

**A STUDY ON TESTING RELIABILITY IN TERMS OF
PARALLEL - FORMS METHOD**

A Thesis Submitted to the Department of English Education
in Partial Fulfilment for the Master of Education in English

Submitted by
Bhawana Thapa

Faculty of Education
Prithivi Narayan Campus, Pokhara
Tribhuvan University, Kirtipur
Kathmandu, Nepal

2010

A STUDY ON TESTING RELIABILITY IN TERMS OF PARALLEL - FORMS METHOD

A Thesis Submitted to the Department of English Education
in Partial Fulfilment for the Master of Education in English

Submitted by

Bhawana Thapa

Faculty of Education

Prithivi Narayan Campus, Pokhara

Tribhuvan University, Kirtipur

Kathmandu, Nepal

2010

T. U. Regd. No.:9-2-43086/2002

Second Year Examination

Roll No.: 480148/2008

Date of Approval of the

Thesis Proposal: Dec. 28, 2009

Date of Submission: Oct.06, 2010

DECLARATION

I hereby declare to the best of my knowledge that this thesis is original, no part of it was earlier submitted for the candidature of research to any university.

Date: _____

Bhawana Thapa

RECOMMENDATION FOR ACCEPTANCE

This is to certify that **Miss. Bhawana Thapa** has prepared this thesis entitled "**A Study on Reliability in terms of Parallel Forms Method**" under my guidance and supervision.

I recommended the thesis for acceptance.

Date:

Dr. Tara Datta Bhatta
Reader
Department of English Education
Faculty of Education
Prithvi Narayan Campus
Pokhara, Nepal

RECOMMENDATION FOR EVALUATION

This thesis has been recommended for evaluation from the following

‘Research Guidance Committee’:

Signature

Mr. Amirman Shrestha

Reader and Head

Department of English Education

Prithvi Narayan Campus,

Pokhara Nepal

.....

Chairperson

Dr. Tara Datta Bhatta (Guide)

Reader

Department of English Education

Prithvi Narayan Campus,

Pokhara Nepal

.....

Member

Mr. Tritharaj Aryal

Reader

Department of English Education

Prithvi Narayan Campus,

Pokhara Nepal

.....

Member

Date:

EVALUATION AND APPROVAL

This thesis has been evaluated and approved by the following 'Thesis Evaluation and Approval Committee':

Signature

Mr. Amirman Shrestha

Reader and Head
Department of English Education
Prithvi Narayan Campus, Pokhara

Chairperson

Dr. Jai Raj Awasthi

Professor
Department of English Education
Chairperson
English and Other Foreign Languages Education
Subject Committee
TU, Kirtipur

Expert

Dr. Tara Datta Bhatta (Guide)

Reader
Department of English Education
Prithvi Narayan Campus, Pokhara

Member

Date:

ACKNOWLEDGEMENTS

I am very much indebted to **Dr. Tara Datta Bhatta**, my honorable teacher as well as my research guide, whose painstaking efforts in providing me with scholarly guidance and constant encouragement and invaluable support both at professional and personal levels made this study feasible to complete in time. His patience, enthusiasm, co-operation, suggestions, and keen interest in this study are ever memorable. His vigorous efforts made me present this research work in this form.

I would also like to express my sincere gratitude to **Mr. Amirman Shrestha**, the head of the department of English Education, P.N.C, Pokhara, for his invaluable suggestions and encouragement.

I am also grateful to **Mr. Tritha Raj Aryal**, **Mr. Bishnu Hari Timilsana**, **Yam Bahadur Chhetri**, and other members of the department for their encouragement, cooperation and suggestions. Similarly, I would like to express my sincere gratitude to Guru of Gurus **Prof. Dr. Jai Raj Awasthi** Chairperson of English and Other Foreign Languages Education Subject Committee TU, Kirtipur for his encouragement and suggestion.

I am extremely thankful to all my dear students, teachers, and principal of Pokhara United Academy, Kalika Higher Secondary School, Shree Amar Singh Higher Secondary School, and Metropolitan Academy. Furthermore, I am equally thankful to my friends **Narayan Sapkota**, **Sulochana Bhandari**, **Binju Gurung**, **Shankar Timilsana**, and **Laxmi Pant**. Finally, deep appreciation is felt for my sisters **Ranu Thapa** and **Amrita Thapa**, brother **Navin Thapa** and all my well-wisher for their encouragement and co-operation in this work.

Date:

Bhawana Thapa

ABSTRACT

This present study entitled 'A Study on Testing Reliability in terms of Parallel Forms Method' aims at finding out the reliability in terms of parallel forms method. The researcher administered pre test and post test to the students of Pokhara United Academy, Kalika Higher Secondary, Shree Amar Singh Higher Secondary and Metropolitan Academy. The total sample of the study was 80 students of the selected schools. The sample of population was done on the basis of the judgmental sampling procedure. It was done with respect to school, gender and item wise. The researcher using primary and secondary sources of data did this research. Results showed that the reliability of parallel-forms method in assessing reading skill is high. The reliability of the test item differs from school to school, item-to-item and even male and female students. Results of item-wise comparison showed that reliability of test in true –false and fill-in-the blanks items was high whereas in matching item, multiple-choice item, rearrange and short answer questions was satisfactory. Moreover, results of true-false item showed the highest reliability whereas short answer questions showed the lowest reliability. Results of personal comparison showed that only a few students got the same marks in both tests. Due to the marks variation among the students, it was concluded that the test item lacked reliability.

The thesis has been divided into four chapters. The first chapter of the study deals with the introduction part, which includes general background, the English language in Nepal, Language teaching and testing, reading skill, testing reading, quality of a good language test, validity, reliability, parallel forms method, review of related literature, objectives of the study and significance of the study. The second chapter deals with the methodology used in the study. It includes source of data, population of study, sampling procedure, tools of data collection, process of data collection, limitation of the study etc. Chapter three is the main part of study that includes analysis and interpretation. Last chapter deals with the findings and recommendations made after the analysis of data.

TABLE OF CONTENTS

	<i>Page No.</i>
Declaration	i
Recommendation for Acceptance	ii
Recommendation for Evaluation	iii
Evaluation and Approval	iv
Dedication	v
Acknowledgements	vi
Abstract	vii
Table of Contents	viii
List of Tables and figures	x
List of Abbreviations	xi
CHAPTER- ONE: INTRODUCTION	1-16
1.1 General Background	1
1.1.1 The English Language in Nepal	2
1.1.2 Language Teaching and Testing	3
1.1.3 Reading Skill	5
1.1.4 Testing Reading	6
1.1.5 Qualities of a Good Language Test	7
1.1.5.1 Validity	7
1.1.5.2 Reliability	8
1.1.6 Parallel- Forms Method	11
1.2 Review of the Related Literature	13
1.3 Objectives of the study	16
1.4 Significance of study	16
CHAPTER-TWO: METHODOLOGY	17-19
2.1. Sources of Data	17
2.2 Population of the Study	17

