

CHAPTER ONE

1. General Background

Language is the principal means used by human beings to communicate with one other. It is dynamic and open system that allows human to communicate their thoughts, feelings, desires, emotions, experiences and ideas. R.H. Robins (1964, 14) defines "languages are a symbol system . . . based on pure or arbitrary convention . . . infinitely extendable and modifiable according to the changing needs of the speakers." This definition says that language uses symbols and such symbols are arbitrary. Language is changed according to the need of the speakers. Language is primarily a pragmatic phenomenon, a symbolic instrument used for communicative purposes.

Language is maker or unmaker of human relationships. It is the use of language that makes a life bitter or sweet. Because of its omnipresence language is of taken for granted. According to Sapir (1978, 8) "Language is a purely human and non-instinctive method of communicating ideas, emotions and desires by means of a system of voluntarily produced symbols." Similarly, A.C. Gimson (1988, 4) defines language as "a system of conventional signals used for communication by a whole community. This pattern of conventions covers a system of significant sound units (the phonemes), the inflexion and arrangement of 'words' and the association of meaning with words." Language is a social phenomenon by means of which we establish the relationship in the society.

1.1 English Language and ELT in Nepal

There are a variety of languages which are used in the world. English is one of the many languages of the world, which is used as an international lingua-franca, an official language in many parts of the world or as the main means of international communication than any other language. It is taken as the richest language in the world because it has richest vocabulary in comparison to other languages. Most of the books are written in English medium and the most dominant language in almost all areas, e.g. mass media, trade, international diplomacy etc.

English has a significant influence in education system of our country. A good number of books, newspapers, magazines are found in English medium in Nepal. It has been taught as a foreign language in all schools in Nepal. In Nepal we mainly need English language for two main purposes:

- i. English as an international language, and
- ii. English is used for academic purpose.

The introduction and development of English education in Nepal is connected with the establishment of Durbar School. It was Jung Bahadur Rana, the first Rana Prime Minister of Nepal who initiated it after he returned from Britain in 1910 B.S. At the very beginning it was confined only to their families. In 1942 B.S., Durbar School was opened for the common people. After the establishment of Tri-Chandra College in 1975 B.S., English became a compulsory subject in higher education. National Educational System Plan (NESP) 2028 introduced English from grade four. Now it is compulsorily taught and learnt from grade one upto Bachelor level. English is used as a medium of instruction in most of the private schools and campus of Nepal.

1.2 Language Teaching and Testing

The origin of modern language testing is not so distant. Before the Second World War, the idea of language testing as a distinct activity scarcely existed. A number of factors contributed to the development of interest in systematic scientific language testing. Basically war time language programmes in United States and else where and the growth of international agencies gave new importance to language teaching projects.

Language teaching and language testing are the two sides of a coin- Language testing always remains incomplete without testing. Whether we claim language teaching governs testing or testing governs language teaching both arguments have their own proper ground for claim. In fact, both go side by side. Regarding the relationship between testing and language teaching Heaton (1985:5) says "both testing and teaching are so closely interrelated that it is virtually impossible to work in either field 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." It is said that language testing is an obedient servant of language teaching. Teaching always guides testing and vice versa. In reality without language teaching there will be no existence of language testing and without testing, there will be no meaning of language teaching. What to teach and what to test are same that is we test what we teach. There is a two way traffic relationship between teaching and testing.

1.3 Qualities of a Good Test

To find out the knowledge and skills of students there are a lot of terms in use such as testing, assessment, evaluation and examination. Assessment is a general term that includes the full range of procedures used to gain information about students learning such as observation, rating of performance or projects and the formation of value judgements concerning learning progress. A test is particular type of assessment that typically consists of a set of questions administered during a fixed period of time under reasonably comparable conditions for all students. Measurement is the assigning of numbers to the result of a test or other type of assessment according to a specific rule. Examination is quite different from the rest in the sense that it is extremely formal and usually conducted at the end of an academic year a twice or year. It's purpose is to grade the students.

There are certain criterias which are to be considered while constructing test. These criterias are known as qualities of a test. A good test should have the following qualities:

1. **Validity:** The validity of a test is the extent to which it measures what it is supposed to measure and nothing else. "The validity of a test is the extend to which the test measures what it is intended to measure." (Harrison 1983: 12)
2. **Reliability:** The reliability of a test is it's consistency. It is a prerequisite for a valid test. That is to say if a test is not reliable then no talk of it's validity. In other words, any test to be valid it should be reliable. The reliability of a test is its consistency in score.

3. **Practicability:** A test must be practicable. It must be fairly straightforward to administer. It refers to the degree to which a test is adaptable in varying situation. It should be fit for the situation so that it can be easily administered. Before designing the test, the test designer should consider the following three things. They are:
 - a) Human resources,
 - b) Material resources and
 - c) Time.
4. **Scorability:** Scorability refers to the proper allocation of marks regarding calculation as well. For example, if there are 4 sub questions in one item bearing 3 marks then it is very difficult to score.
5. **Economy:** Any test to be a good one, it should be economical in all senses. That is to say any good test should be as economical as possible in time and money without desolving the other qualities of a good test. It refers to the construction, administration and scoring of test in minimum cost and labour.
6. **Washback:** Washback is one of the qualities of a good test. In language testing literature the term backwash and washback are being used interchangeably. Hughes (1995: 1) writes "the effect of testing on teaching and learning is known as backwash."

1.4 Subjective Tests and Objective Tests

Tests can be classified variously on the basis of various aspects. On the basis of scoring tests can be classified into two types. They are subjective tests and objective tests.

It is not the test which are subjectives or objectives in themselves, but the process involved in marking an answer determines a test as subjective and objective (Harrison, 1983: 10).

These terms are used to designate two types of scoring. Objective tests are those that are scored rather mechanically without need to evaluate complex performance on a scale. Subjective tests are those that require opinion, judgement on the part of the examiner. In subjective tests there are differences in scoring by different examiners, hence the name subjective. Subjective tests are also called essay type of test. Seven general characteristics that are used as a basis for differentiating between two types of test items are given below:

1. Difficulty of preparing the test item.
2. Adequacy of the sampling of the subject matter.
3. Relative ease with which knowledges and understandings are measured.
4. Study procedures followed by the pupil as he prepares for the test.
5. Originality of the response the pupil must make to the test item.
6. Relative success of guessing correct responses.
7. Difficulty of scoring the pupils responses.

Comparison of essay and objective tests

Characteristics	Essay tests	Objective tests
– Preparation of the test item	Items are relatively easy to construct.	Items are relatively difficult to construct.
– Sampling of the subject matter	Sampling is often limited.	Sampling is usually extensive.
– Measurement of knowledge and understanding	Item can measure both; measurement of understanding is recommended	Items can measure both, measurement of knowledge is emphasized

– Preparation by pupil	Emphasis is primarily on larger units of material.	Emphasise is primarily on factual details
– Nature of response by pupil	Pupil organizes original response	Except for supply test items, pupil select response
– Guessing of correct response by pupil	Successful guessing is minor problem	Successful guessing is major problem
– Scoring of pupil responses	Scoring is difficult, time consuming, and somewhat unreliable	Scoring is simple, rapid and highly reliable

Adapted from "Evaluating Pupil Growth" by J. Stanley Ahmann and Marvin D. Clock 1958.

The usual objections to objective tests are that they are too simple, that they do not require real thinking but simply memory, and that they do not test the ability of the student to organize his thoughts. "When objectives tests of language are properly made, they have two important values, they can test in a short time the entire range of the sound system, or the major grammatical patterns or a representative sample of the vocabulary taught during a whole year or several years. Secondary, they can be scored with speed and ease."

