The data collected in the pre try-out of the Reading Ability Test Battery enabled the investigator to eliminate gross deficiencies in the tests and make necessary alterations in the general directions to the testees on how to write the tests. Thus the test battery was now ready for the try-out.

The try-out is the process through which information about each and every item of the test is obtained. According to Lindquist, the main objectives of the try-out test are:

i) to identify weak or defective items and to reveal needed improvement,

ii) to determine the discriminating power of each individual item

iii) to provide data needed to determine appropriate time limit for the finished test

iv) to identify non-functioning and plausible distractors in multiple choice items

v) to determine the number of items in the test.

vi) to determine the difficulty of each individual item to facilitate selection of items and
vii) to determine the needed improvements in the process of administering the test.

(Lindquist. 1955 pp 250-251).

5.1 Sample of the Try-out Test.

Sampling is the most widely used concept in research to ascertain information required in answering questions about a specific population. It is a process by which relatively small number of individuals or measures of individuals are selected for item analysis.

As the main purpose of the try-out test is to give a final shape to the test material according to the discriminating power and difficulty level of individual items, the sample chosen for the try-out should be capable of supplying the necessary data for this purpose. Therefore, it is always advisable to select the sample from different schools of the same type rather than from one school.

In the present study, for the purpose of the try-out, it was decided to administer the Reading Ability Test Battery to a sample of 200 girl students of the tenth standard of the Public Schools of Delhi. This decision was taken because of the following reasons.

Firstly, as the administration of the test battery was expected to be a time-consuming process, the school authorities were not willing to spare the students of both the ninth and eleventh classes for both the try-out and the final administration of the test. They were willing to
spare the students of tenth standard for the try-out and those of the ninth and eleventh for the final tests.

Secondly, the try-out was proposed to be conducted in August 1991, when the tenth standard students would be neither far advanced nor far behind with their reading ability. The item-analysis criteria applied for item selection after the try-out would account for the difficulty level and discrimination level of each item. Therefore, an item which was suitable for the tenth standard students could be presumed to be suitable for both the ninth and eleventh standard students as well.

Thirdly, if the try-out was to be given to the ninth and eleventh class students separately there would be the practical difficulty of co-ordinating the performance of the students of these two classes with regard to the difficulty and discrimination levels of items and reaching a conclusion on the selection of items suitable for both the classes. The performance of the tenth standard students was expected to provide a sort of median measurement of reading ability, enabling the investigator to construct a test battery suitable for both the ninth and eleventh standard students.

According to Verma (1965): "The size of the sample for the try out also depends on what method or ready reckoner one intends to use for item analysis." For example, if the dichotomous criterion is adopted for item analysis, a convenient size of the sample can be easily worked out. In this method, the top and bottom twenty seven percent of the sample are utilized for which ready reckoners are available. Thus, if a sample of 185 is
taken, the top and bottom twenty seven percent would be fifty cases each. So the proportions of cases passing an item in top and bottom groups can be easily calculated.

Keeping the aforesaid points and also the objectives of the study in mind, a sample of two hundred students was drawn from three different schools in Delhi. It had already been decided to use the dichotomous criterion with the top and bottom twenty seven percent for the purpose of item analysis. For this purpose also a sample of 200 was the optimum number as the investigator could select 185 after discarding incomplete and highly erroneous answer sheets. The table 5.01 shows the administration schedule of the try-out tests of the Reading Ability Test Battery.

Table 5.01: Showing the Administrative Schedule of the Try-out Tests

<table>
<thead>
<tr>
<th>Sl. No.</th>
<th>Name of the School</th>
<th>Number of Students</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>St. Antony's Girls' High School, Hauz Khas, New Delhi</td>
<td>60</td>
<td>7.8.91</td>
</tr>
<tr>
<td>2.</td>
<td>Holy Child Convent Girls' High School, Tagore Garden, New Delhi</td>
<td>70</td>
<td>10.8.91</td>
</tr>
<tr>
<td>3.</td>
<td>Carmel Convent Girls' High School, Chanakyapuri, New Delhi</td>
<td>70</td>
<td>12.8.91</td>
</tr>
</tbody>
</table>
5.2 Administration of the Try-out Test

The dates for the try-out test were fixed in consultation with the principals of the concerned schools as 7th, 10th and 12th of August. This was the convenient time for both the teachers and students as the first terminal examinations were just finished then. Moreover after the commencement of the second term, the school authorities were not willing to disturb the studies of the 10th grade students.