2.3	Sample of Population	18
2.4	Sampling Procedure	18
2.5	Tools of Data Collection	18
2.6	Process of data Collection and Interpretation	18
2.7	Limitations of the Study	19
CHAPTER-THREE: ANALYSIS AND INTERPRETATION		20-34
3.1	Holistic Description of Reliability	21
3.2	School-wise Description of Reliability of Parallel Forms Method	21
3.3	Gender-wise Description of Reliability	23
3.4	Item-wise Description of Reliability	25
3.4.1	Sentence- Matching Items	25
3.4.2	True-False Item	26
3.4.3	Fill in the Blanks Items	27
3.4.4	Multiple Choice Items	29
3.4.5	Correct Order	30
3.4.6	Short Question –Answer	31
3.5	School-wise Description of Reliability	32
3.5.1	Private Schools	33
3.5.2	Government aided Schools	33
CHAPTER-FOUR: FINDINGS AND RECOMMENDATIONS		35-37
4.1	Findings	35
4.2	Recommendations	36

REFERENCES

LIST OF TABLES AND FIGURES

	<i>Page No.</i>
Table 1: Holistic Description of Reliability (Male +Female)	21
Table 2: School- wise Description	22
Table 3: Gender-wise Description	23
Table 4: Sentence- Matching Items	25
Table 5: True-False Items	26
Table 6: Fill in the blanks items	28
Table 7: Multiple Choice Items	29
Table 8: Correct Order	30
Table 9: Short Question –Answer	31
Table 10: Private School	33
Table 11: Government aided School	33
Figure I: Parallel Forms Method	13

LIST OF ABBREVIATIONS

e.t.c	Et.Cetera
i.e.	that is
KHSS	Kalika Higher Secondary School
MPA	Metropolitan Academy
PUA	Pokhara United Academy
r	Reliability
S.N.	Serial Number
SAHSS	Shree Amar Singh Higher Secondary School
St.dev	Standard Deviation

CHAPTER ONE

INTRODUCTION

1.1 General Background

Language is a voluntary vocal system of human communication. Communication means exchange of message, information, or ideas etc. Therefore, language is a means through which information and ideas are exchanged among human beings. We exchange our ideas, thoughts, feelings, desires, emotions, expressions etc by the help of language. In the process of communication, one perceives the clear picture of the whole world through the language. Language is the symbolic system based on pure or arbitrary convention infinitely extendable and modifiable according to the changing needs and condition of the speakers. Language organizes sounds and vocal symbols with meaningful unit. Language is a system of systems. There are four levels in every language. They are phonological, morphological, syntactic, and semantic levels. Language is the most valuable single possession of human race. It is a social phenomenon, which is used in our society to establish the human relationship. It is difficult to trace back the origin of language but it can be assumed that the development of language is as old as human race.

The English language is one of the languages of the Indo-European language family. The Indo-European family is the most widely used language in the world. This is the single largest language family. Indo-European language family has about 150 languages and about three billion speakers. Languages include Hindi and Urdu (400 million), Bengali (200 million), Spanish (300 million), Portuguese (200 million), French (100 million), German (100 million), Russian (300 million), and English (400 million) in Europe and the Americas. The Indo-European languages stretch from the America through Europe to North India. The Indo-European Family is thought to have originated in the forests north of the Black Sea (in what is now Ukraine) during the Neolithic

period (about 7000BC). These people began to migrate between 3500BC and 2500BC, spreading west to Europe, south to the Mediterranean, north to Scandinavia, and east to India. The Indo-European Family is divided into twelve branches, ten of which contain existing languages. The Germanic Branch is one branch of Indo-European family. They include English, the second most spoken language in the world, the most widespread, the language of technology, and the language with the largest vocabulary. (as cited in www.languagefamily.net)

The English language is commonly used language all over the world. Therefore, it is an International language. The English language connects all the countries of the world with each other. English is most important in every field and sector of the world. This language is the language of communication, politics, marketing, trade, medicine, science, and technology. Without the English language, we will not be able to interact with other people who are unknown about our language. The English language for these days becomes a basic need for human beings. People must learn language for survival. Science and Technology is the main part of today's world. To know about these, we need English because about half of the world's books of different subjects have been written in English. Due to the modernization of world, the need and scope of the English language is increased day by day. By these reasons English has the highest position. This language is spoken as a native language or official language in many countries.

1.1.1 The English Language in Nepal

The English language is taught and learnt as a foreign language of global significance. That is why, much priority has been given to it over foreign languages. According to Panday (1999, p. 243) in Nepal, English was first introduced in Durbar school on 27th Asoj, 1910 by Jang Bahadur Rana. The educational system in India has always influenced the Nepalese curriculum when Tribhuvan University started to prescribe its own syllabuses and conduct its own examina-

tions in 1960; there were not much to choose between Patna and Tribhuvan University. In fact, the opening of Tri-Chandra College in 1918 A.D. marks the formal beginning of English in Higher Education in Nepal. In these days, literature used to be focused on literacy text study. Literacy text were perused, analyzed and critically studied. Mainly, reading and writing were the language skills given precedence over others. Nowadays, English is used as a medium of instruction due to the establishment of English medium schools. Although, it is taken as a medium of instruction, there is lack of its proper use. Due attention has not been paid to develop the four skills equally in child. The students from boarding background are able to speak and write frequently but those who are from government aided school background and remote area are not able to speak. Education board always takes this process as an experiment and are unable to provide enough textbooks and materials. Therefore, despite experiments, 'weak' textbooks and various other factors led to the present no-so-happy scenario.

1.1.2 Language Teaching and Testing

Language teaching is an art. It can be defined as the teaching of language for communicative purpose. Language teaching is an application of language theories, methods, and techniques in actual teaching and learning activities. Language teaching involves teaching of both first and second/foreign language.

Testing is an inherent part of teaching. We cannot separate testing from teaching. Both teaching and testing go together. Teaching and testing are the two parts of the same coin. Testing is inevitable for we have to assess the achievement of the students as well as the effectiveness of the programme if self-test may be constructed primarily as a means of assessing the student's performance in the language.

According to Heaton (1975, p. 1) language testing, in the past, was considered as a separate entity from teaching. But both testing and teaching are so closely in-

terrelated that it is virtually impossible to work in either field without being constantly concerned with each other. Testing is conducted for finding out the performance of the students of the purpose of comparison, selection, gradation, etc., testing is used to find out how far learners have learned what the teacher wishes them to learn, whether the learner understands what has been taught or not.

With the help of language testing, the teacher will be able to compare, select, grade and provide certificate among his/her students that is why language testing is important for language teaching. We can also evaluate the effectiveness of teaching style, syllabus, methods, approaches, techniques materials etc. with the help to language testing. The classroom test is concerned with evaluation for enabling teachers to increase their own effectiveness by making adjustments in their teaching to enable certain groups of students or individuals in the class to benefit more. It tries to find out what activities the candidates can do. A well-constructed classroom test will provide the students with an opportunity to show their ability to perform certain task in the language. The test should also enable the teacher to ascertain which parts of the language programme have been found difficult by the class. The testing can be conducted after teaching, during the teaching time or before teaching to know the level of the students. Testing is equally important for remedial measures. Testing offers useful inputs to the teachers to be aware of the effect of his teaching. Testing has become an instrument to contribute to change in the way teachers perform in the classroom, and the way reform takes place in the whole system of education.