1.5 Types of Objective Test

The principal types of objective test items are listed as follow:

- 1) Recall types
 - a. Simple recall
 - b. Completion
- 2) Recognition types
 - a. More common
 - i. Alternative response
 - ii. Multiple choice

- iii. Matching
- b. Less common
 - i. Rearrangement
 - ii. Identification
 - iii. Analogy
 - iv. Incorrect statement

But the most widely used objective test items by classroom teachers are given below:

- 1) Simple recall tests: It is defined as one in which each item appears as a direct question, a stimulus, word or phrase and specific direction. The response must be recalled by the pupil from his past experience rather than merely identified from a list of suggested answer supplied by the teacher.
- 2) Completion: This test is defined as a series of sentences in which certain important words or phrases have been omitted and blanks submitted for the pupil to fill in. A sentence may contain a single blank or it may contain two or more blanks.
- 3) Alternative response tests: An alternative response test is made up of items each of which admits of only two possible responses. The usual form of alternative response test is true false test item, other forms are same /opposite and yes/no.
- 4) Multiple choice tests: A multiple choice test is made up of items each of which presents three or more responses, only of which is correct or definitely better than the others.

- 5) Matching tests: A matching test typically consists of two columns, each item in first column to be paired with word or phrase in the second column upon some basic suggested.

Some Guidelines to Construct Multiple Choice Items

The multiple choice item is generally recognized as the most widely applicable and useful type of objective test item. It can measure a variety of complex outcomes in the knowledge, understanding and application areas. It consists of a problem and a list of suggested solutions. The students are typically requested to read the stem and the list of alternatives to select the one correct, or best alternatives. The correct alternative in each item is called the answer, or key and the remaining alternatives are called distracters also called decoys or foils. The following steps help to construct a good multiple choice items.

1. The stem of the item should be meaningful by itself and should present a definite problem.
2. The item stem should include as much of the item as possible and should be free of irrelevant materials.
3. Use a negatively stated stem only when significant learning outcomes require it.
4. All of the alternative should be grammatically consistent with the stem of the item.
5. An item should contain only one correct or clearly best answer.
6. All distractors should be plausible.
7. The relative length of the alternatives should not provide a clue to the answer.
8. Only one feature at a time should be tested.

1.6 Marking subjective and objective tests

A language test has two major aspects; setting and marking. The first refers to the construction of the instruments designed to elicit responses and the second refers to the evaluation of the responses. The testing consists of pre-administrative and post-administrative activities. Marking is assigning marks to an answerbook according to certain fixed and explicit rules. Marks in themselves, which are simply symbols, have not any intrinsic value but can be used to roughly indicate the level of achievement of a pupil.

There are different approaches to marking for subjective and objective test. Mainly, three approaches are used for marking subjective test. They are:

- 1) Analytical marking, 2) Impressionistic marking and
- 3) Multiple marking

Harrison (1983: 111) mentions the following three ways in marking subjective test.

- a) Rank marking: In this approach the researcher puts all the answer sheets in rank order, from best to worst, by comparing one with another and sorting them, rather like a hand in a game of cards.
- b) Category marking: Another system, the most frequently used, is to assess in categories such as vocabulary, grammar, content and form, allotting, say, five marks to each and awarding a total mark out of twenty.
- c) Mechanical marking: A purely mechanical system of marking continues written work, such as an essay, is to count off sections of, say, eight words, without taking account of sense, and see what

can be given credit: sequence of correct words, vocabulary, verb terms, idioms, phrasal verbs and so on.

In objective test item there is only one distinctly correct or more suitable answer than others. Therefore, an objective test can be quickly and accurately marked by using answer key.

There may or may not be the system of penalty for wrong guessing. Where there is no system of penalty, full mark is given for each correct answer. Where there is a system of giving penalty for guessing, the following formulas are used.

1) For True/False

Total mark = Number of correct answer - number of incorrect answer

2) For multiple choice items

Total marks = Number of correct answer - $\frac{\text{Number of incorrect answer}}{\text{Number of alternative} - 1}$

1.7 What is Item Analysis?

A language test has two major aspects: setting and marking. The first refers to the construction of instruments designed to elicit responses and the second refers to the evaluation of these responses. The testing consists of pre-administrative and post-administrative activity of the test.

Most of the teachers think that the test is finished once the raw marks have been obtained. But the results obtained from subjective test can be used to provide some other valuable information as well. "Even individual items make their own contribution to the total test some contribution more than others and it is the purpose of item analysis to identify those that need to be changed or replaced" (Hughes, 1995: 160).

The result obtained from objective tests can be used to provide valuable information concerning;

- The performance of the students as a group.
- The performance of individual student, and
- The performance of each of the items comprising the test.

"Information concerning the performance of the students as a whole and of individual students is very important for teaching purposes, especially as many test result can show not only the types of error most frequently made but also the actual reasons for the errors being made" (Heaton, 1988: 178).

"A language item is a sample of the performance of students on language problems. Item analysis is the study of validity, reliability and difficulty of test items taken individually as if they were separate tests" (Lado, 1961).

According to Richard et al. (1999: 192) "Item analysis is the analysis of the responses to the items in a test, in order to find out how effective the test items are and to find out if they indicate differences between good and weak students."

"Even individual items make their own contribution to the total test. Some contribute more than others, and it is the purpose of item analysis to identify those that need to be changed or replaced (Hughes, 1995: 160).

Re-examining each item of a test for the purpose of discovering its strength and flaws is known as item analysis. Item analysis customarily concentrates on two vital features of each test item: Its level of difficulty

and its discriminating power. By the former is meant the percentage of pupils who answer correctly each test item, by the latter is meant the ability of the test item to differentiate between pupils who have achieved well and those who have achieved poorly. Item analysis methods are essentially mathematical in nature and can take many forms.

The performance of the test items themselves, is of obvious importance in compiling future tests. Since a great deal of time and effort are usually spent on the construction of good objective test items. It is thus useful to identify those items which were answered correctly by more able students taking the test and badly by the less able students.

Item analysis helps to recognize the difficulty level of an item. It determines whether the same item can be reused or not. Item analysis answers the following questions.

- Can the questions fulfill the expected goal or not?
- Whether the difficulty level is appropriate or not?
- Whether the questions are errorless or not?
- Whether the distracters of multiple questions are effective or not?

These questions are important for the selection and improvement of the test items. The important of item analysis is not limited to the improvement of the items. It has the following advantages:

- i. It provides a basis for efficient class discussion of the test result.
- ii. It provides basis for remedial work.
- iii. Item analysis leads to increase skill in test construction.

"Item analysis results can serve two major purposes. The first and more obvious purpose is that they provide important information

concerning the problems encountered when informal achievement test are built. On the basis of the information obtained the teacher can gain a much better view of the worth of the test he built and used, he can also profit by his mistakes in that he should be able to construct noticeably better test in future. The second use is to diagnosis" (Ahmann, 1958).

Item analysis mainly includes three main aspects of objective questions. They are:

1. Facility value of an item, which simply show how difficult and easy the particular item proves in the test.
2. Item discrimination means the extent to which the item discriminates between the testees, separating the more able testees from the less able.
3. Distractor analysis: Distractors are the incorrect options in each multiple choice item. In item analysis we can find out which distractors attract most of the students and which distractors attract none of the students.

"The main information to be obtained about individual items (both multiple-choice and True/False) is how difficult they are and how will they sortout the better students from the poorer ones" (Harrison, 1983,127).

"Facility value is" a measure of the ease of a test item. It is the proportion of the students who answered the time correctly, and may be determined by the formula:

$$\text{Item facility (IF)} = \frac{R}{N}$$

Where, R = Number of correct answer

N = Number of students taking the test" Richard et. al. (1999.182).

The facility value of a large number of individual items will vary considerably. An average facility value of 5 or 50 percent may be desirable. For example, If 21 out of 40 students obtained the correct answer for one of the items that item would have an index of difficulty of 0.52 percent. This item is at right level of difficulty. This index can be useful when deciding the order of items in a test. It is generally desirable to start the test with easy items.

"The discrimination index of an item indicates the extent to which the item discriminates between the testes separating the more able testes from the less able. The index of discrimination (D) tells us whether those students who performed well on the whole test tended to do well or bad on each item in the test" (Heaton, 1988.179).

There are various methods of obtaining the index of discrimination. The following method has been used to carry out this work.