The school authorities were good enough to spare one full day for the tryout purpose in each school. As the students had to write ten tests, after the first five test, they were given an interval of 30 minutes. This helped the students to relax and take up the rest of the five tests with undiminished spirits.

A very important point a researcher has to heed to is the uniformity in the administration of the test.

According to Thorndike:

"One of the major goals of effective organisation of test administration is the achievement of uniformity of procedure in test administration, scoring and weighting, so that final evaluation of a given individual will be the same no matter where, when or by whom he was tested."

(Thorndike 1949 p. 256)

To ensure uniformity in the whole procedure, the investigator herself administered the tests in all the
schools. However, the willing help of the teachers of the respective schools was made use of for the healthy conduct of the test.

Another point educationists have stressed is that the conditions for the test must be favourable. The test should be given in the familiar environment of the pupils' own classroom. Unnecessary distractions and interruptions should be avoided. All these aspects were taken into consideration when the try-out test was given in all the schools.

Along with the test booklets the students were given separate answer sheets to mark their answers. They were required to write the name of the test and other particulars like their name, age, school and the date of the test in columns provided for that purpose on the answer sheet. On the top right hand corner of the answer sheet in the space provided, they had to mention the starting time, finishing time and the total time taken for each test. To assist the students in keeping this record, a time piece was kept on an elevated position in front of the students.

On the answer sheet, after each question number, four boxes labelled 'a' 'b' 'c' and 'd' were given. (cf. Appendix B). The students were asked to put a cross mark in the box which bears the serial alphabet of their choice. An advantage of this type of arrangement on the answer sheet is that, in case the students wanted to make corrections they could do so by cancelling the first answer and putting a cross mark in the column of their next choice. The students were instructed to cancel a choice by putting three lines across, over the first cross mark.
Detailed instructions on how to write the test were given on the facing sheet of the test booklet. In addition to this, separate directions about how to write individual subtest were also given on top of every subtest with a worked-out example. The instructions on the cover page as well as the directions preceding the exercises were drafted on the basis of the advice obtained from experts.

As the students were not used to this type of test procedure, the investigator read out the instructions and explained everything in detail to the students and clarified all their doubts before the commencement of the test. In all the schools, the investigator could take the students into confidence by keeping an informal, friendly and sympathetic air throughout the test procedure.

5.3 Time-length for the Try-out

As it was a try-out, no time length was fixed. As Traxler (1951) says in his 'Administering and Scoring the Objective Test', sufficient time must be allowed for answering the test, so that almost all the students must be able to answer almost all the items. Hence, it was decided to give ample time to the students for completing the test. As one purpose of the try-out test is to determine the time length for each subtest from the performance of the testees, the investigator took particular care to see that the students mentioned the starting time, the finishing time and the total time taken in the columns meant for them on the answer sheet.
5.4 Scoring of the Try-out Test

There are two important factors which require proper attention in the scoring procedure. They are the method of weighting and the correction for guessing.

5.4.1 Weighting

Usually in school tests items differ in weightage depending upon the difficulty and importance of items. But when different weights are allotted to items, the scoring work becomes complicated, laborious and slow.

Ross suggests:

"The scoring procedure adopted should be fairly simple. As a rule, the best procedure in scoring objective tests is to give one point of credit for each correct response. It is unnecessary to weight the items according to estimated difficulty or importance as it has been empirically proved that almost all pupils will be in the same rank order regardless of the weighting of the individual items."

(Ross. 1956 p. 156)

This view has been endorsed by several other researchers. They have established the fact that there is very high correlation between weighted and unweighted scores of the objective tests. Therefore, the difference between weighted and unweighted scores may be considered negligible. In keeping with their advice, no weightage was allotted in the present investigation. One point was given for each fully correct response of the item.
5.4.2 Correction for Guessing

In multiple choice tests there are possibilities of students making guesses while selecting a response from the alternatives given.

According to Rawat:

"The guessing factor does not present as much of a problem in multiple choice items as in several other types. If the number of alternatives or choices are four or less than four, a correction formula eliminates the guessing."

(Rawat 1965, p. 49)

In the present study, there being equal number of choices for all the items, and there being scope for guessing, the correction formula was applied. The formula is as given below:

\[
RC = \frac{R - W}{K - 1} \quad \frac{R + W}{R + W}
\]

where,

RC = The proportion of correct responses corrected for chance success.
R = The number answering an item correctly.
W = The number answering an item incorrectly.
K = The number of alternative responses in an item.

