be improved by using several judges, each assessing independently. And validity can be improved by requiring the student to write two or three shorter essays, rather than one single long one. Job sample reliability and validity can be improved in the same way.

For determining test "Validity is the degree to which a test measures what it is suppose to measure or can be used successfully for the purposes for which it is intend." Heaton.

The validity of at test is the extent to which the test measures what it intended to measure. Aim to test writing skills e.g. question item:

Is photography an art or science?

ii. Reliability:

Reliability is also a good quality of a test. So to become a good test there should be reliability.

The reliability of a measuring device is high when any variation in the readings taken represent true differences between the individuals who are being tested. Any other variation represents error.

A great deal of the effort which goes into test construction is devoted to reducing error. What might be called extrinsic sources of error are examiner variability and variability of testing conditions. These can be greatly reduced by common sense precautions. Examiner variability is virtually eliminated by objective formats, used properly, and variability of testing conditions are reduced by meticulous care in providing instructions to the test administrator and in formulating the explanations to the candidates, and if necessary by giving some preliminary practice, so that people who have not taken an objective test before are not unduly handicapped.

Reliability simply refers to the consistency of scores on measurement. That is to say a test set to be reliable if the scores remain the same or nearly the same on marking the test on different occasions by the same marker or on the same occasion by different markers. There should consistency of scores.

Reliability is a measure of the degree to which a test gives consistent results. A test is said to be reliable if it gives the same results when it is given on different occasions or when it is used by different people.

Reliability is a necessary characteristic of any good test for it to be valid at all a test must first be reliable as a measuring instrument . If the

lest is administered to the same candidates on different occasions (with no language practice work taking place between these occasions) then to the extent that it produces differing results, it isn't reliable. Reliability measured in this way is commonly referred to as test re-test reliability to distinguish it from mark re-mark reliability to and the other kinds of reliability denotes the extent to which the same marks or grades are awarded if the same test papers are marked by :

- i. Two or more different examiners
- ii. The same examiner on different occasions.

In short, in order to be reliable a test must be consistent in its measurements.

Reliability is of primary importance in the use of both public achievement and proficiency tests and class room tests. Methods of estimating the reliability of individual items in a test will be indicated in the next chapter. However an appreciation of the various factors affecting reliability is important for the teacher at the very outset, since many teachers tend to regard tests as infallible measuring instruments and fail to realize that even the best test is indeed a somewhat imprecise instrument with which to measure language skills.

1.6 Test-Retest Method

This is the most obvious way of obtaining reliability. It is the single form of test is administered twice on the same sample (s) with a reasonable time gap. In this way two administrations of the same test yield the two independent sets of scores. The two sets when correlated, give the value of reliability . This indicates to what extent the examinees retain their relative position as measured of time. A high test score over a given period indicates that the examinee who obtains a low score on the

first administration tends to score low on the second administration and who obtains a high score tends to score high on the second.

This method contains several drawbacks. If the second administration of the test is too soon after the first, then subject one likely to recall items and their responses to them, making the same responses more likely, this making correlation co-efficient to be very high. Similarly, if there is too long gap. the subjects are likely to be motivated more and learn more on the matter, or forgetting may take place in such a case. Thus, it may lead to the possibility of decreasing reliability co-efficient.

1.7 Review of Related Literature

English is taught at all levels of education in Nepal. It is estimated that the teaching of English at the secondary level (grade 9 and 10) accounts for 11.11% of the total academic programmed and the Higher Secondary level (along with existing proficiency certificate level) occupies about 16.68% if teaching and learning time. It is also estimated that some 30% of the education budget at these levels may well go to English' (Kansakar, 1998:71). Further in his article, "in recent years there have major changes in structure and .content of curricula and textbooks at secondary (grades 9 and 10) and higher secondary (Classes XI and XII) level and re-orientation of the Bachelor programme to include an extra year of teaching. Each of these levels, education was envisaged as self terminating with proportionately decreasing enrollment targets. Prior to the present reforms and innovations, it was estimated that 40% of lower secondary students took up secondary studies and only 19% of text books, manageable class size and effective evaluation techniques. Students continue to learn very little English in classes and hardly any English outside the English classroom. Besides, hand, we are still

suffering from the lack of good co-ordination between teaching and testing.

Teaching and testing are so closely interrelated that it is virtually impossible to work in either field without being constantly concerned with the other. Tests may be constructed primarily as devices to reinforce learning and to motivate the students, or primarily as a means of assessing the students' performance in the language. We find many differences in a traditional way of testing the same aspect of language when it is compared with the new methodology. As quoted by Heaton (1977:4) says 'the traditional examination may tell us relatively little about the students general fluency and ability to handle the target language, although it may give some indication of the student's ability in some of the skills he need as a students'. So most people probably have visions of students in a foreign language classroom poring over a written examination. Only few of them may know what exactly a language test is a device that tries to assess how much has been learnt in a foreign language course or some part of the course."

Thus testing is equally important as teaching is. Test can be subdivided into three major types:

- i. Goal based test.
- ii. Mode based test.
- iii. Approach based test.