1. Arrange the scripts in rank order according to raw scores and the top and bottom third of raw score are found, the middle third is ignored.
2. Count the number of those candidates in the upper group answering the first item correctly, then count the number of lower group candidates answering the item correctly.
3. Subtract the number of correct answers in the lower group from the number of correct answer in the upper group.
4. Divide this difference by the total number of candidates in one group.

$$D = \frac{\text{Correct U} - \text{Correct L}}{n}$$

D = Discrimination, n=Number of candidates in one group

U = Upper half and L=Lower half

Thus, D is the difference between the proportion passing the item in U and L.

5. Proceed in this manner for each item.

Where multiple choice items are used, it is necessary to analyse the performance of distractors. Distractors which do not work, i.e. are chosen by very few candidates, make no contribution to test reliability. Such distractors have to be replaced by better ones, or the item has to be replaced by better ones, or the item has to be otherwise modified or dropped. We can simply count how many students choose each alternative.

Another aspect of item analysis is item-test correlations. The assumption is that the people who do best on the whole test should do best on any particular item. Item-test correlation checked out the correlation between total scores and scores on a particular item. But it requires computer. Now a day, item analysis has become a sophisticated field. In recent years new methods of analysis have been developed. Hughes mentions all these above terms under the heading of item response theory, and the form of it so far most used in language testing is called "Rassch analysis." But efficient use of these analysis powerful microcomputers are necessary.

1.8 Literature Review

A number of researchers have carried out several studies in the field of language testing. Most of the studies deal with the various aspects of language testing such as, 'A study on marking an English answer book,' "Effectiveness of discrete point test as a measure of English," "washback effect of examinations." So far as the research works carried out the item analysis of objective marking is concerned they are very limited.

Sharma (1999) carried out a study on "A study in marking an English answer book." The researcher has attempted to find out the nature and degree of variation in marking on SLC Compulsory English answer book in terms of different variables. The researcher concluded that variation in marking is found to occur not only across but also within group of various kinds. Similarly more variation in marking is found to occur in subjective test items than in objective one.

Dhaka (2000) conducted a research on "Effectiveness of Discrete point test and integrative test as measure of English language proficiency: A comparative study." The finding was that the students performed better on textbook material test than on non-textbook test on discrete point test and the result of integrated test was also as it was resulted in discrete point test. But it did not show which of the test was better.

Sah (2001) has carried out research on "Effectiveness of objective and subjective type tests for Grade Ten." The finding was that the performance of the students was better in objective test than in subjective test. The performances of the students in both subjective and objective test were better in the test book material than non-textbook material.

Gautam (2001) conducted a research on "An Analysis of subjective marking." His study is concerned with the variation in marking subjective answers by different examiners. The finding shows that there occur variations in marking in subjective test. More over the research shows that the higher the designation (academic position) of the examiners, the more strict they are in marking. Similarly, female examines in general are more lenient than the male examiner.

Neupane (2004) carried out research on "Washback effect of examination: A case of Communicative English." The finding was that the examination was not seemed to have helped to promote the communicative abilities of the students and students' participation in the classroom in was very low. Similarly, Teaching was teacher centered.

Baral (2004) carried out research on "Item analysis of multiple choice objective test: A practical study." The research concluded that items with good facility value need not necessarily have good discrimination index.

Ojha (2005) attempted to find out the "Content Validity of ELT Theories and Method Exam at B.Ed. level." The research finding was that tests have high content validity in terms of coverage principle but tests have low content validity in terms of weighting principle.

Although the studies mentioned above are related to language testing and item analysis, no research has been done on the item analysis of B.Ed. 2nd year's course "ELT Theories and methods." Thus, this study differs from the rest of studies carried out in the department till present date and the researcher hopes that this research with or a fresh research in the department.

1.9 Objectives of the Study

The general objective of the proposed study was to carry out the item analysis of objective test item of B.Ed. 2nd year "ELT Theories and methods" in the year from 2055 to 2059.

The specific objectives of the study are:

- To find out the index of difficulty,
- To find out the index of discrimination,
- To find out the power of distracters of each objective questions,
- To find out the repetition of objective questions,
- To suggest some pedagogical implication,

1.10 Significance of the Study

As stated in the objectives, the study attempts to weigh-up the statistical efficiency of multiple-choice objective test. So it will be important to the language teachers, test designer and the students. This study is significant because it throws light on how to make objective test more effective.

1.11 Definition of the Specific Terms

Facility Value: The proportion of students responding correctly to an item gives its facility value. There can be no strict rule about what range of facility value are to be regarded as satisfactory, easy or difficult. It depends on what the purpose of the test is. But the researcher has used the following criteria to carry out the work.

Easy Item: An item with a facility value above 0.7.

Average Item: An item with a facility value falling between 0.3 to 0.7.

Difficult Item: An test item with facility value below 0.3.

Discrimination index: An extent to which an item discriminates between the testees, separating more able testees from less able. It measures how performance on one test item correlates to the performance in the test as a whole. It ranges from +1 through -1.

- Perfect discriminator: A case when 100% testees at the upper level correct a particular item and none of the testees at lower level correct the same item.
- Negative discriminator: An item, which discriminates in the wrong direction.
- Moderate discriminator: An item with discrimination index falling between 0.3 to 0.7.
- Poor discrimination: An item with discrimination index falling below 0.3. 0 is also taken as poor.
- Stem: The initial part of each multiple-choice item.
- Key: The correct option in each multiple-choice item.
- Distracters: Incorrect options in each multiple choice items.
- Options: The choices from which students select their answer.
- Omit: If the students have not answered an earlier item, the unanswered item is called omit. Any items following the last item attempted are designed "not reached."
- Non-functioning: A distractor, which attract none of the testees.

CHAPTER TWO

2. Methodology

This chapter deals with methodology adopted during this research work. The details of the methodology are as follows.

2.1 Sources of Data

In the preparation of this dissertation both primary and secondary sources of data have been used: primary source was used for data collection and secondary source was used to facilitate the study.

2.1.1 Primary Source

The primary sources of data were the B.Ed. 2nd year students. However, the sample population consisted of 80 students from different campuses of Kathmandu valley.

2.1.2 Secondary Source

The researcher studied the related books, journals, reports and theses to facilitate the study.

2.2 Population of the Study

The population of the study consisted of B.Ed. 2nd year students studying at different campuses of Kathmandu valley.

2.3 Sample Population

The sample population of study consisted of 80 students of B.Ed. 2nd years. The students were selected from four different campuses of Kathmandu. The selected campuses for the population are given below.

1. Mahendra Ratna Campus, Tahachal,
2. Sanothimi Campus, Bhaktapur,
3. Kathmandu Shiksha Campus,
4. Gramin Adarsha College, Nepaltar. The researcher selected 20 students from the each campus.

2.4 Tools for Data Collection

The major tools for the collection of data were the test items. Five sets of items consisting of altogether 100 multiple choice objective test items asked in the examination of B.Ed 2nd year in the year from 2055 to 2059 B.S. of the course "ELT Theories and Methods" were used to elicit data from the students. The test items were answered by ticking /√/ the correct answer.

2.5 Process of Data Collection

First of all four campuses were selected from the Kathmandu, two representing private campuses and two representing the government campuses. The researcher then visited the selected campuses and with the cooperation of the campus authorities, selected 20 students from each campus. Then he administered the test on the students systematically following a step wise procedure as follows.

- The students were arranged in seats maintaining a considerable distance between adjacent ones.
- He explained to the students his purpose of taking the test. He told them that the test was not to make them pass or fail. Its purpose is just to analyze the facility value and index of discrimination of the test item.

- He gave necessary instructions clearly. After making it sure that all of them know that what to do, he distributes the question paper.
- Their response is recorded just by ticking.
- He collected the test papers from all of them and thanked them for their participation.

2.6 Limitation of the Study

- The study was limited to carry out item analysis of multiple choice objective tests asked in the exam B.Ed. 2nd years course "ELT Theories and Methods" in the year from 2055 to 2059.
- The study was limited to 80 students.

CHAPTER THREE

ANALYSIS AND INTERPRETATION OF DATA

In this section, the marks of the students were analyzed. Firstly, the marks secured by them were tabulated. On the basis of the table scores, the analysis and interpretation were carried out as effectively and accurately as possible.

In this section, analyses of multiple-choice items were done in terms of:

- The index of difficulty
- The index of discrimination
- The power of distracters
- The repetition of the item within the years from 2055 to 2059

While examining the difficulty of the test, test items were divided into three categories.