Language plays an important role in learning and communication. Since language is the medium of instruction of communicating meaning in the most learning activities. Language comprises four skills: listening, speaking, reading, and writing i.e. LSRW. Listening and speaking skills are basic to language learning. The more the students listen and speak, the more they learn to communicate. Reading and writing are secondary skills, which are used to commu-

nicate and express the acquired knowledge. Thus, the language is such a skill through which children can communicate words to survive in the society. Learning of language not only facilitates communication but it also serves as a powerful medium of acquisition of knowledge. Although, all the language skills are equally important, for the present study only reading skill was taken.

1.1.3 Reading Skill

Reading skill can be defined in many ways. It is the amalgamation of visual and non-visual experience or behaviour, reading is decoding print, “deciphering” print, or reading is for understanding, interpreting or making sense of a given text etc. How every reading in the class room for easy understanding can be defined as making sense of a given text, the process during which all other possible meaning of “reading” include themselves.

Reading may mean many things and it takes on a different meaning in different contexts. We read in order to obtain information that is presented in a right form, but the nature of the information obtained demand definition that is more explicit. Information means the content, which is cognitive. We read referential material in order to obtain information with which to operate in our environment.

Reading and reading comprehension are interchangeably used in a language, because understanding a text means comprehending it. Reading comprehension seeks only relevant information. Reading comprehension is extracting required information from the written text as efficiently as possible. Reading comprehension varies according to the purpose of reading and type of the text, reading comprehension is the combination of some reading sub- skills.

According to Munby (1979 as cited in Khaniya, 2005, pp.142 - 45),the following are the sub-skills, which exist in reading comprehension:

-) Recognizing the script of language.

-) Deducing the meaning and use of unfamiliar lexical items.
-) Understanding explicitly stated information.
-) Understanding information not explicitly stated.
-) Understanding conceptual meaning.
-) Understanding the communicating value (function) of sentences and otherness.
-) Understanding relations between the parts of a text through lexical cohesion devices.
-) Understanding cohesion between parts of a text through grammatical cohesion devices.
-) Interpreting text by going outside it.
-) Recognizing indicators in discourse.
-) Identifying the main points or important information in a piece of discourse.
-) Distinguishing the main idea from supporting details.
-) Extracting salient points to summarize (the, text, an idea, etc).
-) Selecting extraction of relevant points from a text.
-) Basic reference skills.
-) Skimming
-) Scanning
-) Scanning to locate specifically required information.
-) Transcoding information for a diagrammatic display.

1.1.4 Testing Reading

Testing reading depends on the purpose of test and level of learners. It starts with the recognition of script of a language to complex reading like understanding conceptual meaning, understanding relations within the sentence etc. Therefore, we must be aware of testing reading. Before testing the learners question setter must be familiar with reading sub-skills as well which are involved in reading skills also. It is useful to include a variety of text type for

reading comprehension as far as possible. While testing reading we must try to test only reading skill nothing else. For example, essay type questions test learner's writing skill that is why they are not suitable for reading skill. We can use different test for testing reading. Some of them are as follows:

-) Multiple-choice item
-) Matching item
-) Completion item
-) Rearrangement item
-) Short answer questions etc.

1.1.5 Qualities of a Good Language Test

Several things should be taken into consideration while constructing a test. Below are some qualities, which are essential to develop test items. Some scholars point only a few qualities and some pointed many but there are some qualities, which are common to all. These qualities are as follows.

- a. Validity
- b. Reliability
- c. Administrability
- d. Scorability
- e. Economy

Validity and Reliability are the most important qualities of a good test. Thus, it is important to mention about the validity and reliability.

1.1.5.1 Validity

Kumar (1999, p. 137) defines validity as the ability of an instrument to measure what it is designed to measure. Validation of an exam would imply whether the exam is sufficient to elicit the ability to perform the given tasks that become the main concern in validation process. Validity measures how far the information it provides is accurate. According to Hughes (1939, p. 22), "A test is said to be valid if it measures accurately what it is intended to measure." Similarly, for

Heaton (1975, p. 159), the validity of a test is “The some extent to which it measures what it is supposed to measure and nothing else”.

According to this view, a measure will be valid if it does what it intends to do. Validity relates to whether a research instrument is measuring what it sets out to measure.

Therefore, validity can be defined as the outcome of the test which reveals its aim i.e. purpose of a test. A valid test reflects the real image of test what is intending to measure. Every test, whether it is a short, informal classroom test or a public examination, should be as valid as the test constructor can make it. In order to achieve this goal, objectives of the test should be clearly stated, breaking them down in the skills and abilities and define them in separate items and assess them in situations which are closely related to the real life situations in which they will be used.

There are five types of validity. They are as follows:

- a. Face validity
- b. Construct validity
- c. Concurrent validity
- d. Predictive validity, and
- e. Construct validity

Although validity and reliability are equally important qualities of a good test, for the present study only reliability was taken.

1.1.5.2 Reliability

Reliability refers to the consistency of scores or performance of the same or similar test administered within a reasonable time. It is a matter of the extent to which we can believe that the performance is true, how likely it is that the performance will be repeated next time. Reliability can be dealt with at two levels: test and retest of students and marking and remarking to the examiners. If the scores of two tests are similar or almost similar at that time we can say that the

test is more reliable. “To be reliable, a test must be consistent in its measurement” Heaton (1975, p. 162). It is assumed that the performance of the examinee remains more or less the same if the examinee is asked to repeat the same exam of the similar types. According to Kumar (1966, p. 140), “Reliability is the degree of accuracy or precision in the measurements made by a research instrument.” If we administer an instrument under the same or similar conditions to the same population and obtain similar results, then only it is reliable “The more similar the scores would have been, the more reliable the test is said to be” (Hughes 1989, p. 29).

Reliability is of primary importance in the use of both public achievement and proficiency test and classroom tests. Reliability is necessary characteristics of any good test for it to be valid. The consistency in the performance is determined by carrying out some statistical analysis that is the reliability coefficient. The factors that contribute to the reliability of the test are as follows:

-) Homogeneity of items if a test has the test items testing more or less the same trait, the test will have high reliability.
-) A test with high discriminating power items will produce high reliability.
-) Variability of group-students with a wide range of ability will yield high reliability.
-) Sufficient test taking time will give high reliability.
-) A test with less freedom of choice will yield high reliability.
-) A test with unambiguous item will have high reliability.
-) Objectivity in scoring will give high reliability.
-) Length of the test also contributes to produce high reliability; longer the test, higher the reliability.

(Source: Khaniya 2005, p. 109)

According to Heaton (1975, pp. 162-163), the following are some factors, which may affect the reliability of a test.

-) The extent of the sample of material selected for testing where as validity is concerned chiefly with the content of the sample, reliability is concerned with the size. The larger the sample (i.e. more tasks the testees have to perform), the greater the probability that the test as a whole is reliable-hence the favoring of objective test, which allows a wide field to be covered.
-) The administration of the test is the same test administered to different groups under different conditions or at different times. Clearly, this is an important factor in deciding reliability, especially in tests of oral production and listening comprehension.
-) Test instructions: Are the various tasks expected from the testees made clear to all candidates in the rubrics?
-) Personal factors such as motivation and illness.
-) Scoring the test: One of the most important factors affecting reliability. Objective tests overcome this problem of marker reliability, but subjective tests are sometimes faced with it, hence the importance of the work carried out in the fields of the multiple marking of compositions and in the use of rating scales.