Approach based test is the main concern of the researcher. Approach based tests are fixed with same approach. Such tests are divided into two categories, discrete point test and integrative test.

We find some of the reliable and suitable points to our context in Rivers opinion. Rivers (1968: 213) says 'languages have been taught to

students down the centuries by a variety of methods, and with a competent teachers, the student learns what the teacher feels it is important for him to learn. No matter what-method is in vogue or is officially advocated, the individual teacher who is professionally, and to what he feels to be appropriate for the particular class he is teaching, having regard to their age level, their situation and their educational needs. Any method ceases to be efficient when it is applied inflexibility, according to set procedures, in every situation.'

Thus, to raise the standard of English in Nepal; students level of knowledge, their age, situation and educational needs are the focal points to be considered while designing the course, teaching methodology and testing devices.

Sthapit (1984) says 'in Nepalese context, for the development of general proficiency in the English language, reading and writing skills should be given especial emphasis " In response to this statement the new curriculum has been developed which highly emphasized on communicative approach of teaching and learning the English and speaking skills. It is hoped the new concept of teaching and learning English may bring a positive change in Nepal.

Kimberley at UCLA (1972) suggests that 'Cloze Tests are not generally sensitive to any practice effect and therefore they should be quite applicable to routine classroom studies with repeated testing. As Kimberley suggests Cloze Test Frequent practice, if this test is new to our students the performance can be terribly low in the beginning.

Shiwakoti (1996) in his M.Ed, thesis entitled An analysis of the reading proficiency of the secondary school students of Jhapa District focused on cloze test, which was not used as a means of test in NG aided schools and private schools. His study found that private school students

performed better than HMG aided schools. The reason he pointed out was that the private school students were better facilitated by the extra supplementary materials. It was also justified by his study that the urban school students performed better than the rural schools students. His comparative study on textbook material and non textbook material showed the better performance of both types of schools on non textbook than in textbook materials. The reason he added was though the textbook material was taken from grade ten, it was also not taught to the students and it was new materials for them.

Khanal; (1997) in his M.Ed, thesis entitled "A study on the effectiveness of the Cloze Test over conventional objective test in testing reading comprehension in English concludes that both private and NG aided school students showed better performance in objective test than in close test on both textbook and non-textbook materials. The reason behind their low performance in Cloze test was the novelty of the test. Neither they were involved in practice in such tests nor such tests were administered in exam ever before. In respect of the overall performance on the close test in combined. (TM and NTM) text private schools were ahead of the public schools by 38.03% average mark. The comparative analysis concluded that private schools showed better performance on objective tests and close test than the public schools. However, both types of schools were found to have secured relatively higher and lower marks in objective tests and the Close test respectively. His calculation of coefficient of correlation between the objectives test and Cloze tests showed very positive and high. The study also found that if the objectives tests reflected a good result presumably the Close test would also do the same and vice-versa.

The above literatures provide us the knowledge of the Test conducted by many researches. In comparison, discrete type of test has been carried out lower in number by the researchers. It is the matter of curiosity to a Nepali researcher to watch the usefulness, implications of discrete point test since this type of test has not been carried out ever before. But some of the foreign researchers as Harris (1969). Clark (1972), Value (1967: 77) Heaton (1977) have done many tests in discrete point test. Similarly. Davis (1977) and Lado (1961) suggest that vocabulary items, grammar items and so on. the elements of language can be tested separately from discrete point test."

1.8 Objectives of the Study

The main aim of the research is to study of the effectiveness of the test retest method as to measure the reliability in test design. For this following objectives are proposed:

- a. To find out the effectiveness of test- retest method to measure the reliability of the test.
- b. To provide some pedagogical implications and insights for designing the test items

1.9 Significance of the Study

The findings of this research will be beneficial to who are concerned and involved in designing the test at different levels. It will be helpful to improve the test items. The findings will be beneficial to the textbook writer, curriculum designers, subject experts, examiners, classroom teacher and language learners too. This research will provide there insights the reliability on designing the test items.

CHAPTER II

METHODOLOGY

This chapter describes the design of the plan and procedures which were used to carry out the research by the researcher to achieve the mentioned objectives of the study. In this study, the researcher attempted to find out the effectiveness of test-retest method to find the reliability of the test items. The analysis was done on the based on the scored marks by the students of the grade ten of Kathmandu and Dhading district.

2.1 Sources of Data

The researcher used the both primary and secondary sources of data.

2.1.1 Primary Sources

The primary sources were the students of grade ten of Dhading and Kathmandu districts.

2.1.2 Secondary Sources

A part form primary sources of the data the researcher studied different books, thesis, articles, journals etc. which are related to the research work and the documents which facilitate the study.

2.2 Population of the Study

The population was consisted the students of the grade ten of Dhading and Kathmandu district.