- 1) Easy item: Item with FV of above 0.7
- 2) Average item: Item with a FV falling between 0.3 to 0.7
- 3) Difficult items: Item with a FV below 0.3

Similarly, while examining the discrimination index of the items, the items are divided into five categories:

- 1) Perfect discriminators: Items having a discrimination of index of +1.
- 2) Good discriminators: Items having a discrimination index above 0.7.

- 3) Moderate discriminators: Items having a discrimination index falling between 0.3 to 0.7.
- 4) Poor discriminators: Items with a discrimination index of below 0.3.
- 5) Negative discriminators: Items which discrimination the wrong direction to -1.

Table 1: Compiled form of the extended "Item analysis" of test retest result of the test items asked in the year 2055

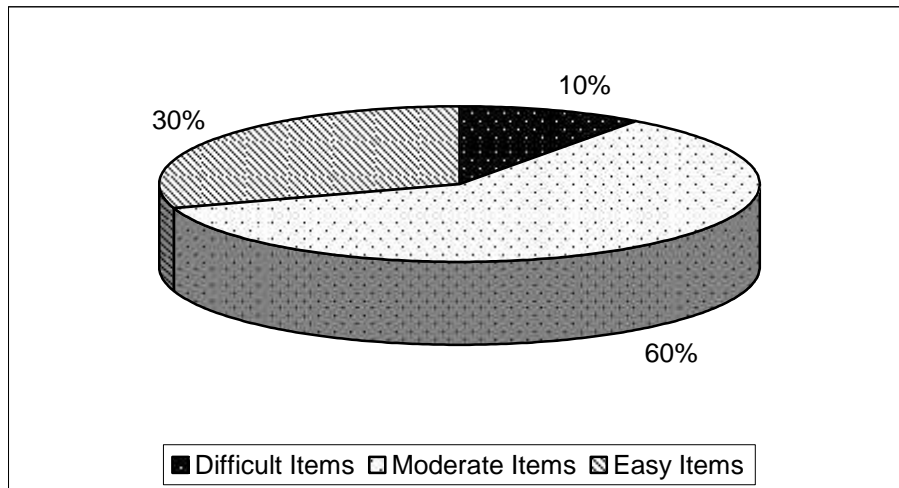
Item Number	Difficulty index, (FV)	Discrimination index	Power of distracters				No response	Repeated items
			A	B	C	D		
1.	0.86	0.26	69	4	3	3	1	2056
2.	0.62	0.26	6	7	50	13	4	
3.	0.68	0.40	3	10	55	12		
4.	0.62	0.53	10	50	4	15	1	2056
5.	0.77	0.13	62	5	7	5	1	
6.	0.73	0.26	59	6	8	7		
7.	0.53	0.26	43	10	12	10	5	
8.	0.65	0.46	8	12	6	52	2	2058 2056
9.	0.47	0.33	38	12	15	10	5	
10.	0.81	0.33	5	65	10	0		2056 2058 2059
11.	0.52	0	9	42	9	20		
12.	0.38	0.6	18	22	31	9		
13.	0.67	0.46	54	12	7	7		
14.	0.63	0.20	3	9	15	51	2	2059
15.	0.83	0.33	67	6	0	6	1	2059
16.	0.68	0.40	7	10	7	55	1	
17.	0.76	0.33	14	61	0	5		
18.	0.53	0.66	11	7	42	19	1	
19.	0.36	0.13	29	17	8	24	2	
20.	0.53	0.06	2	3	43	30		2059

Difficulty Index (FV)

Difficult Items = 10%

Moderate Items = 60%

Easy Items = 30%



Discrimination Index (DI)

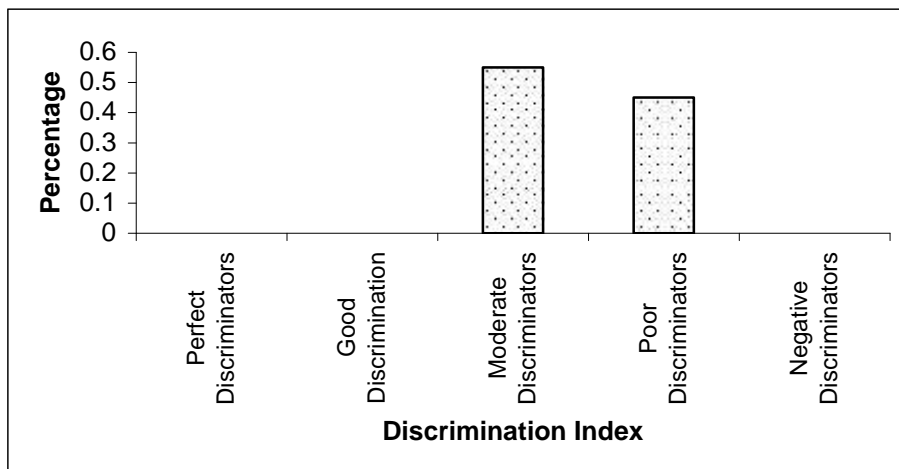
Perfect Discriminators = 0

Good Discriminators = 0

Moderate Discriminators = 55%

Poor Discriminators = 45% (0 is also counted as poor)

Negative Discriminators = 0



The table and diagram in the preceding pages show the recorded measures of the test, viz. difficulty index, discrimination index, power of the distracters, items not responded and test items repeated within the years from 2055 to 2059 B.S.

Facility Value

To examine the test in terms of difficulty index, 10% items appeared to have been easy, 60% items (12 items out of 20) seemed to have right level of difficulty and 30% items (6 out of twenty) were difficult.

Discrimination Index

While examine the test in terms of discrimination index, none of the items were seemed to have a perfect and even a good discrimination index, while 55% items were moderate discriminators and 45% items were found to be poor discriminators. None of the items were found to discriminate in the wrong direction.

Power of the Distracters

To analyze the distracters, distracter 'D' of item no. 10, 'C' of item no. 15 and 'C' of item no. 17 were found non functioning distracters. Distracters 'A' of item no. 3, 14 and 20 and distracters 'B' of item no. 5, and 20 appeared to be very poor.

Repetition of the Test Items

While examining the repetition of the test items, item no. 1, 4, 8 and 10 have been adopted in the year from 2056 and item no. 10 has been adopted form the year 2058 and item no. 10, 14, 15 and 20 have been

adopted from the year 2059. Here, item no. 10 has been repeated three times and 35% items were repeated item in a test item.

Table 2: Compiled form of the extended "Item analysis" of test retest result of the test items asked in the year 2056

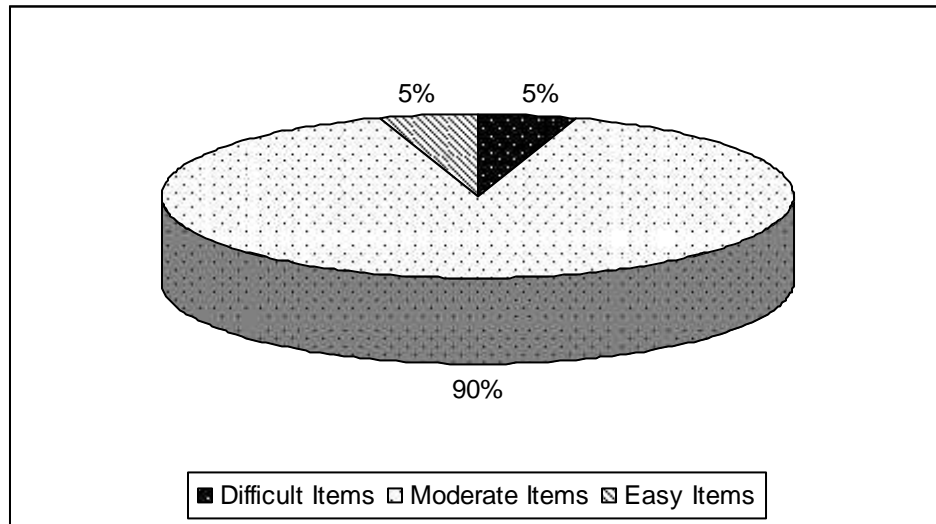
S.N.	Facility Value	Discrimination index	Power of distracters				No response	Repeated Items 2055-2059
			A	B	C	D		
1.	0.60	0.53	19	10	3	49		2059
2.	0.68	0.40	3	11	55	11		2058
3.	0.48	0.40	39	13	15	13		2059
4.	0.62	0.53	13	50	11	6		2056,2058,2059
5.	0.63	0.60	51	7	19	0	3	2058
6.	0.40	0.33	21	8	19	32		2059
7.	0.53	0.46	12	21	43	3	1	2059
8.	0.40	0.66	32	18	15	12	3	
9.	0.45	0.26	19	19	36	6		
10.	0.25	0.06	15	20	34	9	2	
11.	0.47	0.66	17	38	10	15		2055
12.	0.57	0.53	19	9	6	46		
13.	0.57	0.40	19	2	46	13		2055
14.	0.65	0.40	8	12	6	52		
15.	0.63	0.13	3	6	19	51	1	2058
16.	0.52	0.46	15	42	12	11		2056, 2058
17.	0.61	-0.13	22	49	6	3		
18.	0.75	0.33	8	6	60	6		2059
19.	0.37	0.33	14	30	8	28		
20.	0.62	0.20	50	9	6	15		