Regarding the methods of establishing the reliability of a test, several methods are discussed:

-) One method of measuring the reliability of a test is to readminister the same test after a lapse of time.
-) Next method of measuring the reliability of a test is by dividing the marks of a single test into two halves and correlating them.
-) Another means of estimating the reliability of a test is by administering parallel forms of the test to the same group.
-) Moreover, the reliability of the whole test can be estimated by using

the formula.
$$r_{xy} = \frac{N}{N-1} \left(1 - \frac{m(N \sum m)}{N \sum x^2} \right)$$

Where,

N = the number of items in the test,
m = the mean score on the test for all the testees,
X = the standard deviation of all the testees' scores
 r_{xy} = Reliability

Heaton, (1975, p 164).

According to Kumar (1996, pp. 141 – 42) there are three methods to find our reliability of test. They are:

- a. Test - retest method.
- b. Split - half method.
- c. Parallel forms method.

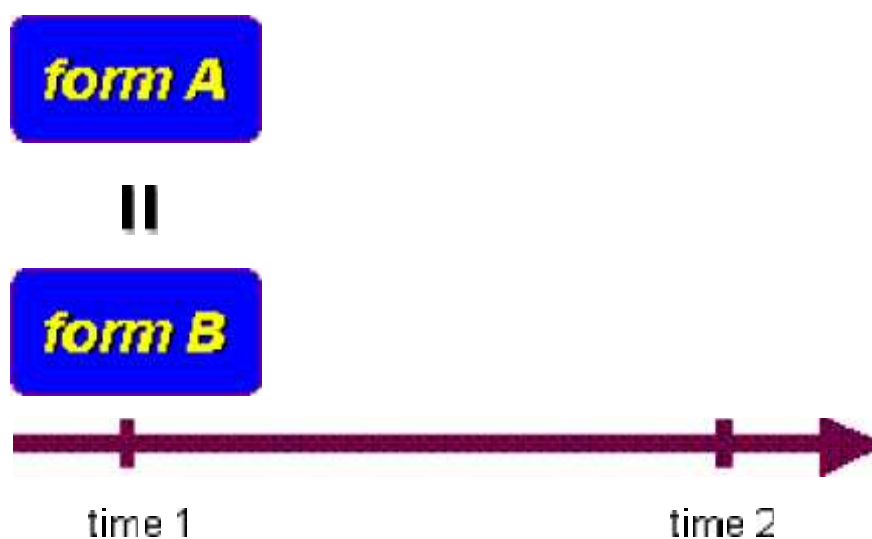
For the present study only parallel form method was chosen.

1.1.6 Parallel- Form Method

This is one of the ways of obtaining reliability. In this method, the reliability of the test is worked out by administering parallel forms of the test to the same group. This assumes that two similar versions of a particular test can be constructed, such tests must be identical in the nature of their sampling, difficulty, length, rubrics, etc. Only after a full statistical analysis of the tests, they can safely be regarded as parallel. If the correlation between the two tests is high (i.e. if the results derived from the two tests correspond closely to each other), then the tests can be termed reliable. In parallel forms, reliability, we first have to create two parallel forms.

One way to accomplish this is to create a large set of questions that address the same construct and then randomly divide the questions into two sets. We administer both instruments to the same sample of people. The correlation between the two parallel forms is the estimate of reliability. One major problem with this approach is that you have to be able to generate many items that reflect the same construct. This is often no easy feat. Furthermore, this approach assumes that the randomly divided halves are parallel or equivalent. Even by chance, this will sometimes not be the case. The parallel forms approach is very similar to the split-half reliability described below. The major difference is that parallel forms are constructed so that the two forms can be used independent of each other and considered equivalent measures. For instance, we might be concerned about a testing threat to internal validity. If we use Form A for the pretest and Form B for the post-test, we minimize that problem. It would even be better if we randomly assign individuals to receive Form A or B on the pre-test and then switch them on the post-test. With split-half reliability, we have an instrument that we wish to use as a single measurement instrument and only develop randomly split halves for purposes of estimating reliability. (as cited in www.socialresearchmethods.net)

Figure1: Parallel- Forms Method



1.2 Review of the Related Literature

This part presents review of related studies. The purpose of this part is to review research studies or other relevant proposition in the related area of the study and to identify the trends of research. There are a number of studies related to reliability. Among them, some studies were studied to present this study. The reviews of those studies are presented below.

Cyril (2006) conducted a study on “Establishing test form and individual task comparability: a case study of a semi-direct speaking test”. He investigated the extent to which three forms of the General English Proficiency Test, Intermediate Speaking Test (GEPTS-I) are parallel in terms of two types of validity evidence: parallel-forms reliability and content validity. The three trial test forms, each containing three different task types (read-aloud, answering questions and picture description), were administered to 120 intermediate-level EFL learners in Taiwan. The performance data from the different test forms were analyzed using classical procedures and Multi-Faceted Rasch Measurement (MFRM). Various checklists were also employed to compare the tasks in different forms qualitatively in terms of content. The results showed that all three-test forms were statistically parallel overall and Forms 2 and 3 could be considered parallel at the in-

dividual task level. Moreover, the checklists identified sources of variation to account for the variable difficulty of tasks in Form 1. Results of the study provide insights for further improvement in parallel-form reliability of the GEPTS-I at the task level and offer a set of methodological procedures for other exam boards to consider.

Likewise, **Batia and Nation** (1999) quoted in the article “A vocabulary-size test of controlled productive ability” that it is important in the design of the vocabulary component of a teaching program that teachers are able to discover the state of their learners’ vocabulary knowledge. It is also important that researchers can draw on a variety of vocabulary measures to investigate the nature of vocabulary growth. This study focuses on a controlled production measure of vocabulary consisting of items from five frequency levels, and using a completion item type like the following. The garden was full of flowers. The controlled-production vocabulary-levels test was found to be reliable, valid (in that the levels distinguished between different proficiency groups) and practical. There was a satisfactory degree of equivalence between two equivalent forms.

In Nepal, so many studies have been carried out in the field of reliability in the Department of English Education. Some of them are as follows.

Shiwakoti (1996) in his M.Ed, thesis entitled, ‘An analysis of the reading proficiency of the secondary school students of Jhapa District’ focused on closed test. He found that private school students performed better than government-aided school students. He pointed out the reasons that the extra supplementary materials better facilitated the private school students than government aided school students. The urban students perform better than the rural schools students.

Khanal (1997) had done a study entitled, “A study on the effectiveness of the close test over conventional objective test in testing reading comprehension in English.” He found that the private schools performed better than the public schools in both objectives and the close test, it is not surprising that the private school performed better in non-textual materials and it was found more effective and reliable. So, it was recommended for the school of Nepal.

Baral (1999) had done a study on “Measurement of readability of a course in general English.” He found that the difficulty level of forty-six passages in general English and ranks them in terms of their readability level.

Adhikari (2007) carried out a study on, “The effectiveness of test retest method to measure reliability of the test item.” He found out that objective questions were more reliable than subjective questions. Long subjective questions like essay writing, paragraph writing, conversation and story writing lack the reliability. In his study, he suggested that objective test items should be focused rather than subjective test item.