The correct proportion for an item according to this formula can be directly read off a table (cf. Guilford 1965 p. 421).
In the present study to apply the correction on individual raw scores, a response count of correct (R) responses, incorrect (W) responses and omitted (O) responses was made.

The item analysis data sheet was prepared keeping in view the requirements of both the qualitative and quantitative analysis of the test items. In addition to the total count of 'R', 'W' and 'O', a count of responses to each distractor was made. This enabled the investigator to identify ambiguous, non-functioning or implausible distractors, so that necessary improvement could be made to items through such a qualitative analysis.

Thus, the procedure for item analysis having been decided, the next step was the evaluation of the answer sheet. The investigator evaluated the answer sheets of 200 students giving one point to every correct response.

5.5 Arranging the Test booklets

In the present study dichotomous criterion with top and bottom twenty seven percent cases was decided to be used for item analysis. Therefore, a sample of 185 booklets was to be studied statistically. The incomplete test booklets and those with numerous erroneous responses were discarded and 185 test booklets out of 200 were selected. These 185 booklets were arranged in descending order of score, i.e., one with the highest score on the top and one with the lowest score at the bottom. These booklets were then numbered 1 to 185.
Kelley (1939) has shown that by comparing approximately the upper and lower twenty seven percent of the total group, the most accurate determination of item validities or internal consistencies can be obtained. Elimination of the middle forty six percent leads to results that are more consistent from sample to sample, than those obtained from using all available cases.

Accordingly the test booklets were divided into three groups.

1. The upper group called 'U' - 27% of 185, i.e., 50 booklets 1-50.
2. The middle group called 'M'. 46% of 185, i.e., 85 booklets, 51-135.
3. The lower group called 'L' - 27% of 185 i.e., 50 booklets 136 - 185.

Group 'U' represented high achievers and Group 'L' represented low achievers.

5.6 Item Analysis

The reliability and validity of a test mostly depends on the nature and type of items which are included in it. The effectiveness of an item in an achievement test depends on the following factors:

a. the discriminating power of an item with reference to both validity and internal consistency.
b. the difficulty level of an item and
c. the plausibility of the distractors.
Item analysis is the simple statistical way of checking each item in an achievement test usually consisting of multiple-choice items. This process provides information about the discrimination index, difficulty value and curricular validity of an item. According to Ebel (1954): "Item analysis reveals certain weaknesses such as excessive difficulty, low discriminating power or poorly functioning distractors." Thus item analysis helps the test constructor to eliminate poor items, i.e., items which are too easy and too difficult and which do not discriminate between the better and the poorer students. It also helps to identify the implausible distractors.

This information, remarks Guilford (1965) is valuable for several reasons. It acts as a check on the test writer's subjective judgement in the selection of the test items. Such checks are desirable even if the test writer is an expert, because with such checking the test writer learns to improve in his art. He learns how examinees react to items in general and to each item in particular. It also helps him in diagnosing the nature of individual difficulties of the students and to give remedial instructions accordingly. Also, the test writer gains new insights into the kind of items that best suits one particular kind of test and can formulate new hypotheses concerning the nature of the ability being measured.

5.6.1 Difficulty value

The difficulty of a test item is usually expressed in terms of the number or percentage of pupils answering the item correctly. There does not seem to have a consensus among experts about the optimum difficulty level of items
in an achievement test. According to Hawkes et al. (1936), "the average difficulty of all items should be about fifty percent." But it is always desirable to include both easy and difficult items along with items with the level of fifty percent pass. An item which is answered by all pupils and an item which is not answered by all the pupils should not be included in a standard test.

In the present study, the difficulty value of the items was determined by using the data obtained from 27 percent upper and lower groups. The difficulty values were directly read off the Item Analysis Chart prepared by Harper, Gupta and Sangal (1962).

5.6.2 Discrimination Index

Another trait of a good achievement test is that its items must be able to discriminate the achievements of various pupils. The discriminating value of an item according to Bean (1963) is the degree to which any single item separated the superior individuals from the inferior ones in the trait being measured.

In the present study, for the calculation of discrimination value Kelley's (1939) method of upper 27 percent and the lower 27 percent of the sample was adopted. Discrimination values were also read off directly from Harper's Item Analysis Chart.