Difficulty Index (DI)

Difficult Items = 5%

Moderate Items = 90%

Easy Items = 5%



Discrimination Index (DI)

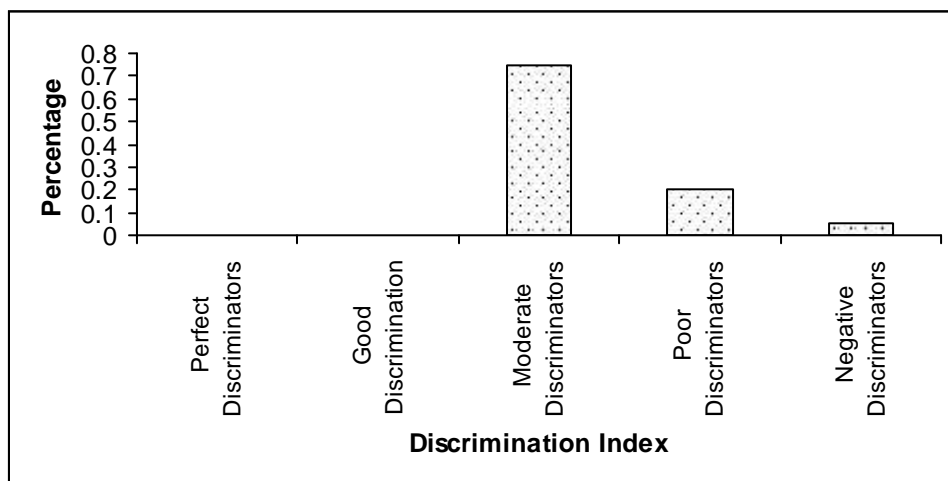
Perfect Discriminators =

Good Discriminators =

Moderate Discriminators = 75%

Poor Discriminators = 20%

Negative Discriminators = 5



The table and diagram in the preceding pages show the recorded measures of the test, viz. difficulty index, discrimination index, power of the distracters, items not responded and test items repeated within the years from 2055 to 2059 B.S.

Facility Value

To examine the test in terms of facility value, 5% items (one out of twenty) appeared easy item, 90% items (eighteen out of twenty) were seemed to have been moderate or right level of difficulty and 5% items in a test were found difficult.

Discrimination Index

To examine the test in terms of discrimination index, none of the items in a test were found to have a perfect and even a good discrimination index. 75% items in the test appeared to be moderate discriminators, 20% items were poor discriminators and 5% items were negative discriminators. Item no. 17, having facility value of 0.61 has failed to discriminate between the testees for it has negative discrimination index of -0.13.

Power of the Distracters

To analyze the distracters, distracter 'D' of item no. 5 seemed to be non-functioning. Similarly distracter 'D' of item no. 7, 17 and distracter 'A' of item no. 2 and 15 and distracter of item no. 13 were also non-functioning.

Repetition of the Test Item

Most of the test items asked in the year 2056 were adapted from the year 2058 and 2059. 55% items have been repeated. Item no. 1, 3, 4, 6, 7 and 18 have been adapted from the year 2059. Similarly, item no. 2,

5, 15 and 16 have been adapted from the year 2058 and item no. 11 and 13 have been adapted from the year 2055.

Table 3: Compiled form of the extended "Item analysis" of test retest result of the test items asked in the year 2057

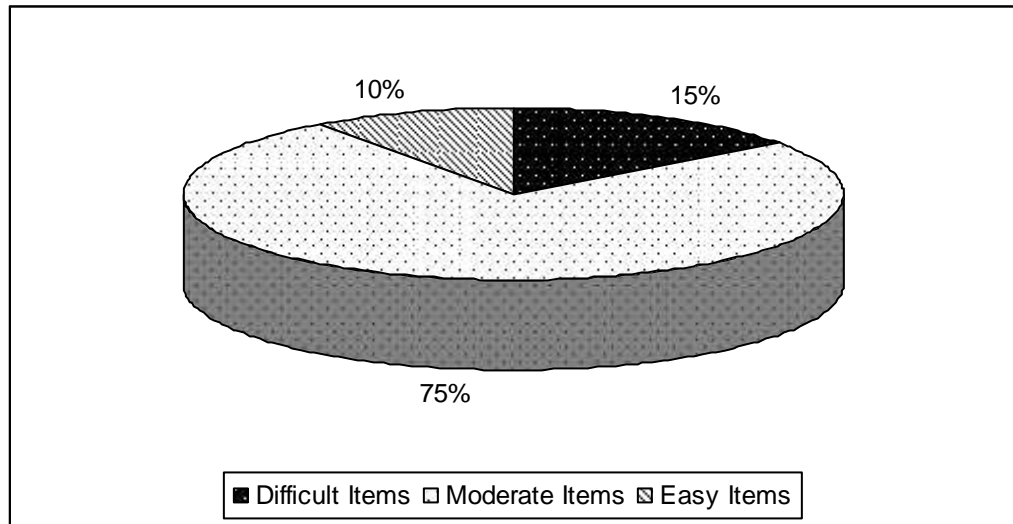
S.N.	Difficulty Index	Discrimination index	Power of distracters				No response	Repeated Items
			A	B	C	D		
1.	0.66	0.40	6	15	6	53		
2.	0.56	0.33	18	45	5	11	1	
3.	0.53	0.27	9	6	43	22		
4.	0.42	0.40	24	34	8	13	1	2059
5.	0.48	0.66	8	23	11	38		2058
6.	0.52	0.20	9	9	17	42	3	2059
7.	0.23	0.20	17	31	13	19		2059
8.	0.28	0.33	36	12	23	8	1	
9.	0.51	0.33	41	8	24	7		
10.	0.65	0.53	5	10	52	13		
11.	0.55	-0.06	30	3	44	3		2059
12.	0.30	0.13	19	19	18	24		2059
13.	0.75	0.26	17	0	3	60		
14.	0.41	0.20	0	28	17	33	2	2058
15.	0.48	0.20	38	23	11	8		2058
16.	0.61	0.26	12	5	49	14		
17.	0.26	0.13	8	21	25	21	5	
18.	0.51	0.46	23	0	16	41		2058
19.	0.66	0.60	11	11	53	5		2055,2058,2059
20.	0.76	0.40	6	6	61	7		

Difficult Index (DI)

