CHAPTER III

Construction of Achievement Test in General science:

Construction of an Achievement Test, as any other construction requires deep consideration of many important factors: -

1. Planning the test
2. Preparing the test
3. Trying out the test
   a) Pre-try out
   b) Try out
   c) Item analysis
   d) Administration of the final test

Planning:

This is a very important factor if a testing programme is well planned, it saves wastage of time, overlapping and confusion. It must have well defined objectives, must specify the nature of the pupils to be tested, and indicate how the test scores would be put to use. ("Good Tests do not just happen, nor are they the result of a few moments of high inspiration or exhaustion. On the contrary the process is calm deliberate, and time consuming .... it cannot be emphasised too strongly that the actual process of test construction must be preceded by careful planning if the test is to be successful. The test will be no better than the quality of the thinking
that goes into it. In planning the test consideration
must be given to the nature of the objective to be mea-
sured, the purpose it is to serve and the conditions under
which it will be used. 1

2. Preparing the Test.

Vaughn suggests, "an achievement test in courses of
study to some extent at least be based upon what the pupils
were actually taught rather than upon what some one may
think should be taught". 2 The preparation of the test
must take into account not only the content taught but con-
sider also the behavioural modifications any programme of
instruction is expected to bring about. That is to say
an attempt should be made to ensure curricular validity.
This then should keep in view the precise objectives to be
measured and the areas of instruction from which they are
to be measured. Hence in accordance with the above prin-
ciples the following objectives were listed for measurement
in this investigation.

1. Knowledge of fundamental concepts and principles
   of general science.

2. Reasoning and interpretation of scientific data.

3. Application of scientific knowledge.

1. Ross and Stanley: "Measurement in To-days' Schools"
   pages 140-141.

These objectives were spread over the subject matter field as follows.

<table>
<thead>
<tr>
<th>Subject</th>
<th>109</th>
<th>77</th>
<th>97</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physics</td>
<td></td>
<td></td>
<td></td>
<td>263</td>
</tr>
</tbody>
</table>

Lindquist suggests, that in the construction of an achievement test, a number of text books in the subject be analysed to secure a tentative test of topics to be tested and the importance to be given to each topic. In addition to this a detailed study of the curriculum, the opinion of experts, a critical study of the examination papers and standardized tests in the field is inevitable. Hence in this investigation the following text books of Science prescribed by the department were analysed for the selection of topics for constructing test items.

3. High School Science by Gregory and Hodges
4. High School Chemistry by T.S. Krishna Iyengar
6. Introduction to Biology by N.S. Veerappa.
7. Introduction to Biology by Sundareshan

The syllabus in General Science enclosed in the
Appendix A was critically analysed with regards the importance to be given to each topic. In addition to this, a detailed study of the (1) public examination papers, (2) class examination papers, (3) standardized and semi-standardized tests, (4) notes of lessons of subject teachers, (5) class notes of students in the subjects was made. The opinion of the experienced teachers in the line was taken and the investigators experience as a teacher was made use of in constructing test items.

Last but the most important of all was to consider the reports of the Zakir Hussain Committee, Radhakrishna University Commission Report, C.R. Reddy's Report and the Lakshmanaswamy Mudaliar Committee's Report for the selection of the objectives and the type of examination to be adopted.

The task of writing the test items was taken up after the above mentioned consideration. "Item writing is an art. It requires an uncommon continuation of special qualities. It is mastered only through extensive and carefully supervised practice. As item writing is essentially creative, just as there can be no set formula for producing a good story or a good painting, so there can be no set of rules that will guarantee the production of good test items."

Though according to Eliot Robert there can be no set rules on item writing, it is done according to certain conditional principles.