2.3 Sample Population

It is obvious that sampling is one of the most important aspect of the any research study. Sample population is a smaller representation of a large whole. For this study the researcher randomly selected the grade ten

students from Kathmandu and Dhading districts. The students were selected from Arunodaya Higher Secondary School, Pharping, Kathmandu, Taudaha Rastriya Higher Secondary School, Kirtipur, Kathmandu. The ten students were selected from Bageshwori Higher Secondary School, Malekhu and Nil Kantha Higher Secondary School, Dhading Bensi of Dhading district.

2.4 Tools for Data Collection

The researcher prepared the test items based on class ten curriculum. The researcher prepared 80 marks test items. Out of the 80 marks the test items, ten marks from unseen passage, 15 marks from vacancy announcement, 12 marks from application writing or essay writing, 6 marks from paragraph writing, 6 marks from story writing and 11 marks from grammar. For grammar, 6 questions from tense, 6 questions from reported speech, 4 questions from voice, 4 questions from sentence and 2 questions from causative verbs. Each grammar question carries 0.5 marks with objective questions.

2.5 Process of Data Collection

Data collection was accomplished by visiting the designated school and was administered the tests with the help of the subject teacher of the concerned school by the researcher himself. Before administering the test items necessary instruction was given to the students. Both objective and subjective questions were prepared from the textbook of grade ten. Test was administered to the same group of students right after the 15 days of the first administered time. The obtained marks were calculated, compared and tabulated. Finally, comparison was made between first scoring and second scoring.

2.6 Limitations of the Study

The study was limited to the following:

1. This research was limited to the students of grade ten of Kathmandu and Dhading district.
2. This study was limited to effectiveness of the test-retest method to measure the reliability of test items.
3. The analysis of the secured marks was limited to the students of Arunodaya Higher Secondary School, Pharping, Taudaha Rastirya Ma.Vi., Kirtipur of Kathmandu district and Bageshwori Higher Secondary School, Malekhu and Nil Kantha Higher Secondary School, Dhading of Dhading district.

CHAPTER III

ANALYSIS AND INTERPRETATION

3.1 Introduction

This chapter deals with the interpretation and analysis of the data that were collected from primary sources. The researcher collected data by preparing questions based on the textbook of grade ten. Data were collected by the test taken to the students of grade ten of Kathmandu and Dhading districts. The collected and information were analyzed under four different headings.

- Analysis of the first and second time marks secured by the students of Arunodaya Higher Secondary School, Pharping, Kathmandu.
- Analysis of the first and second time marks secured by the students of Tauadaha Rastirya Secondary School, Kirtipur.
- Analysis of the first and second time marks secured by the students of Bageshwori Higher Secondary School, Malekhu, Dhading.
- Analysis of the first and second time marks secured by the students of the Nil Kanatha Higher Secondary School, Dhading Bensi, Dhading.
- Comparative analysis of the researcher.
- Researcher's own analysis.

3.2 Analysis of the First and Second Time Marks Secured by the Students of the Arunodaya Higher Secondary School Pharping, Kathmandu

Students are the target group of whole educational activities. The entire educational activities are organized to address the students. Curriculum is designed, textbooks are written, teaching materials are

selected, teachers are selected and whatever are done, should be according students level, interest, experiences, priori knowledge and their competence level. So their performance should be measured with reliable test items.

The researcher prepared 80 marks written text items including subjective and objectives. The test items were administered to the students of the grade ten. The same questions paper was administered to the same group of the students right after 15 days.

The following table shows the marks variation between two times administered test items of the students of the Arunodaya Higher Secondary School, Pharping.

Table No. 1
Marks Secured by the Students of Arunodaya Higher Secondary School, Pharping

S.N.	Name	F.S.M	S.S.M.	Marks variation
1	Sabitri Maharjan	68	70	2
2	Chirag Tamang	66	68	2
3	Mamata K.C.	67	69	1
4	Dev Raj Gasmev	58	60	2
5	Kiran Balami	48	52	4
6	Shanti Tamang	52	52	0
7	Rojina Shrestha	56	57	1
8	Nanita Khatri	52	51	-1
9	Subash Bhurtel	41	38	-3
10	Shreejana Shrestha	42	44	-2

F.S.M= First secured marks, S.S.M. = Second secured marks

The table given above shows that the competence and performance level of the students is variation. The students have secured 41 to 68

marks in the first administered test. The same question was administered right after 15 days. In the second time administered test students have secured 38 to 70 marks. Among the students three students got two marks more than earlier. Two students got 1 mark more, one student got 4 marks more and only one student got equal 52 marks both time. One student got one mark less and one student got three marks less than earlier.

The secured marks does not support the reliability aspects of the test items. Although 10 percent students got equal marks whereas 40 percent students got reliable marks. But 10 percent students got relatively high marks whereas 20 percent students got relatively low marks. So table No. 1 does not support the reliability aspect of the test items.