Difficulty Items = 15%

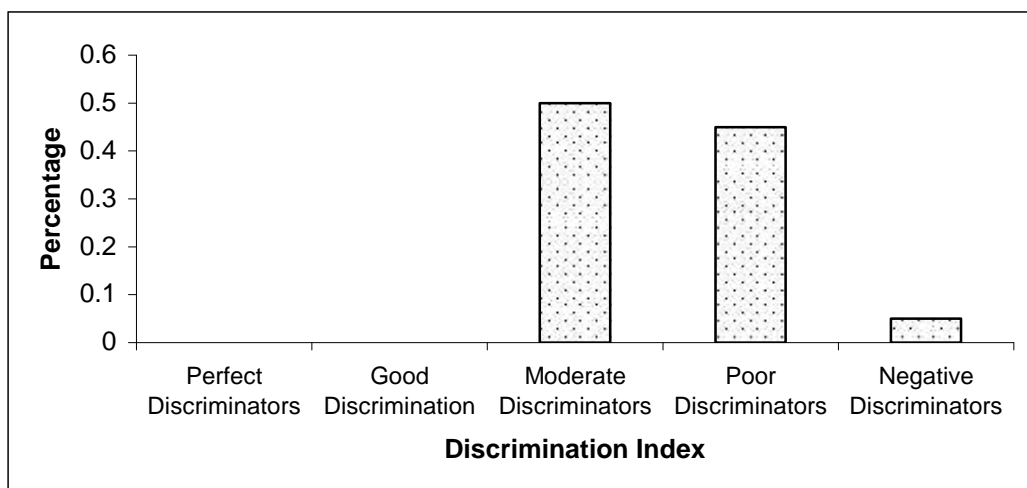
Moderate Items = 75%

Easy Items = 10%



Discrimination Index

1. Perfect Discriminators = 0
2. Good Discriminators = 0
3. Moderate Discriminators = 50%
4. Poor Discriminators = 45%
5. Negative Discriminators = 5%



The table and diagram in the preceding pages show the recorded measures of the test, viz. difficulty index, discrimination index, power of the distracters, items not responded and test items repeated within the years from 2055 to 2059 B.S.

Facility Value

To examine the facility value of the test items, 15% items were difficult and 75% items (fifteen out of twenty) were moderate and 10% items (two out of twenty) were easy items.

Discrimination Index

While examine the discrimination index, none of the items were appeared to have been perfect discriminators, 50% items (ten out of twenty) appeared to have been moderate discriminators and 5% items appeared to have been negative discriminators.

Power of the Distracters

If we analyze the power of the distracters, distracter 'A' in item no. 14, distracter 'B' in item no. 13 and distracter 'd' in item no. 11 were seen to be non-functioning distracters. Other distracters seemed functioning satisfactorily.

Repetition of the Items

To examine the repetition of the test items, 50% items (ten out of twenty) were adapted from the year 2055 to 2059. Item no. 5, 14, 15, 18 and 19 were adapted from the year 2058 and item no. 4, 6, 7, 11, 12, and 19 were adapted from the year 2059.

Table 4: Compiled form of the extended "Item analysis" of test retest result of the test items asked in the year 2058

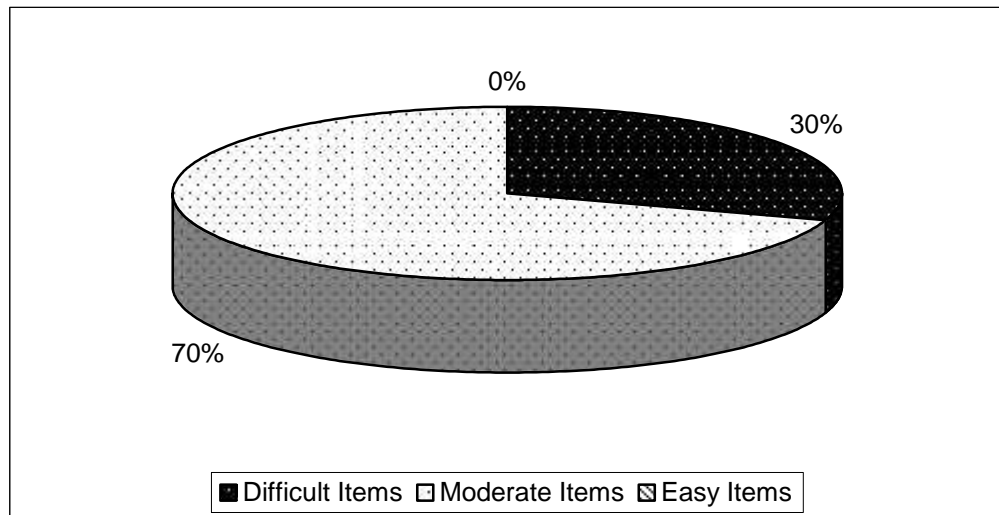
S.N.	Difficulty Index	Discrimination index	Power of distracters				No response	Repeated Items
			A	B	C	D		
1.	0.75	0.13	7	60	6	7		
2.	0.51	0.26	23	0	16	41	2057	
3.	0.66	0.53	11	11	53	5		
4.	0.71	0.40	3	10	57	10	2056	
5.	0.43	0.33	8	11	35	26	2057	
6.	0.71	0.60	20	3	0	57		
7.	0.53	0.26	8	16	13	43		
8.	0.41	0.33	0	28	17	33	2	2057
9.	0.37	0.40	17	30	26	7		
10.	0.47	0.13	38	23	11	8		2057
11.	0.53	0.46	7	3	27	42	1	2056
12.	0.65	0.60	8	12	6	52	2	2055,2056
13.	0.67	0.06	3	8	54	12	3	2057
14.	0.81	0.20	5	65	10	0		2056,2059
15.	0.80	0.40	64	11	5	0		
16.	0.43	0.13	16	20	35	8	1	
17.	0.33	0.13	8	28	27	12	5	
18.	0.70	0.20	56	8	6	4	6	
19.	0.57	0.06	12	8	46	9	5	
20.	0.67	0.13	12	8	54	6		

Difficulty Index (FV)

Difficult Items = 30%

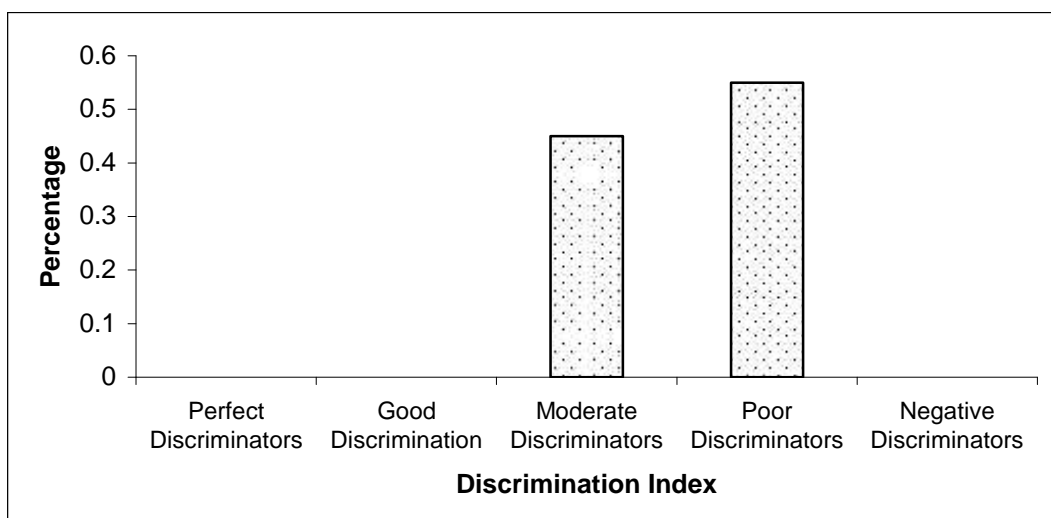
Moderate Items = 70%

Easy Items = 0%



Discrimination Index

1. Perfect Discriminators = 0
2. Good Discriminators = 0
3. Moderate Discriminators = 45%
4. Poor Discriminators = 55%
5. Negative Discriminators = 0%



The table and diagram in the preceding pages show the recorded measures of the test, viz. difficulty index, discrimination index, power of the distracters, items not responded and test items repeated within the years from 2055 to 2059 B.S.

Facility Value

To examine the facility value of the test items, 30% items (six out of twenty) were easy items, 70% items (fourteen out of twenty) moderate and none of the items was difficult item.

Discrimination Index

To examine the discrimination index of the test items, none of the items were perfect discriminators and Good discriminators. 45% items (nine out of twenty) were moderate discriminators and 55% items (eleven out of twenty) were poor discriminators. None of the items were negative discriminators.