The language must be simple, appropriate and within the understanding capacity of the pupils. The sentences must not be very long. Each item should allow not more than one answer. There should be no room for ambiguity. There must be minimum writing on the part of the students as far as possible answers must be secured on one side of the paper for easy scoring. "In test construction the prime requisite from the point of scoring is that those pupils reaction to the test which are to be scored be as simple, abbreviated and controlled as possible and the reactions have a definite spatial location". The preliminary draft should include more items than needed in the final test. This facilitates culling out the appropriate from the inappropriate ones at a later stage. This necessitates that every type of tests must be one and a half times as long as the final one. As far as possible the get up of the present test items is in conformity with the above mentioned principles. 283 items were constructed they were spread over the various types of tests as shown in the table below.

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>True false</td>
<td>60</td>
</tr>
<tr>
<td>2</td>
<td>Modified true false</td>
<td>30</td>
</tr>
<tr>
<td>3</td>
<td>Multiple choice</td>
<td>30</td>
</tr>
<tr>
<td>4</td>
<td>Matching</td>
<td>21</td>
</tr>
<tr>
<td>5</td>
<td>Completion</td>
<td>30</td>
</tr>
<tr>
<td>6</td>
<td>Classification</td>
<td>25</td>
</tr>
<tr>
<td>7</td>
<td>Analogies</td>
<td>20</td>
</tr>
<tr>
<td>8</td>
<td>Enumeration</td>
<td>31</td>
</tr>
<tr>
<td>9</td>
<td>Problem</td>
<td>15</td>
</tr>
<tr>
<td>10</td>
<td>Diagram</td>
<td>21</td>
</tr>
</tbody>
</table>

These items were subjected to the critical scrutiny of expert subject teachers and improved on the basis of their suggestions. These items were assembled into the form of a battery. Thus a battery of ten tests was prepared. Instructions were given at the beginning of each and samples were worked out.

The next step in the programme of test construction is to subject the test to a few trials. This is an important step in test construction and standardisation. This way this can be standardized both for its content and method of administration in order to facilitate proper evaluation in its final form regarding its quality and the quality of pupil responses.

Trying out the Test.

The next step is to try the test on a representative sample of the examinees.

According to Conard the following purposes are served by a try out.

1. To identify weak or defective items and to reveal needed improvements. To identify ambiguous, indeterminate, implausible distractors, over difficult and over easy items.

2. To determine the difficulty of each individual item,

---

in order that a selection of items may be made that will show item difficulty, appropriate to the purpose of the finished test.

3. To determine the discriminating power of each individual item.

4. To provide data to determine the number of items to be included in the final test.

5. To find out the time limits for the finished test.

6. To discover needed improvements in the mechanics of test taking in the directions to examinees regarding the method of recording their responses.

7. To determine the inter-correlations among the items in order to avoid overlap in item selection.

Pre-Try-out.

By pre try out is meant a sample try out on a small sample of examinees for the purpose of finding out gross defects and shortcomings in the test with no idea of analyzing the data for individual items. It is just a step to find out how the pupils react to the test. The sample may consist of half a dozen to 100 examinees, or a few adults who try to put themselves in the position of students for whom the test is intended.

This test was preliminarily tried on 40 students of
High School II Year of the Maharaja's High School, Mysore.  
(The time taken for answering the test was about one hour and 45 minutes). The opinion of the boys about the test was taken. They found the test to be interesting and not very difficult. In addition to this the test was administered to the 20 T.C.H. and 20 T.C.L. students of the Government Training College for Men, Mysore, who had studied the same topics. They also found the test to be interesting and of average difficulty.

Findings of the Pre-try-out.-

The test on the whole was found to be neither easy nor very difficult. There was no sign of disgust or indifference in the testees on the other hand a lot of interest was evinced by them in answering the test. Some vague and difficult items were detected. Out of the 283 items given 17 were discarded and the remaining 266 items were grouped under 10 sub-tests. Instructions were given at the beginning of each sub-test and examples were worked out. A copy of the booklet is given in the appendix.

Try-out.

This trial of the test as has been already pointed out is to find out the various kinds of defects in the test items, instruction, type of administration, time length required, difficulty value and discriminating index of each item. Lindquist suggests that if one try out does not point out
all the defects present in a test it may be given two or more try outs. This means physical and financial burden. This can be avoided if necessary precautions are taken to administer the test as far as possible under ideal conditions.