3.3 Analysis of the First and Second Time Marks Secured by the Students of the Tauadaha Rastriya Secondary School, Kirtipur

Table No. 2

Marks Secured by the Students of Tauadha Rastriya Secondary School, Kirtipur

S.N.	Name	F.S.M	S.S.M.	Marks variation
1	Sujita Khadka	62	64	2
2	Mandira Shrestha	64	65	1
3	Rojina Adhikari	58	58	0
4	Buddhina Shrestha	60	61	1
5	Raju Thapa	40	38	-2
6	Raj Kapur Maharjan	26	28	2
7	Shanti Koirala	44	46	2
8	Rojina K.C.	54	52	-2
9	Sunita Kunwar	36	36	0
10	Hira Lama	28	32	4

F.S.M= First secured marks, S.S.M. = Second secured marks

The table given above shows that the knowledge level of the students is varied from one to another. The students have secured 26 to 64 in the first administered test and 28 to 65 in the second administered test. The secured marks seemed varied right after 15 days. Although 20 percent students got same marks whereas 30 percent students get two marks more than first administered test. 20 percent students got only one marks more than first administered test. 10 percent students secured 40 marks more than first administered test. But 20 percent students secured two marks less than first administered test.

The secured marks in the table No. 2 does not support the reliability aspects of the test items. 20 percent of students got equal marks but 10 percent students got four marks more where as 20 percent students got less marks. So table No.2 or item 3.3. does not support the reliability aspect of test items.

3.4 Analysis of the First and Second Time Marks Secured y the Students of the Bageshwori Higher Secondary School, Richoktar, Malekhu, Dhading

Table No. 3

Marks Secured by the Students of the Bageshwori Higher Secondary School, Richoktar, Malekhu, Dhading

S.N.	Name	F.S.M	S.S.M.	Marks variation
1	Shiva Prasad Tripathi	56	58	2
2	Prabin Pandey	51	53	2
3	Ramesh Silwal	52	55	3
4	Ishma Kadel	49	49	0
5	Madhav Khatiwada	41	43	2
6	Ramesh Regmi	38	42	4
7	Nisha Rawal	43	43	0
8	Ramesh Prasad Silwal	39	39	2
9	Goma Damai	35	38	3
10	Ranjiv Oli	36	39	3

F.S.M= First secured marks, S.S.M. = Second secured marks

The table given above shows that the knowledge level of the students is differ from one to another. The students have secured 35 to 56 marks in the first administered test whereas 38 to 58 in the second administered test. The same test paper was administered right after 15 days, 20 percent students got equal marks, 40 percent students got 2 marks more, 30 percent students got 3 marks more and 10 percent students got four marks more.

The secured marks in the table No. 3. does not support the reliability aspect of the test items.

3.5 Analysis of the First and Second Time Marks Secured by the Students of the Nila Kantha Higher Secondary School, Dhading Bensi, Dhading

Table No. 4

Marks Secured by the Students of the Nila Kantha Higher Secondary School, Dhading Bensi, Dhading

S.N.	Name	F.S.M	S.S.M.	Marks variation
1	Ashok Shrestha	56	57	1
2	Elisha Adhikari	54	52	-2
3	Rajan Burlakoti	51	49	-2
4	Hari Burlakoti	48	52	4
5	Rita Maharjan	26	28	2
6	Ritu Rizal	36	36	0
7	Dipesh Pandey	44	45	1
8	Ganga Ghimire	28	28	0
9	Anjana Subedi	38	40	2
10	Indu K.C.	52	55	3

F.S.M= First secured marks, S.S.M. = Second secured marks

The table shows that the knowledge level of the students is varied from one to another. The students have secured 26 to 56 marks in the first administered test whereas students have secured 28 to 57 marks. The secured marks seemed varied right after day. According to the table 20 percent students got the equal marks 20 percent students got two marks higher than first administered test, 20 percent students have got 1 marks more than first administered test, 10 percent students have got 3 marks more, 10 percent students have got 4 marks more than first administered test whereas 20 percent students have got two marks less than first administered test.

The secured marks in the table no.4 does not support the reliability aspect of test items lacks the reliability. It is concluded that item No. 4.5 does not support the reliability aspect.

3.6 Comparative Analysis of the Researcher

To find out the effectiveness of the test-retest method the researcher has conducted alternative method also. Test-retest method is an estimate of the reliability of a test administered by the extent to which a test gives the same result if it is administered at the different times. Besides that researcher has used alternative method to find out the reliability of the test result. For the alternative method researcher had prepared two sets of parallel questions and administered the students of Kathmandu and Dhading districts.

Table No. 5
First Time Marks Secured by the Students of Arunodaya H.S.
School, Pharping

S.N.	x	f	fx
1	68	1	68
2	66	1	66
3	67	1	67
4	58	1	58
5	56	1	56
6	52	2	104
7	48	1	48
8	42	1	42
9	41	1	41
		N = 10	550

$$M = \frac{\sum fx}{N} = \frac{550}{10} = 55$$

In the first administered test the mean score of the test is 55.

Table No. 6
Second Time Marks Secured by the Students of Arunodaya H.S.
School, Pharping

S.N.	x	f	fx
1	70	1	70
2	69	1	69
3	68	1	68
4	60	1	60
5	57	1	57
6	52	2	104
7	51	1	51
8	44	1	44
9	38	1	38
		N = 10	561

$$M = \frac{\sum fx}{N} = \frac{561}{10} = 56.1$$

In the second administered test the mean score of the test is 56.1.