Power of the Distracters

If we analyze the power of the distracters, distracter 'A' in item no. 8, distracter 'B' in item no. 2, distracter C in item no. 6 and distracter 'D' in item no. 14 and 15 were non-functioning. Distracter 'A' of item no. 4, 13, distracter 'B' of item no. 6 and 11 were non-functioning since they attract less than 5% respondents.

Repetition of the Test Items

Most of the test items asked in the year 2058 were repeated from the year 2057. Item no. 2, 5, 8, 10 and 13 were repeated from the year 2057. Item no. 4, 11 and 14 were repeated from the year 2056. Item no.

12 has been repeated two times in the year 2055 and 2056. As a whole 45% items have been repeated.

Table 5: Compiled form of the extended "Item analysis" of test retest result of the test items asked in the year 2059

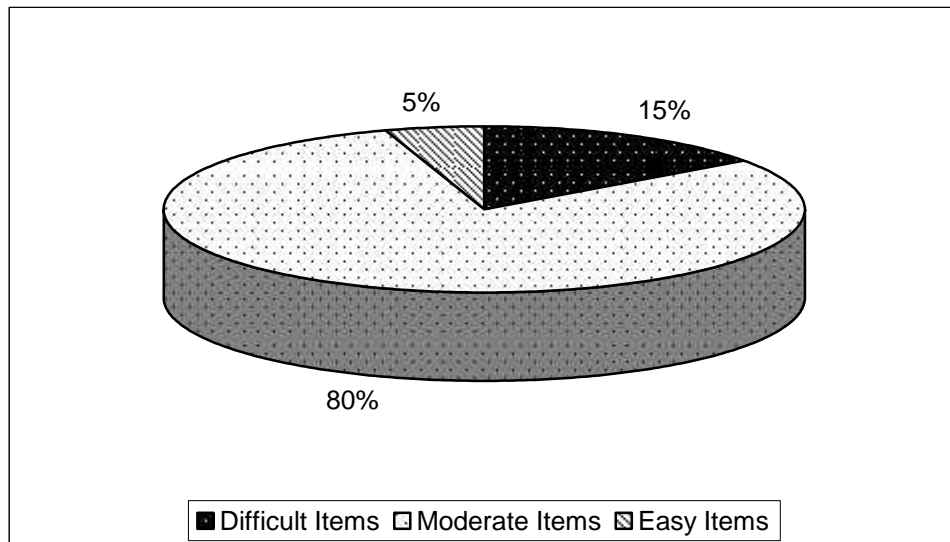
S.N.	Difficulty Index	Discrimination index	Power of distracters				No response	Repeated Items
			A	B	C	D		
1.	0.66	0.46	11	11	53	5		2056
2.	0.53	-0.06	6	7	14	43		
3.	0.26	0	2	22	30	21	5	2055
4.	0.62	0.20	18	3	50	9		2056
5.	0.58	0.20	23	47	5	5		2057
6.	0.55	0.26	10	21	5	44		
7.	0.37	0	30	9	32	9		
8.	0.30	0.13	19	19	18	24		2057
9.	0.55	0.13	30	44	3	3		2057
10.	0.60	0.47	2	3	48	27		2055
11.	0.51	0.40	12	41	21	6		
12.	0.47	0.33	38	27	3	12		
13.	0.83	0.47	67	6	0	6	1	
14.	0.55	-0.13	8	44	20	8	1	
15.	0.23	0	19	41	8	6	6	
16.	0.53	-0.02	5	3	23	43	6	
17.	0.53	-0.06	10	11	43	10	6	
18.	0.32	0.06	14	16	18	26	6	
19.	0.20	-0.13	8	14	16	28	14	
20.	0.35	0.33	28	25	8	13	6	2055,2056,2058

Difficulty Index (FV)

Difficult Items = 15%

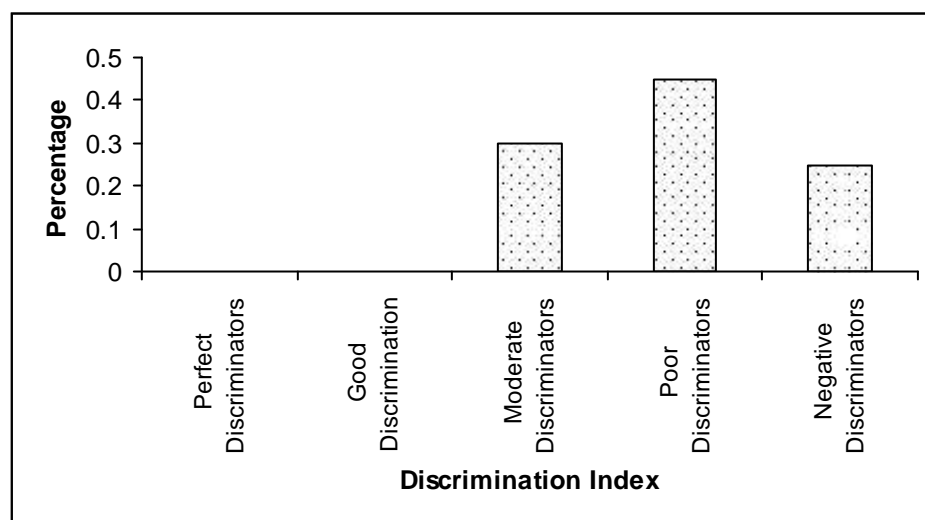
Moderate Items = 80%

Easy Items = 5%



Discrimination Index

1. Perfect Discriminators = 0
2. Good Discriminators = 0
3. Moderate Discriminators = 30%
4. Poor Discriminators = 45%
5. Negative Discriminators = 25%



The table and diagram in the preceding pages show the recorded measures of the test, viz. difficulty index, discrimination index, power of the distracters, items not responded and test items repeated within the years from 2055 to 2059 B.S.

Facility Value

To examine the facility value of the test items, 15% items were difficult items, 80% items were moderate items and 5% items were easy items.

Discrimination Index

To examine the discrimination index of the test items, none of the items were perfect discriminators and good discriminators. 30% items were appeared to have been moderate discriminators, 45% items were poor discriminators and 25% items were appeared to be Negative discriminators.

Power of the Distracters

To analyze the individual distracters, distracter 'A' of item no. 3, 10 distracter 'B' of item no. 4, 16 distracter 'C' of item no. 9, 12 and distracter 'D' of item no. 9 were found poor discriminators. Distracter 'C' of item no. 13 was found non-functioning distracters. Other distracters were found to function satisfactorily.

Repetition of the items (within the year from 2055-2059)

Item no. 3, 10 and 20 have been repeated from the year 2055, item no. 1, 4 and 20 were repeated from the year 2056, item no. 5, 8 and 9 were repeated from the year 2057. Item no. 20 was repeated three times

in the year 2055, 2056 and 2058. 40% items (eight out of twenty) were repeated as a whole.

Table 6: Difficulty Index and Discrimination Index of the tests referred to in the Table from 1 to 5