Shrestha (2007) carried out a study on, “Reliability using test -retest method.” She found out that the reliability of test-retest method in assessing writing skill is high. Moreover, in writing test the reliability in controlled writing is high. The reliability of test retest method in sentence matching, guided writing and free writing is also high. She suggested that test retest method should be used to ensure the reliability of test items in assessing writing skill.

Dahal (2009) carried out a study on “A study on Reliability in terms of split -half method.” He found that the reliability of split-half method in assessing reading skill was high. Moreover, the reliability of the test item differed from school to school item to item and even between male and female student.

By the above mentioned detail, we know that there were many studies in this field but no one tried to carry out the research on reliability in terms of parallel forms method. Therefore, the researcher has selected this topic for the present study.

1.3 Objective of the Study

The study was carried out with the following objectives:

1. To find out the effectiveness of parallel forms method to measure reliability of the reading skill.
2. To compare the reliability of the parallel forms method in terms of following variables:
 - a. School wise
 - b. Gender wise
 - c. Item wise
3. To suggest some pedagogical implications.

1.4 Significance of Study

Any interested candidates of language testing will take benefit from this work. It will also be significant to the language teachers, students, researchers, question setters, examiners etc. It will play an important role in the area of language teaching and testing. Hopefully, it will be equally important for those who are involved in the area of psychology, research methodology and so on.

CHAPTER TWO

METHODOLOGY

Methodology is the process to find out new facts and information about a particular subject matter. Methodology helps to find out effective and reliable conclusion that is why it is considered as a vehicle for carrying out any information. To accomplish this, the following methods and procedures were employed.

2.1 Sources of Data

To accomplish the study, the following sources of data were used:

2.1.1 Primary Source

Test scores of 80 students of class nine who had successfully passed class eight from selected four schools were the primary sources of data.

2.1.2 Secondary Sources

The secondary sources of data were the related books, like Lado (1961), Heaton (1975), Huges (1989), Khaniya (2005), etc., and articles, journals, magazines, e-mail internet and the theses approved by the Tribhuvan University and all the related materials which were available in print and electronic media.

2.2 Population of the Study

The population of the study was eighty students of class nine who had successfully completed the examination of class eight from selected schools of Pokhara valley.

2.3 Sample Population

For every study, we need sample population. For this study, the sample population of 80 students was taken from Shree Kalika Higher Secondary School, Amar Singh Higher Secondary School, Pokhara United Academy and Metropolitan Academy, Pokhara.

2.4 Sampling Procedure

The researcher employed judgemental sampling procedure, which is one of the useful non-probabilities sampling designs. For which the researcher selected four schools, among them two were government aided schools and two were private schools. Similarly, 10 girls and 10 boys from each school were selected by random sampling procedure.

2.5 Tools of Data Collection

The researcher collected two sets of parallel test forms herself based on District Level questions (see appendix 1). The test items were designed to assess only reading skill. One passage for each test was selected. The test included both subjective and objective type of questions carrying 30 marks each. Matching, true-false, fill in the blanks, re-arranging and short answer questions were included in the question.

2.6 Process of Date Collection

The researcher visited the selected schools (Shree Kalika Higher Secondary School, Amar Singh Higher Secondary School, Pokhara United Academy and Metropolitan Academy, Pokhara) to collect the data. She explained the purpose of her visit and objectives of her study to the school authorities. When she got the permission, she went to the classroom. She introduced herself and gave necessary instructions to the students about the test. She conducted the test by the help of subject teachers of those schools. One hour thirty minutes time was given to the students. Most of the students finished the tests within the given

time. After finishing the tests, she collected the answer sheets. She prepared the answer key and checked the answer sheet. She calculated the marks.

2.7 Limitation of Study

The researcher could not cover a wide range of validity and reliability, so this study had the following limitations

-) This research was limited to only four schools of Pokhara Valley. (i.e. Shree Kalika Higher Secondary School, Amar Shing Higher Secondary School, Pokhara United Academy and Metropolitan Academy, Pokhara).
-) Only the students of grade-9 were taken.
-) Only 80 students were the sample population of the study.
-) Only parallel method is used in this study.
-) Only reading skill was measured.
-) The test items were prepared from grade 8 English Test book.

CHAPTER THREE

ANALYSIS AND INTERPRETATION

The most important part of any research is the analysis and interpretation of the data. Thus, this chapter is concerned with the analysis and interpretation of the data obtained from the students. For the present study, two reading passages with objective and subjective questions were selected. The test was administered to the students of the selected schools and scores were calculated. The different statistical tools were used to analyze and interpret the data. The various statistical tools used to analyze and interpret the data are: descriptive statistics i.e. Mean, Standard Deviation, and Reliability.

The researcher has analyzed the data with respect to the following variables:

- a) School wise
- b) Gender wise
- c) Item-wise

To analyze data, the researcher has used the following formula:

$$a. r_{xy} = \frac{N(\sum xy) - (\sum x)(\sum y)}{\sqrt{[N \sum x^2 - (\sum x)^2][N \sum y^2 - (\sum y)^2]}}$$

Where,

r_{xy} = correlation between x and y

N = number of given scores

x and y = individual scores

- b. Correlation coefficient

$r_{xy} < 0.3$ poor correlatives

$0.3 < r_{xy} < 0.7$ moderate/ Satisfactory

$0.7 < r_{xy}$ highly correlated

3.1 Holistic Description of Reliability

Eighty students of grade nine of four selected schools in Pokhara valley were involved for this study. Both subjective and objective test of 30 marks were used in each school. The following table 1 presents the holistic description of reliability.

Table No. 1
Holistic Description of Reliability

	Test 1	Test 2
Numbers of testees	80	80
F.M.	30	30
Mean	25.7	24.13
Reliability	0.8	

The above table 1 shows that there were eighty students as a whole and full mark of the test was thirty in both tests. As per the table, the mean of the test 1 is 25.7 and test 2 is 24.13. The reliability of the test is 0.8 which means the reliability of parallel forms method is high. It means as a whole, all the testees of class nine have done well.

3.2 School-wise Description of Reliability of Parallel Forms Method

Four schools were selected to obtain this information. The following table presents the descriptive statistics of the tests and reliability of parallel forms method.

Table No. 2
School- wise Description

S.N	Name of school	No.	Test 1		Test 2		r
			Mean	St. Dev	Mean	St. Dev	
1	PUA	20	27.85	1.98	26.95	2.25	0.77
2	KHSS	20	20.2	4.76	17.9	3.8	0.75
3	SAHSS	20	26.8	1.43	25.25	1.06	0.46
4	MPA	20	26.45	2.52	26.6	.2.66	0.14

(Individual scores of each student are given in appendices III, IV and V)

The above table shows that there were twenty students who took the tests from PUA. The mean of the test1 is 27.85 with the 1.98 standard deviation and test 2 is 26.95 with the 2.25 standard deviation and the reliability is 0.77 which means the reliability of the parallel forms method is highly correlated.

Twenty students of KHSS were given the two parallel forms test to find out the reliability of the parallel forms method in reading skill. As per the table, there were twenty students in the school, who attended the tests. From the score they got, the mean of the test1 is 20.2, test 2 is 17.9 standard deviation of test 1 is 4.76; test 2 is 3.8 and the reliability is 0.75 that means the reliability of the parallel forms method is highly correlated.