Due to the mean variation the test result does not fully support the reliability aspect.

Table No. 7

**First Time Marks Secured by the Students of the Taudaha Rastriya
S. School, Kirtipur**

S.N.	x	f	fx
1	64	1	64
2	62	1	62
3	60	1	60
4	58	1	58
5	54	1	54
6	44	1	44
7	40	1	40
8	36	1	36
9	28	1	28
	26	1	26
		N = 10	432

$$M = \frac{\sum fx}{N} = \frac{432}{10} = 43.2$$

In the first administered test mean score of the test is 43.2.

Table No. 8
Second Time Marks Secured by the Students of the Taudaha
Rastriya S. School, Kirtipur

S.N.	x	f	fx
1	65	1	65
2	64	1	64
3	61	1	61
4	58	1	58
5	52	1	52
6	46	1	46
7	38	1	38
8	36	1	36
9	32	1	32
10	28	1	28
		N = 10	450

$$M = \frac{\sum fx}{N} = \frac{450}{10} = 45$$

In the second administered test the mean score of the test is 45.

Due to the mean or average marks variation the test items lacks the reliability.

Table No. 9

**First Time Marks Secured by the Students of the Bageshwari H.S.
School, Malekhu, Dhading**

S.N.	x	f	fx
1	56	1	56
2	52	1	52
3	51	1	51
4	49	1	49
5	43	1	43
6	41	1	41
7	39	1	39
8	38	1	38
9	36	1	36
10	35	1	35
		N = 10	440

$$M = \frac{\sum fx}{N} = \frac{440}{10} = 44$$

In the first administered test, mean score of the test is 44.

Table No. 10

**Second Time Marks Secured by the Students of the Bageshwari H.S.
School, Malekhu, Dhading**

S.N.	x	f	fx
1	58	1	58
2	55	1	55
3	53	1	53
4	49	1	49
5	43	2	86
6	42	1	42
7	39	2	78
8	38	1	38
		N = 10	459

$$M = \frac{\sum fx}{N} = \frac{459}{10} = 45.9$$

In the second administered test, mean score of the test is 45.9.

It is found that the scored marks are varied between first and second administered test. The mean or average marks is also varied so table No. 10 lacks the reliability of test items.

Table No. 11

**First Time Marks Secured by the Students of the Nil Kantha H.S.
School, Dhading Bensi, Dhading**

S.N.	x	f	fx
1	56	1	56
2	54	1	54
3	52	1	52
4	51	1	51
5	48	1	48
6	44	1	44
7	38	1	38
8	36	1	36
9	28	1	28
10	26	1	26
		N = 10	433

$$M = \frac{\sum fx}{N} = \frac{433}{10} = 43.3$$

In the first administered test, the students of Nil Kantha H.S. secured the 43.3 marks as the mean marks.

Table No. 12
Second Time Marks Secured by the Students of the Nila Kantha H.S.
School, Dhading Bensi, Dhading

S.N.	x	f	fx
1	57	1	57
2	55	1	55
3	52	2	104
4	49	1	49
5	45	1	45
6	40	1	40
7	36	1	36
8	28	2	56
		N = 10	452

$$M = \frac{\sum fx}{N} = \frac{452}{10} = 45.2$$

In the second administered test mean score of the test result is 45.2.

It is concluded that the secured marks is varied from first administered to second administered test. The mean or average marks is also varied so table No. 11 and 12 do not support the reliability of the test item.

3.7 Researcher's Own Analysis

Language teaching and testing are the two sides of a coin. So testing is the evaluation of the over all teaching and students competence and performance. Testing is the complex phenomenon. It is not easy to construct the test items to represent the all kinds of students. Before administering the real test teacher should prepare representative test

paper. So that all geographical, social and mental level easily take part in such kinds of test.

There are different ways to make the test reliable. The researcher has considered the test, retest method to find the effectiveness of test items. Besides the test-retest method, alternative, split half internal consistency and mark-remark are also popular to find reliability of the test.

To find out the reliability of the test items through test-retest method the researcher has prepared 80 marks objective and subjective test items. Out of the 80 marks 10 marks was from seen passage. By over all examination the students got elastic marks. 5 marks question was asked form poem and the marks which secured by the students in poem was quick constant. 15 marks question was asked form unseen passage in this question. Students have scored constant marks. 15 marks was asked from vacancy announcement in this question students have scored constant marks. 12 marks essay writing was asked but the marks of the varied and not seen consistency. So essay writing lacks the reliability of the test items.

Likewise 6 marks paragraph writing was asked but in this question also students answer seem variation and elasticity so the marks also seem variation. So paragraph writing also lacks the reliability of the test items. Six mark close test was also asked, the students answer also seem variation there too. But in comparison with essay writing and paragraph writing story writing has got a bit more reliability.

The textbook for grade ten has been designed according to communicative approach. So only 11 marks grammar questions were asked dealing with tense, reported speech, voice, causative verb and

conditionals. Twenty two objective questions were asked dealing with 5 marks for each questions. Out of twenty two questions reported speech and voice questions were quite difficult and student choose the guess answer only so it lacks the consistency.