The item analysis chart prepared by Harper et al. (1962) is a sophisticated and time-saving device as it eliminates the necessity for tabulating all of the available data. Kelley (1939) as mentioned earlier had
shown that the product moment correlation between a test item score and the total score on the test can be estimated by using the top and bottom 27 percent tails. Flanagen (1939) prepared tables for the estimation of the product-moment correlation from the 27 percent tails of the distribution. Davis (1946) translated the non-linear \( r \) into a linear scale of discrimination values by using Fisher's \( z \). Thus Davis translated Flanagen's tables into tables from which his linear discrimination index could be read. He also added a difficulty index to his tables so that both discrimination and difficulty could be read at the same line. Davis's difficulty index is based on a linear transformation of the percent passing. Davis calculated percent passing as the mean of the percent passing in the high and low groups. Fan (1954) has shown that the \( p \) values calculated by this method are accurate only when the item-criterion correlation is extremely low. Using Pearson's Tables, Fan estimated the \( p \) values for various levels of item-criterion correlation. Harper et al. (1962) translated Davis's discrimination index into a chart. Instead of using the difficulty values in Davis's tables, they took Fan's difficulty \( p \)'s and translated them into Davis's linear difficulty index, thus making it possible to read the discrimination index and difficulty value of an item at the same line.

### 5.7 Plausibility of Distractors

The frequencies of responses to distractors in respect of the Upper Group and the Lower Group give a clear idea about the plausibility of distractors.
The following are the criteria for good distractors:

a. All distractors should be able to attract some examinees.

b. No distractors should attract more examinees than the correct choice.

c. An incorrect choice should attract lesser number of examinees from the upper group than the lower group.

The responses were examined in respect of the above criteria. In the item analysis data sheet a count of response to each distractor was given. It was discovered that some distractors had attracted almost negligible number of testees. Such distractors were improved. In a few cases, the wrong distractors had attracted more pupils from the upper group, thus making the items negatively discriminative. Such items were dropped (see Appendix C).

5.8 Item Selection

Lindquist suggests that the following principles should be observed while selecting the items for the final test:

(i) A difficulty index should be computed for each item.

(ii) A discrimination index should be computed for each item.

(iii) The number of items desired at each level of difficulty should be estimated.

(iv) All the try-out items should be separated into groups indicated in the outline of the test.
(v) The entire group of test items should be examined to detect unnoticed overlapping of choices.
(vi) The choice-by-choice item analysis data for each item should be studied.
(vii) The items should be grouped approximately arranged in order of difficulty.

(Lindquist 1955 pp. 313-315)

In selecting the items for the final test, in the present study, all these points were kept in mind. Items ranging between twenty-eight percent to seventy-four percent difficulty values and sixteen percent and above discrimination values were selected for the final run.

In the first vocabulary test, out of 35 items, 26 items having suitable difficulty and discrimination values were selected. The second and third Vocabulary Tests contain 25 items each. These two tests were taken from Grover's (1990) Reading Ability Test Battery. In the fourth Vocabulary Test, out of 35 items only 25 items were retained and in the fifth vocabulary test 28 items were retained out of 35. In the sixth Vocabulary Test the number of items retained for the final test was 25 out of 32. The seventh Vocabulary Test consisted of 23 items selected out of 30 items at the try-out stage. In the eighth Vocabulary test 24 items were retained out of 35 try-out items.

Similarly in Comprehension Tests, in the first subtest 25 items were retained out of 32 and in the second 18 items out of 25 were retained. In Comprehension Subtests III, 16 items were retained, out of 20 and in Subtests IV,
18 out of 20 were retained. Comprehension subtests V, VI and VII were taken from Grover (1990). Subtests V and VI consisted of 25 items each. Subtest VII consisted of ten passages, each with a set of three items.

All the selected items in each subtest were rearranged in an ascending order of difficulty values, the first item of each subtest being the easiest one.

5.9 Determining the Time-length.

Before finalising the tests to be administered to a large sample, it is very essential to fix appropriate time limits for answering each test. The time limit for the test depends upon the purpose of the test and upon the ability and experience of the pupils. Researchers have agreed on the point that adequate time should be granted as far as possible to all pupils to consider all items in the test. Hawkes et al. (1935) have suggested that in general achievement tests, the time allowance should be so adjusted that at least 75 percent of the pupils will have time to consider all items in each test. Menzel (1952) has observed that in any test requiring ingenuity, ample time should be given for about two-thirds of the examinees to finish all they are capable of finishing. Ruth (1929) suggests that time should be liberally granted to testees so that ninety percent can attempt all items within their power. It has been experimentally proved by Ghosh (1958) that with increasing time allowance, the mean scores of the group go on increasing, until we arrive at a certain time limit after which additional time allowance brings no change in mean score. From the results we find that mean score obtained under the time limit does not significantly differ from the mean score obtained under unlimited time.
Therefore, it was decided to fix liberal time limits for the final test. Time taken by all the students in each subtest was noted down separately from their try out test. The period within which 90 percent of the students could complete the test was calculated using the formula:

\[ \text{Time allowed} = \frac{T \times N_f}{N_t} \]