We find that there were twenty students in the school, who attended the tests from SAHSS. From the score they got, the mean of the test1 is 26.8, test 2 is 25.25, standard deviation of test 1 is 1.43, and test 2 is 1.06 and the reliability is 0.46 that means the reliability of the parallel forms method is moderate.

Twenty students took the tests from MPA. From the score they got, the mean of the test1 is 26.45, test 2 is 26.6, standard deviation of test 1 is 2.52, test 2 is 2.66, and the reliability is 0.14 that means the reliability of the parallel forms method is poor.

3.3 Gender-wise Description of Reliability

To obtain gender wise description of the male and female students of the same selected schools, the tests were administered to the students of each school i.e. ten were boys and ten were girls. The table 3 presents the gender wise description.

Table No. 3
Gender-wise Description

S.N	Name of school	Sex	No	Test 1		Test 2		r
				Mean	St.dev	Mean	St.dev	
1	PUA	F	10	29	0.81	26.9	0.98	0.6
		M	10	26.7	2.16	27	3.12	1
2	KHSS	F	10	21.1	3.60	17.7	2.62	1.71
		M	10	19.3	5.75	18.1	4.95	2.09
3	SAHSS	F	10	26.2	1.22	25.5	0.7	0.21
		M	10	27.4	1.42	25	1.33	0.73
4	MPA	F	10	25.1	2.68	25.1	2.50	0.1
		M	10	28.4	1.71	26.6	2.39	0.55

(Individual scores of each student are given in appendices vii -xiv)

The above table shows that ten female students attended the tests from PUA. From the score they got, the mean of the test1 is 29, test 2 is 26.9, standard deviation of test 1 is 0.81, test 2 is 0.98, and the reliability is 0.6, which means the reliability of the parallel forms method is moderate.

Ten male students attended the tests from PUA. From the score they got, the mean of the test1 is 26.7, test 2 is 27, standard deviation of test 1 is 2.16, test 2 is 3.12, and the reliability is 1 that means the reliability of the parallel forms method is highly correlated.

According to the table, ten female students attended the tests from KHSS. From the score they got, the mean of the test 1 is 21.1, test 2 is 17.7, standard deviation of test 1 is 3.60, test 2 is 2.62, and the reliability is 1.71 that means the reliability of the parallel forms method is highly correlated.

Ten male students of KHSS were given the test to find out the reliability of parallel forms method of reading skill. From the score they got, the mean of the test 1 is 19.3, test 2 is 18.1, and standard deviation of test 1 is 5.75, test 2 is 4.95, and the reliability is 2.09, which mean the reliability of the parallel forms method is highly correlated.

Ten female students of SAHSS were given the parallel form test to find the reliability of reading skill. From the score they got the mean of the test 1 is 26.2, test 2 is 25.2, standard deviation of test 1 is 1.22, test 2 is 0.7 and the reliability is 0.21, which means the reliability of the parallel forms method is poor.

There were ten male students in SAHSS who attended the tests. From the score they obtained the mean of the test 1 is 27.4, test 2 is 25, standard deviation of test 1 is 1.42, test 2 is 1.33, and the reliability is 0.73, which means the reliability of the parallel forms method, is highly correlated.

Ten female students attended the tests from MPA. From the score they got, the mean of the test 1 is 25.1, test 2 is 26.4, standard deviation of test 1 is 2.68, and test 2 is 2.50 and the reliability is 0.1 which means the reliability of the parallel forms method is poor.

Ten male students attended the tests from MPA. From the score they got, the mean of the test 1 is 28.4, test 2 is 26.2, standard deviation of test 1 is 0.71, test 2 is 2.39, and the reliability is 0.55 that means the reliability of the parallel forms method is moderate.

3.4 Item-wise Description of Reliability

Five test items were administered to the students of the same selected schools to find out reliability of parallel forms method. Item -wise description is presented below:

3.4.1 Sentence- Matching Items

The students of the selected schools were given same test to find out reliability of parallel forms method in matching item of the students. The table is presented below.

Table No. 4
Sentence- Matching Items

S.N	Name of school	No	Test 1		Test 2		r
			Mean	St.dev	Mean	St.dev	
1	PUA	20	4.9	0.44	3.9	1.02	0.59
2	KHSS	20	2.1	1.20	1.62	0.98	1.01
3	SAHSS	20	4.9	0.44	2.95	0.22	0.41
4	MPA	20	4.4	1.14	4.5	0.88	0.47

As per the table, PUA has different mean and standard deviation in sentence matching items. It has been found that the mean of test 1 is 4.9, test 2 is 3.9 and the standard deviation of test 1 is 0.44, test 2 is 1.02 and the reliability of tests is 0.59, which means the reliability of sentence matching items is satisfactory.

KHSS has different mean and standard deviation in sentence matching items. It has been found that the mean of test 1 is 2.1, test 2 is 1.62 and the standard deviation of test 1 is 1.20, test 2 is 0.98 and the reliability of tests is 1.014, which means the reliability of sentences matching items, is highly correlated.

Likewise, SAHSS has different mean and standard deviation in sentence matching items. The researcher has found that the mean of test 1 is 4.9, test 2 is

2.95 and the standard deviation of test 1 is 0.44, test 2 is 0.22 and the reliability of tests is 0.59 that means the reliability of sentences matching items is satisfactory.

Similarly, MPA has different mean and standard deviation in sentence matching items. The researcher has found that the mean of test 1 is 4.4, test 2 is 4.5 and the standard deviation of test 1 is 1.14, test 2 is 0.88 and the reliability of tests is 0.47 that means the reliability of sentences matching items is satisfactory.

Students of the selected schools were involved to obtain the information.

Among them, the students of KHSS have the highest reliability i.e. 1.01 in sentences matching item. It means the reliability of parallel –forms method is high.

3.4.2 True-False Item

To find out reliability of parallel form method in true-false item of the students, the test was administered to the students of the selected schools. The table is presented below.

Table No. 5
True-False Item

S.N	Name of school	No	Test 1		Test 2		r
			Mean	St.dev	Mean	St.dev	
1	PUA	20	4.9	0.30	4.95	0.22	0.86
2	KHSS	20	4.2	0.83	4.15	1.56	0.38
3	SAHSS	20	4.95	0.22	4.9	0.22	0.6
4	MPA	20	4.9	0.30	4.95	0.22	0.86

According to the table 4, PUA has different mean and standard deviation in true-false items. It has been found that the mean of test 1 is 4.9, test 2 is 4.95

and the standard deviation of test 1 is 0.30, test 2 is 0.22 and the reliability of tests is 0.86, which means the reliability of true-false items is high.

KHSS has different mean and standard deviation in true-false items. The researcher has found that the mean of test 1 is 4.2, test 2 is 4.15, and the standard deviation of test 1 is 0.83, test 2 is 1.56 and the reliability of tests is 0.38, which means the reliability of true-false item, is satisfactory.

SAHSS has different mean and standard deviation in true-false items. The researcher has found that the mean of test 1 is 4.95, test 2 is 4.9 and the standard deviation of test 1 is 0.22, test 2 is 0.22 and the reliability of tests is 0.6 that means the reliability of true-false items is satisfactory.