But theoretical and subjective questions lacks the reliability than objective questions.

CHAPTER IV

FINDINGS AND RECOMMENDATIONS

The main expectation of this study was to determine the effectiveness of test-retest method to measure the reliability of the test items. Regarding to the study the researcher himself observed and read the class ten textbook for preparing question paper. The researcher prepared 80 marks test items for the students of grade 10. The researcher administered the test item to different schools of Kathmandu and Dhading district for the students of the grade ten. 80 marks questions were based on textbook for grade 10. Both subjective and objective questions were prepared. Equal marks parallel questions was also prepared to measure the reliability of the test items. After analysis and interpretation of the data obtained from the students of the grade ten by both test-retest and alternative method the following findings were drawn by the researcher.

4.1 Findings

The findings of the research work were mentioned below:

1. It is found that textual questions remain more reliable than non-textual questions. In class ten more than 50 percent questions are asked non-textually so student's answer remain bias as well as lacks the consistency. In comparison with subjective question objective questions remain more reliable. Due to the marks variation between first and second time administered test, the test items lacks the reliability but alternative method remain quite reliable.
2. Due to the marks variation of the students the test items can't address the level of the students. Here also objective questions were more reliable than subjective questions. Again non-textual

questions lacks the reliability, on the top of that essay writing, paragraph writing and story writing lacks the reliability. But alternative method seemed more reliable than the test re-test method.

3. The students performance and competence level seemed too much variation. Some students got 2 to 4 marks more than the first administered test whereas some students got equal marks. Due to the marks variation it can be concluded that the test items lacks the reliability. As like other, the subjective test items lacks the reliability rather than objective test items. Non-textual questions like 'unseen passage', 'vacancy announcement' lacks the reliability. Likewise, essay writing, paragraph writing and story writing seemed more unreliable. But comparatively alternative test items seemed more reliable than test re-test method.
4. In the same way 10 students were selected to measure the reliability of the test items through test re-test method as well as alternative method. The students competence and performance level of the students seemed for variation. Some students got 1-4 marks more than first administered test, some students got equal marks and 20 percent students got two marks less than first administered test. Due to the marks variation t can be concluded that the test items lacks the reliability. As like earlier subjective test items lacks reliability in comparison with objective test items. The non-textual questions lacks the reliability. Long subjective questions like essay writing, paragraph wiring, conversations and story writing lacks the reliability. But alternative method seemed more reliability than test-retest method.

5. It is found that the question paper is designed beyond the limited course so while preparing test items; the limitation should be within the designed course.
6. It can be concluded that piloting or sampling test is compulsory before administering the real test. So that all level, interest, competence and performance of the students can be addressed.
7. To make the test items more reliable TM should be focused rather than NTM.
8. In order to make the test items more reliable other methods should be applied like: alternative, split half and internal consistency. So that pure test items can be designed.
9. While designing the test item purpose of the test should be clear so that a test can be measured what it is expected to test.
10. It can be concluded that gender should be equally address so that all boys and girls can express their performance freely. It means test items should not be paradoxical and gender bias.
11. Special trainings and workshops are required for the teacher so that test items design can be uniformity.

4.2 Recommendations

By help of the students performance of the students of the grade ten, researcher's analysis and the findings the following suggestions and recommendations have been made.

- Test items should be designed to address the common competence and performance of the students.

- While designing test items maxims of teaching should be followed like: simple to complex, known to unknown, more interesting to less interesting and concrete to abstract.
- Paradoxical questions should be avoided.
- Test items should be piloted before administering test.
- It is also recommended that learners should not be confined to the limited prescribed course only. Teaching and testing need to be reformed and modified taking into account the sufficient provision to be made for exercises on test reflecting the features of English language. In order to increase their outlook of English students need to have confrontation with TM as well as NTM.
- Test items should be designed to decrease the biasness of the answer or express their own analysis as well as the biasness of examiners (teachers).
- While preparing test items gender biasness should be reduced so test items should address the both girls and boys equally.
- It can be recommended that objective test item should be focused rather than subjective test items.
- To check the answer sheet answer key should be prepared for both subjective and objective test items so that biasness can be reduced.
- It is also recommended that those test items can be used differently in different tests regarding their purpose.
- The test-designer must know the methodology a techniques of different tests. A good test designing avoids the students form dilemmatic keys for the test.

- Rubrics should be clear so that students may understand the questions and will be more directed towards the content.
- In order to develop language ability of the students the development and improvement of teaching/learning strategies, test devices are vital component. The language teacher and test designer must know what language teaching and testing is.
- Regarding to the present study, a language test is a device that tries to assess how much has been learned in a foreign language course or some part of a course. So the teaching learning materials determines the testing devices.
- After sampling the test items necessary changes should be made before administering the real test.