Where,

\[ T = \text{Time period within which 90 percent of the testees could complete a test.} \]
\[ N_f = \text{Number of items in the final form.} \]
\[ N_t = \text{Number of items in the try-out form.} \]

In this way, the time limit within which 90 percent of the pupils could complete the tests was calculated separately for all the tests.

The details of the selected items and the time limit for each subtest are given in Table 5.02.

Table 5.02 Showing the number of items selected and the time limit for each test

<table>
<thead>
<tr>
<th>Sl. No.</th>
<th>Test</th>
<th>No. of items</th>
<th>Time limit (Minutes)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Pre</td>
<td>Try</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.</td>
<td>Vocabulary Test I (Recognition of Word Meaning)</td>
<td>40</td>
<td>35</td>
</tr>
<tr>
<td>2.</td>
<td>Vocabulary Test II (Synonyms)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Sl. No.</td>
<td>Test</td>
<td>No. of items</td>
<td>Time limit (Minutes)</td>
</tr>
<tr>
<td>--------</td>
<td>----------------------------------------------------------------------</td>
<td>--------------</td>
<td>---------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pre-Try-Out</td>
<td>Try-Out Selec-ted</td>
</tr>
<tr>
<td>3</td>
<td>Vocabulary Test III (Antonyms)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>Vocabulary Test IV (Identification of a Word to replace a phrase or idea)</td>
<td>40</td>
<td>35</td>
</tr>
<tr>
<td>5</td>
<td>Vocabulary Test V (Recognition of Words in context)</td>
<td>40</td>
<td>35</td>
</tr>
<tr>
<td>6</td>
<td>Vocabulary Test VI (Recognition of a Keyword for a Proverb)</td>
<td>35</td>
<td>32</td>
</tr>
<tr>
<td>7</td>
<td>Vocabulary Test VII (Recognition of the Meaning of a Phrasal Verb)</td>
<td>35</td>
<td>30</td>
</tr>
<tr>
<td>8</td>
<td>Vocabulary Test VIII (Recognition of Diction and Usage)</td>
<td>40</td>
<td>35</td>
</tr>
<tr>
<td>9</td>
<td>Comprehension Test I (Recognition of the Paraphrase of the Keyword)</td>
<td>40</td>
<td>32</td>
</tr>
<tr>
<td>10</td>
<td>Comprehension Test II (Recognition of Idiomatic expressions)</td>
<td>30</td>
<td>25</td>
</tr>
<tr>
<td>11</td>
<td>Comprehension Test III (Recognition of Sentence Meaning)</td>
<td>24</td>
<td>20</td>
</tr>
<tr>
<td>Sl. No.</td>
<td>Test Description</td>
<td>No. of items</td>
<td>Time Limit</td>
</tr>
<tr>
<td>--------</td>
<td>--------------------------------------------------------</td>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pre</td>
<td>Try-Selected</td>
</tr>
<tr>
<td>12.</td>
<td>Comprehension Test IV (Recognition of the Meaning of Proverbs)</td>
<td>25</td>
<td>20</td>
</tr>
<tr>
<td>13.</td>
<td>Comprehension Test V (Reading for Inference)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>14.</td>
<td>Comprehension Test VI (Reading for General Significance)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>15.</td>
<td>Comprehension Test VII (Reading to note Details)</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

On the whole, the Reading Ability Test Battery consisting of two sets - 8 Vocabulary Tests and 7 Comprehension Tests - was calculated to be completed in 3 hours and 3 minutes. The time limit for each subtest was decided to be given on the top right hand corner of each subtest. It was also decided to request the pupils to keep the time schedule strictly, though over-strictness was to be avoided.

An estimated time of fifteen to twenty minutes was expected to be utilised for distributing the question papers and answer sheet to the testees, for giving them the general directions and for clarifying their doubts. An
interval of 30 minutes in between the two sessions of the testing programme was also decided to be granted to students. Altogether administration of the Reading Ability Test Battery was expected to be completed within 4 hours.