S.N.	2055		2056		2057		2058		2059	
	FV	DI	FV	DI	FV	DI	FV	DI	FV	DI
1.	0.86	0.26	0.6	0.53	0.66	0.40	0.75	0.13	0.66	0.46
2.	0.63	0.26	0.68	0.40	0.56	0.33	0.51	0.26	0.53	0.06
3.	0.68	0.4	0.48	0.40	0.53	0.27	0.66	0.53	0.26	0
4.	0.62	0.53	0.62	0.53	0.42	0.40	0.71	0.40	0.62	0.20
5.	0.77	0.13	0.63	0.60	0.48	0.66	0.43	0.33	0.58	0.20
6.	0.73	0.26	0.4	0.33	0.52	0.20	0.71	0.60	0.55	0.26
7.	0.53	0.26	0.53	0.46	0.23	0.20	0.53	0.26	0.37	0
8.	0.65	0.46	0.4	0.66	0.28	0.33	0.41	0.33	0.30	0.13
9.	0.47	0.33	0.45	0.26	0.51	0.33	0.37	0.40	0.55	0.13
10.	0.81	0.33	0.25	0.06	0.65	0.53	0.47	0.13	0.6	0.47
11.	0.52	0	0.47	0.66	0.55	-0.06	0.53	0.46	0.51	0.40
12.	0.38	0.6	0.57	0.53	0.3	0.13	0.65	0.60	0.47	0.33
13.	0.67	0.46	0.57	0.40	0.75	0.26	0.67	0.06	0.83	0.47
14.	0.63	0.20	0.65	0.40	0.41	0.20	0.81	0.20	0.55	-0.13
15.	0.83	0.33	0.63	0.13	0.48	0.20	0.80	0.40	0.23	0
16.	0.68	0.40	0.52	0.46	0.61	0.26	0.43	0.13	0.53	-0.02
17.	0.76	0.33	0.61	-0.13	0.26	0.13	0.33	0.13	0.53	-0.06
18.	0.53	0.66	0.75	0.33	0.51	0.46	0.70	0.20	0.32	0.06
19.	0.36	0.13	0.37	0.33	0.66	0.60	0.57	0.06	0.2	-0.13
20.	0.53	0.06	0.62	0.20	0.76	0.40	0.67	0.13	0.35	0.33
Total	12.63		10.8		10.13		11.71		9.54	
Avg.	0.63		0.54		0.51		0.59		0.48	

CHAPTER FOUR

FINDINGS AND RECOMMENDATIONS

Year Specific Findings

2055

1. Statistical identity of the test,
 - a. 10%, 60% and 30% items in the test are found difficult, moderate and easy respectively.
 - b. 55% and 45% test items are found to be moderate discriminators and poor discriminators. None of the items are found to be perfect and negative discriminators.
2. Statistically efficient items are given below.

Item no.	Facility Value	Discriminator Index
3	0.68	0.40
4	0.62	0.53
8	0.65	0.46
9	0.47	0.33
10	0.81	0.33
13	0.67	0.46
15	0.83	0.33
16	0.68	0.40
17	0.76	0.33
18	0.53	0.66

3. 35% items (seven out of twenty) are found to be repeated in the year 2055.
4. The average facility value of the test is 0.63.

2056

1. Statistical identity of the test.
 - a. 5%, 90% and 5% items in the test are easy, moderate and difficult respectively.
 - b. 75%, 20% and 5% items are moderate, poor and negative discriminators respectively. None of the items discriminates perfectly.
2. Statistically efficient items with their facility value and discriminator index are given below

Item no.	Facility Value	Discriminator Index
1	0.60	0.53
2	0.68	0.40
3	0.48	0.40
4	0.62	0.53
5	0.63	0.60
6	0.40	0.33
7	0.53	0.46
8	0.40	0.66
11	0.47	0.66
12	0.57	0.53
13	0.57	0.40
14	0.65	0.40
16	0.52	0.46
18	0.75	0.33

3. Item no. 4 has been repeated in the year 2056, 2058 and 2059. 60% items (twelve and of twenty) are repeated.
4. The average facility value of the test is 0.54.

2057

1. statistical identity of the test,
 - a. 15%, 75% and 10% items in the test are difficult, moderate and easy respectively.
 - b. 50%, 45% and 5% items have moderate, poor and negative discriminators.
2. Statistically efficient items are given below:

Item no.	Facility Value	Discrimination Index
1	0.66	0.40
2	0.56	0.33
4	0.42	0.40
5	0.48	0.66
9	0.51	0.33
10	0.65	0.53
18	0.51	0.46
19	0.66	0.60
20	0.76	0.40

3. Item no. 20 is repeated in the year 2055, 2058 and 2059. 50% items are repeated in the year 2057.
4. The average facility value of the test is 10.11.

2058

1. Statistical identity of the test,
 - a. 30%, and 70% items in the test are easy and moderate. None of the items are found difficult items.
 - b. 45%, and 55% items are moderate and poor discriminators. None of the items are perfect and negative discriminators.
2. Statistically efficient items are given below:

Item no.	Facility Value	Difficulty Index
3	0.66	0.53
4	0.71	0.40
5	0.43	0.33
6	0.71	0.60
8	0.41	0.33
9	0.37	0.40
11	0.53	0.46
12	0.65	0.60
15	0.80	0.40

3. 45% (nine out of twenty) items are repeated items.
4. The average facility value of the test is 0.58.

2059

1. Statistical identity of the test,
 - a. 15%, 80% and 5% items in the test are difficult, moderate and easy respectively.
 - b. 30%, 45% and 25% items have moderate, poor and negative discriminators respectively.
2. Statistically efficient items are given below,

Item no.	Facility Value	Discrimination Index
1	0.66	0.46
10	0.60	0.47
11	0.51	0.40
12	0.47	0.33
13	0.83	0.47
20	0.35	0.33

3. Item no. 20 is repeated from the year 2055, 2056 and 2058. 40% items have been repeated.
4. Test has the average facility value of 0.47.

4.1 General Findings

1. 13% items were difficult, 75% items were moderate and 12% items were easy in total.
2. Perfect discrimination is impossible.
3. 51% items were moderate discriminators, 42% items were poor discriminators and 7% items were negative discriminators.
4. 8.75% (35 out of 400) distracters were poor distracters.
5. 46% items were repeated as a whole.
6. Too easy and too difficult items fail to discriminate between the testees.
7. Items with good facility value need not necessarily have good discrimination index.
8. 48% items in the year from 2055-2059 are statistically satisfactory.

4.2 Recommendations

After the findings of the study, the researcher would like to make the following recommendations.

1. Only the statistically satisfactory items should be included in the final version of the test.
2. Multiple-choice items should be arranged in the order of increasing difficulty.
3. Items listed in the year-wise finding of the work are statistically efficient which could be used safely for the future tests.
4. Only pre-tested items should be used repeatedly.

REFERENCES

- Allen J.P.B. and Davies, 1977, **Testing and Experimental Methods**, Edinburgh Course in Applied Linguistics, Vol. 4, Oxford University Press.
- Baral, R.P., 2004, **Item Analysis of Multiple Choice Objective Test: A Practical Study**, M.Ed. Thesis, Department of English Education, T.U., Kirtipur.
- Bhattarai, G.R., 2001, **A Thematic Analysis of Research Reports**, Kathmandu, Ratna Pustak Bhandar.
- Brown, James, **Objectives Tests, Their Construction and Analysis**, A Practical Handbook for Teachers, London, Longman.
- Ross, C.C., 1941, **Measurement in Today's School**, Prentice Hall, Inc. New York.
- Dhakal, T., 2000, **Effectiveness of Discrete Point Test and Integrative Test as a Measures of English Language Proficiency**, M.Ed. Thesis, Department of English Education, T.U.
- Gautam, P., 2001, **An Analysis of Subjective Marking**, M.Ed. Thesis, Department of English Education, T.U.
- Giri, R., 2003, **Language Testing: Then and Now**, Journal of NELTA, Vol. VIII, Kathmandu, Nepal English Language Teachers' Association.
- Harrison, A., 1991, **A Language Testing Handbook**, ELTS/Longman.

- Heaton, J.B. 1985, **Writing English Language Tests**, FLBS A practical Guide for Teachers of English as a Second or Foreign Language.
- Hughes, A., 1983, **Current Development in Language Testing**, Academic Press.
- Hughes, A., 1989, **Testing for Language Teachers**, Cambridge University Press.
- J. Stanley, et. al., 1958, **Evaluating Pupil Growth**, Boston, Allyn and Bacon.
- Jon Clark and et. al., 1971, **Essential of Testing**, Addison Wesley Publishing Company, London.
- Khaniya, T., 2005, **Examination for Enhanced Learning**.
- Lado, R. 1993, **Language Testing**, Longman.
- Neupane, M., 2004, **Washback Effect of Examination: A Case of Communicative English**, M.Ed. Thesis, Department of English Education, T.U.,
- Ojha, D., 2005, **Content Validity of ELT Theories and Methods Exam at B.Ed. Level**, M.Ed. Thesis, Department of English Education, T.U.
- Richards, J.C. et. al., 1999, **Longman Dictionary of Applied Linguistic**, Longman.
- Weir, C. 1998, **Communicative Language Testing**, U.K., Prentice Hall.