REFERENCES

- Abbott, G. et al. (1981). *The Teaching of English as an International Language: A Practical Guide*. London: Collins ELT.
- Allen, J.P.B. and S.P. Corder (eds.) (1973). *Techniques in Applied Linguistics*. The Edinburgh Course I. Applied Linguistics Vol. 1.2, 3. Oxford University Press. Oxford.
- Brown, H.D. (1994). *Principles of Language Learning and Teaching*. Prentice Hall.
- Carter, R. and D. Nunan (eds.) (1994). *Introducing Listening*. Penguin English.
- Crystal, D. (1991). *A Dictionary of Linguistics and Phonetics*. Black well.
- Davies, A. (1977). "The Construction of Language Tests" in Allen J.P. and A Davies (eds.) *Testing and Experimental Methods* OUP, Oxford.
- Davies, F. (1995). *Introducing Reading*. Penguin English.
- Harmer, J. (1987). *Teaching and Learning Grammar*. Longman. London.
- Harrison, A. (1983). *A Language Testing Handbook*. ELTS. Macmillan Press. London.
- Heaton, J.B. (1975). *Writing English Language Tests*. London: Longman.
- Hughes, A. (1989). *A Language Testing Handbook*. ELTS. London: Macmillan Press.
- Jakeman, V. and M. Clare (1996). *Cambridge Practice Tests for IELTS*. Cambridge: CUP.
- Lado, R. (1961). *Language Testing: The Construction and Use of Foreign Language Tests, A Teacher's Book*. Longman.
- Oller, J. (1988). *Language Tests for Schools*. London: Lonman.
- Rai, V.S. (1998). *English Language Teaching: Materials and Practice..* Kathmandu: Bhundipuram Prakashan.
- Richard et al. (1985). *The Longman Dictionary of Applied Linguistics*. London: Longman.
- Shiwakoti (1996). *An Analysis of the Reading Proficiency of the Secondary School Students: An Unpublished Thesis*. T.U., Kirtipur.

APPENDIX-A

MODEL QUESTIONS

Class: Ten

F.M.: 80

Subject: C. English

P.M.: 26

Attempt all questions.

1. Read the passage below and do the activities that follows:

Suren Majhi sat on the sandy bank of the Sun Koshi river. He had been fishing all day but had caught nothing. He was tired and the twilight was changing into darkness. Suddenly, in the distance over the river, Suren saw a bright object. It might be a helicopter, he thought, probably going to Diktel, but there was no noise. It came closer, and took on the shape of a saucer. It landed on the bank about 50 meters away.

Suren got up and walked towards the object. Once he had seen a film of a spaceship, probably this was one. It might have come from some other planet. As he went closer, a door opened, and Suren was drawn into flashing lights. He was frightened and almost jumped when he heard a voice saying, "Welcome to our spaceship, Suren. Don't be afraid. Sit down on the couch behind you."

"How do you know my name?" Suren asked.

"We are from another planet and we have been observing Earth for several months. Our devices have recorded radio and television broadcasts, and we have learned to speak English, since it is the most widely-used language. By using a long-distance listening device we learnt your name and that you are the best student in your school. We have chosen you to pass on an important message to the leaders of your nation and of the world."

A. Answer these questions.

- i. What was Suren doing when the spaceship descended?
- ii. Where did the spaceship come from and where was it going to?
- iii. Why did the people in the spaceship pick Suren up?
- iv. What did they want Suren to do?
- v. What did Suren think when he saw the object?

B. Put these sentences in the correct order.

- i. They welcomed Suren in their spaceship.
- ii. Suren was chosen to pass on an important message.
- iii. Suren saw a bright object like helicopter.
- iv. Twilight was changing into darkness.
- v. The spaceship might have come from another planet.
- vi. Suren was drawn into the vessel.

C. Write 'T' for the true and 'F' for the false statements.

- i. The fisherman had caught plenty of fish.
- ii. Suren went to the vessel to watch what there were.
- iii. When he heard his name from the spaceship he was happy.
- iv. Spaceship had the devices that had recorded radio and TV broadcasts.

2. Read the poem below and do the activities that follow.

Exile

My Shade is striped.
My air is barred
(As it was before;
Only here the bars
Do not bend with the wind.)

My breath is short
My claws are blunt
With long pacing on stone ground.

All day I xmeel
Zebra.. gazelle...
But I am full of easy meal.
I have no appetite.
Only at dusk –
When the monkeys mock from stone trees
And the buck (soft noses sniffing the air.
For the scent they no longer fear)
Come down to the mudless pool to drink –
Only a dusk I dare to dream
Eyes half closed, bending the bars
Against the sun
And plotting my return.

A. Decide if the following sentences are true or false.

- i. The speakers of the poem is not a man.
- ii. The poem is about the animals in zoo.
- iii. The poet pleads for freedom of animals.
- iv. The tiger in the cage is the symbol of confinement.

B. Answer these questions.

- i. Why are the buck no longer afraid?
- ii. Why is there no mud in the pool where monkeys drink water?
- iii. Who is the speaker in the poem?