Similarly, MPA has different mean and standard deviation in true-false items. The researcher has found that the mean of test 1 is 4.9, test 2 is 4.95 and the standard deviation of test 1 is 0.30, test 2 is 0.22 and the reliability of tests is 0.8636 that means the reliability of true-false items is high.

Students of the selected schools were involved to obtain the information. Among them, the students of PUA and MPA have the highest reliability i.e. 0.86 in true-false item. It means the reliability of parallel –forms method is high.

3.4.3 Fill-in-the Blanks Items

To find out reliability of parallel forms method in true-false item of the students, the test was administered to the students of selected schools. The table is presented below.

Table No. 6
Fill in the Blanks Item

S.N	Name of school	No	Test 1		Test 2		r
			Mean	St. dev	Mean	St. dev	
1	PUA	20	4.7	0.47	4.5	0.51	0.60
2	KHSS	20	4.3	0.80	3.15	0.55	0.66
3	SAHSS	20	4.5	0.68	4.8	0.52	0.60
4	MPA	20	4.75	0.44	4.6	0.50	0.12

The above table shows that PUA has different mean and standard deviation in fill in the blanks items. The researcher has found that the mean of test 1 is 4.7, test 2 is 4.5 and the standard deviation of test 1 is 0.47, test 2 is 0.51 and the reliability of tests is 0.60, which means the reliability of fill in the blanks items is satisfactory.

KHSS has different mean and standard deviation in fill in the blanks items. The researcher has found that the mean of test 1 is 4.3, test 2 is 3.15, and the standard deviation of test 1 is 0.80, test 2 is 0.55 and the reliability of tests is 0.66, which means the reliability of fill in the blanks, is satisfactory.

Likewise, SAHSS has different mean and standard deviation in fill in the blanks items. The researcher has found that the mean of test 1 is 4.5, test 2 is 4.8 and the standard deviation of test 1 is 0.68, test 2 is 0.52 and the reliability of tests is 0.60 that means the reliability of fill in the blanks item is satisfactory.

Similarly, MPA has different mean and standard deviation in fill in the blanks items. The researcher has found that the mean of test 1 is 4.75, test 2 is 4.6 and the standard deviation of test 1 is 0.44, test 2 is 0.50 and the reliability of tests is 0.12 that means the reliability of fill in the blanks items is poor.

Students of the selected schools were involved to obtain the information. Among them, the students of PUA, KHSS, and SAHSS have moderate reliability in fill in the blanks item.

3.4.4 Multiple Choice Items

To find out reliability of parallel forms method in Multiple-choice item of the students, the test was administered to the students of selected schools. The table is presented below.

Table No.7
Multiple Choice Item

S.N	Name of school	No	Test 1		Test 2		r
			Mean	St. dev	Mean	St. dev	
1	PUA	20	4.2	0.61	4.55	0.51	0.53
2	KHSS	20	3.85	1.23	3.85	0.74	0.34
3	SAHSS	20	4.85	0.36	3.9	0.30	0.24
4	MPA	20	3.9	0.64	4.55	0.51	0.46

The above table shows that PUA has different mean and standard deviation in multiple-choice item. The researcher has found that the mean of test 1 is 4.2, test 2 is 4.55 and the standard deviation of test 1 is 0.61, test 2 is 0.51 and the reliability of tests is 0.34, which means the reliability of multiple-choice items is satisfactory.

KHSS has different mean and standard deviation in multiple-choice items. The researcher has found that the mean of test 1 is 3.8, test 2 is 3.85 and the standard deviation of test 1 is 1.23, test 2 is 0.74 and the reliability of test is 0.34, which means the reliability of multiple-choice item, is satisfactory.

Likewise, SAHSS has different mean and standard deviation in multiple-choice items. The researcher has found that the mean of test 1 is 4.85 test 2 is 3.9 and

the standard deviation of test 1 is 0.36, test 2 is 0.30 and the reliability of tests is 0.24 that means the reliability of multiple-choice items is poor.

Similarly, MPA has different mean and standard deviation in multiple-choice items. The researcher has found that the mean of test 1 is 3.9, test 2 is 4.55 and the standard deviation of test 1 is 0.64, test 2 is 0.51 and the reliability of tests is 0.46, which means the reliability of multiple-choice items is satisfactory.

Students of four selected schools were involved to obtain the information. From the information, we found that all the schools have moderate reliability.

3.4.5 Correct Order

To find out the reliability of parallel forms method in correct order item of the students, the test was administered to the students of selected schools. The table is presented below:

Table No. 8
Correct Order Item

S.N	Name of school	No	Test 1		Test 2		r
			Mean	St. dev	Mean	St. dev	
1	PUA	20	4.65	1.18	4.65	1.18	0.1
2	KHSS	20	3.5	1.70	3.8	1.79	0.54
3	SAHSS	20	4.2	1.00	4.95	0.22	1.62
4	MPA	20	4.55	1.23	4.35	1.03	0.38

Table 7 shows that PUA has different mean and standard deviation in correct order. The researcher has found that the mean of test 1 is 4.65, test 2 is 4.65 and the standard deviation of test 1 is 0.18, test 2 is 0.18 and the reliability of tests is 0.1, which means the reliability of correct order is poor.

KHSS has different mean and standard deviation in correct order. The researcher has found that the mean of test 1 is 3.5, test 2 is 3.8 and the standard

deviation of test 1 is 1.70, test 2 is 1.79 and the reliability of test is 0.54, which means the reliability of correct order, is satisfactory.

Likewise, SAHSS has different mean and standard deviation in correct order. The researcher has found that the mean of test 1 is 4.2 test 2 is 4.95 and the standard deviation of test 1 is 1.00, test 2 is 0.22 and the reliability of tests is 1.62 that means the reliability of correct order is high.

Similarly, MPA has different mean and standard deviation in correct order. The researcher has found that the mean of test 1 is 4.55, test 2 is 4.35 and the standard deviation of test 1 is 1.23, test 2 is 1.03 and the reliability of tests is 0.38, which means the reliability of correct order is satisfactory.

Students of four selected schools were involved to obtain the information.

Among them SAHSS has high reliability. KHSS and MPA have satisfactory and PUA has poor reliability.

3.4.6 Short Question –Answer

To find out reliability of parallel forms method in short question answers item of the students, the test was administered to the students of selected schools.

The table is presented below.

Table No.9

Short Question –Answer Item

S.N	Name of school	No	Test 1		Test 2		r
			Mean	St. dev	Mean	St. dev	
1	PUA	20	4.5	0.68	4.5	0.82	0.27
2	KHSS	20	2.25	1.51	1.35	0.81	0.26
3	SAHSS	20	4.05	0.82	3.7	0.68	0.49
4	MPA	20	4.05	0.88	3.7	1.26	0.18

According to the table 8, PUA has different mean and standard deviation in short question answer. The researcher has found that the mean of test 1 is 4.5, test 2 is 4.5 and the standard deviation of test 1 is 0.68, test 2 is 0.82 and the reliability of tests is 0.27, which means the reliability of short question answer is poor.

KHSS has different mean and standard deviation in short question answer. The researcher has found that the mean of test 1 is 2.25, test 2 is 1.35 and the standard deviation of test 1 is 1.51, test 2 is 0.81 and the reliability of test is 0.26, which means the reliability of short question answer, is poor.