3. Read the letter below and do the activities that follow.

35/768, Veldangi

Jhapa, Nepal

10th Feb., 2002

To

The Headmaster

Mahendra Ratna High School

Anarmani, Jhapa

Sir,

With reference to your advertisement in the Purwanchal of yesterday. I beg to apply for the post of an English teacher. As regards my qualifications and experience I bed to lay down as follows.

- i. I passed B.A. (English) and B.Ed. (English) examinations of the Tribhuvan University in 1997 and 1999, respectively and got a second class.
- ii. Teaching is the most coveted profession for me. I have been teaching in English School, a private school Devakul Veldangi, for the last four years. I can give my best in this profession.
- iii. My father was also a teacher. His noble spirit of social service inspired me to become a teacher. Hence, I take delight in teaching the students who are to become future citizens of our country.

I am a young man of 24 years, possessing an excellent physique and a sound moral character. Copies of necessary testimonials are enclosed here with for you kind perusal.

In view of the above facts, I pray that I may please be given a chance to serve in your institution. I may assure you that I'll spare no pains in the satisfactory discharge of my duties.

Hoping for favourable reply, I remain sir.

Yours faithfully

Shekhar Rijal

A. Write the words from the above letter which are opposite in meaning to the following words:

- i. question
- ii. inferior

B. Answer the following questions:

- i. What is the candidate applying for?
- ii. What is the experience of the candidate?
- iii. Why does the candidate want to be a teacher?
- iv. How is the candidate's moral character?
- v. Do you know which date the advertisement was published?

C. Write 'T' for true and 'F' for false statements:

- i. The letter is sent from Anarmani, Jhapa.
- ii. Shekhar promises to please the Headmaster with his fair service.
- iii. The name of the school where he is teaching now is Mahendra Ratna Secondary School.

D. Fill in the gaps with appropriate words:

- i. The advertisement was published in
- ii. This letter is addressed to
- iii. The candidate's qualification is
- iv. It is a specimen of an

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

Career opportunity

A highly reputed business house is in need of experienced, career oriented, hardworking and painstaking personnel's for the post of Marketing Executive. Candidate must have got Bachelors degree in Management studies.

Performance might be given to individuals with a motorbike please mail you CV inclusive of a recent photograph within 7 days of the publication date to the following address.

The Manager
P.O.B. 2162, Kathmandu, Nepal

A. Answer the following questions:

- i. Who might get preference for the post?
- ii. What are the requirements to be submitted along application?
- iii. Where can the petition be mailed to?
- iv. How many post are in required?

B. From the advertisement find the words that are close in meaning to the following words/phrases.

- i. Company
- ii. One who manage the office at all.....
- iii. Staff.....
- iv. Priority to be given

C. Read the advertisement again and fill the blanks accordingly.

This is an example of an for the post of A person having minimum qualifications required in the advertisement can for the post. Preference might get the candidate who has with him. Application must be submitted within 7 days of the

5. Write an essay in about 180-200 words on 'AIDS: A Fatal Disease.'

6. Write a few paragraph on Nature has given more gifts than calamities to mankind?

7. Develop a short composition depending on the outlines given about Eskimos of regarding their past and present candidates.

Many Eskimos..... not live as they used in houses..... factories for American Oil Companies..... Many become Americanized..... were modern and fashionable dresses schooling speaking English Now no hunting..... buy foods from shops..... even buy frozen meat and vegetables.

8. Rewrite the following sentences selecting the correct word from the brackets:

- i. Columbus..... America more than 400 years ago.
(discovered/has discovered/ have been discovered)
- ii. Have you ever by ship? (traveled/been traveling/travel)
- iii. She told me good news, the sentence is.... (past perfect continuous/simple past/present perfect)
- iv. I telephone you when he back. (comes/will come/came)
- v. My friend..... the Prime Minister yesterday. (see/have seen/aw)

- vi. The earth ... round the sun. (move/moves/has moved)
- vii. She asked him (where he lived/where did he lived/where does he live?)
- viii. He said to me, "where..... now?" (are you going/you were going/were you going?)
- ix. I said "...." (don't go/not to go/ won't go)
- x. She asked ... (please help me, to help her, helped me)
- xi. He said that he..... (has finished homework/had finished homework/ had been finished homework).
- xii. I replied that I came form the Himalaya when she said to me "." (Where you come form/Where do you come from? Where do I come from?)
- xiii. Unless you your experience, I won't improve. (practice/will practice/ practise)
- xiv. Whether she loves me or not. I her. (love/will love/ would love)
- xv. Whether they invited me or not I the meeting. (attend/will attend/ would attend)
- xvi. Unless the country.... its people educated, development is impossible (gets/ will get/would get)
- xvii. Shanti won the first prize. The passive is (A plan was made to repair the bridge/ A plan had been made to repair the bridge/A plan was being made to repair the bridge).
- xviii. A game will be played on computer. The active is (Arya will play a game on the computer/Arya will have played a game on the computer/Arya will be playing a game on the computer.)
- xix. 18 years ago I (were born/was born/ had been born)
- xx. were you taught English? (who, by whom, whose)
- xxi. Did she get her pen (fill/filled/to fill)
- xxii. I will have my watch..... (repair/repared/to repair)