Likewise, SAHSS has different mean and standard deviation in short question answer. The researcher has found that the mean of test 1 is 3.4 test 2 is 3.6 and the standard deviation of test 1 is 0.82, test 2 is 0.68 and the reliability of tests is 0.49 that means the reliability of short question answer is moderate.

Similarly, MPA has different mean and standard deviation in short question answer. The researcher has found that the mean of test 1 is 4.05, test 2 is 3.7 and the standard deviation of test 1 is 0.88, test 2 is 1.26 and the reliability of tests is 0.18, which means the reliability of short question answer is poor.

Students of four selected schools were involved to obtain the information. All the four schools have moderate reliability.

3.5 School-wise Description of Reliability

In this section, school-wise description of reliability with respect to private and government-aided schools is given. Four selected schools were involved to obtain this information. Among them, two were private and two were government-aided schools.

3.5.1 Private School

The following table presents the information of reliability of parallel form method of private school.

Table No. 10
Private School

S.N	Name of school	No	Test 1		Test 2		r
			Mean	St.dev	Mean	St.dev	
1	PUA	20	27.85	1.9808	26.95	2.2	0.77
2	MPA	20	26.45	2.52	26.6	2.66	0.14

From the above, table it is seen that there were twenty students in each school, who attended the tests. From the score of the students of PUA got, the mean of the test1 is 27.85, test 2 is 26.95, standard deviation of test 1 is 1.98, and test 2 is 2.2 and the reliability is 0.77, which means the reliability of the parallel forms method is highly correlated.

Similarly, MPA students got the mean of the test1 26.45, test 2 is 26.6, standard deviation of test 1 is 2.52, and test 2 is 2.66 and the reliability is 0.14 which means the reliability of parallel forms method is poor.

3.5.2 Government aided Schools

The students of selected two government schools were involved in the test to find out reliability in terms of parallel forms method. The table is presented below.

Table No. 11
Government School

S.N	Name of school	No	Test 1		Test 2		r
			Mean	St. dev	Mean	St. dev	
1	KHSS	20	20.2	4.76	17.9	3.86	0.75
2	SAHSS	20	26.8	1.43	25.25	1.06	0.30

From the above table it is found that there were twenty students in each school, who attended the tests. From the score KHSS got, the mean of the test1 is 20.2, test 2 is 17.9, standard deviation of test 1 is 4.76, and test 2 is 3.86 and the reliability is 0.75 which means the reliability of the parallel forms method is highly correlated.

Similarly, SAHSS students got the mean of the test1 26.8, test 2 is 25.25, standard deviation of test 1 is 1.43, and test 2 is 1.06 and the reliability is 0.30 which means the reliability of parallel forms method is poor.

CHAPTER FOUR

FINDINGS AND RECOMMENDATIONS

4.1 Findings

The purpose of the present study was to determine the reliability of test using parallel forms method in assessing reading skill of grade nine. In the process of research, the researcher selected the questionnaire based on the district level question. After the analysis and interpretation of the data obtained from the subjects, the researcher drew the following findings:

1. The reliability of parallel-forms method in assessing reading skill is high.
2. The reliability of the test item differs from school to school, item-to-item and even male and female students.
3. The reliability of the test was high in both private and government aided school.
4. In gender comparison, the reliability of the test administered to the male students was found to be higher than that of female students.
5. In gender-wise comparison, the reliability of test of male and female students of KHSS and male students of PUA and SAHSS was high. The reliability of female students of PUA and male students of MPA was poor. The reliability of test of male students of KHSS was the highest.
6. Gender issues were not addressed equally.
7. In item-wise comparison, the reliability of test in true –false item and fill in the blanks item was high whereas in matching item, multiple-choice item, rearrange and answer these questions was satisfactory. Out of them true-false item showed the highest and answer the questions showed the lowest reliability.

8. In personal comparison, only a few students got the same marks in both tests. Due to the marks variation among the students, it was concluded that the test item lacked the reliability.
9. It was found that the purpose of test was not clear.
10. Untrained teachers were found in making test items.

4.2 Recommendations

By the help of the performance of the students of the grade nine, the researcher has analyzed the findings and following recommendations have been made.

4.2.1 Recommendation for Pedagogical Implications

1. To find out the reliability of reading test, the parallel –forms method can be used.
2. Parallel –forms method can be used in schools to ensure the reliability of test item.
3. In gender - wise also, parallel –forms method should be used to find out the reliability assessing reading skill. The researcher's findings show that the female students need more practice in every exercise.
4. It can be concluded that gender issues should be equally addressed so that all boys and girls can do freely. It means test item should not be paradoxical and gender biased..
5. Parallel –forms method can be used in different test items to find out their reliability.
6. The instruction should be clear so that the students may understand the question, which helps to increase the reliability of test item.
7. It can be recommended that objective test item should be focused rather than subjective test item.
8. To check the answer sheet, answer key should be prepared for both subjective and objective test items so that biases can be reduced.
9. The designer be informed of the various techniques that can be used while designing tests.

10. Regarding the present study, a language test is a device that tries to assess how much has been learned in a foreign language course. So the teaching learning materials determines the testing devices.
11. After sampling the test items necessary changes should be made before administering the real test.
12. 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.
13. Special trainings and workshops are required for the teacher so that test item design can have uniformity. These are some recommendations...

REFERENCES

- Adhikari, S. B. (2007). *The Effectiveness of test re-tests method to measure reliability of the test item*. Unpublished M.Ed Thesis. Tribhuvan University. Kathamdu, Nepal.
- Batia L. & Nation, P., (1999). A vocabulary-size test of controlled productive ability. *Language testing*. Vol. 16. No. 1. 33-51.
- Cyril, J. W. (2006). Establishing test form and individual task comparability: a case study of a semi-direct speaking test. *Language Testing*. Vol. 23. No. 2. 167-197.
- Dahal, M. P. (2009). *A study on reliability in terms of split-half method*. Unpublished M.Ed Thesis. Tribhuvan University. Kathmandu, Nepal.
- Heaton, J. B. (1975). *Writing English language test*. CLBS: Longman.
- Hughes, A. (1989). *Testing for language teacher*. Cambridge: CUP
- Khanal, U. R. (1997). *A study on the effectiveness of the close test over conventional objectives test in testing reading comprehension in English*. Unpublished M.Ed. Thesis. Tribhuvan University. Kathmandu. Nepal.
- Khaniya, T. R. (2005). *Examination for enhanced learning*. Kathamdu: Millennium Publication.
- Kumar, R. (1996) *Research methodology*. London: SAGE Publication.
- Lado, R. (1961). *Language testing*. London: Longman.

Panday,B. (1999). *Teaching english: Theory of method*. Kathmandu: M.K. Publisher and Distributers.

Siwakoti, P. P. (1996). *An analysis of the reading proficiency*. Unpublished M.Ed. Thesis. Tribhuvan University. Kathmandu, Nepal.

Shrestha, I. (2007). *Reliability using test-retest method*. Unpublished M.Ed. Thesis. Tribhuvan University. Kathmandu, Nepal.

[http:// www.socialresearchmethods.com](http://www.socialresearchmethods.com). Retrieved 8 Feb. 2010.