The sample of the try-out must be similar to the one intended for administering the final test. The sample must be efficient that is to say, the sample must give valuable information about the population. The sample must be of adequate strength. But it is not the number tried that is important. The areas or schools from which it is drawn is of importance. A sample of 200 students from many schools drawn from different strata of society is much better than taking the entire sample from one or two schools.

The precautions to be taken into consideration while administering the test, the conditions under which the test is tried and the directions to the examinees should be similar to these while administering the final test.

**Time Length**

Sufficient time must be allowed for answering the test so that a great majority of the testees must be able to answer almost all the items. The time required for answering the test items should not be usually longer than the
school period to avoid the difficulties in the problems of management.

Motivation.

The active cooperation of the testees is essential if the testing programme is to be a success. To secure their cooperation they have to be motivated. There are various methods of motivating. Arthur Taxler suggests that the pupils must be informed beforehand the purpose with which the test is given and the influence it will be having on the students.

Guessing.

There is no agreement on the method to be adopted in directing pupils with regard to guessing. Some suggest that strict instructions must be given not to guess. If all follow this instruction the problem of guessing would be eliminated. But inspection of these instructions some would surely guess. This way those who also could have guessed correctly and got the advantage stand to lose.

Some others suggest that every pupil must be asked to answer every item whether he knows it or not. This allows free and uniform guessing. After applying the correction formula the true score can be gauged. This method makes students careless and adopt loose thinking.

Guessing may be of two types, wild guessing and honest guessing. In wild guessing or dishonest guessing the guessor does not know the correct answer. He takes a chance by choosing some response wildly. Whereas in honest or intelligent guessing the pupil has partial knowledge and he guesses a right response with the help of his partial knowledge. Some suggest that pupils may be asked to guess intelligently. But it is not possible to know who have guessed wildly and who have done it intelligently. The application of the correction formula will not be justified in the case of those who have guessed intelligently.

Directions to the examinees.-

A study by Weidemann and Meivens indicated that the nature of the directions may have considerable effect upon test scores. They tried five sets of directions - for giving tests involving true-false and indeterminate statements and found significant differences in the resulting scores. The directions must be clear and concise.

In this investigation the test (the battery of 10 sub-tests consisting of 266 items) was tried on 172 students of High School I Year at four High Schools in Mysore City, situated in different areas of the City, representing the different strata of society, the Maharaja's, Vidyavardhaka, Meivens, and Weidemann.  

---

Christ the King's Convent and the Shepherds Convent. The schools were requested well in advance to make proper arrangements for the conduct of the test on an examination level.

The cooperation of the Heads of the institutions and the teachers was made use of for smooth conduct of the test. The pupils were informed that they would be having a new experience in answering this new type of test and answering this would revise their portion in General Science and this would help them in their promotional examination which would come off in a couple of days.

Examples given at the beginning of each sub-test were worked out on the board with their help. Thus the method of answering each type of test was explained. After this they were given the test booklets with strict instructions not to open them till they were so asked. They were asked to fill in the blanks on the facing sheet and read the instructions carefully. Instructions were given that they should not guess as this would reduce their score. They were given the opportunity of clarifying their doubts and continued that they should not ask any question after they begin to answer. They were asked to begin on hearing the signal 'start.' A time piece was kept on the table. The time taken for general instruction and working examples on the board was noted to the
be half an hour. They were asked to note the time they start and finish each type of test. The total time taken by about 90 percent of the boys was found out to be about an hour and a quarter. The details of the time taken are given below.

**Statement showing the time taken for answering different Tests.**

<table>
<thead>
<tr>
<th>No. of items</th>
<th>Time taken</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 True false</td>
<td>50</td>
</tr>
<tr>
<td>2 Modified true false</td>
<td>30</td>
</tr>
<tr>
<td>3 Multiple choice</td>
<td>30</td>
</tr>
<tr>
<td>4 Matching</td>
<td>21</td>
</tr>
<tr>
<td>5 Completion</td>
<td>25</td>
</tr>
<tr>
<td>6 Classification</td>
<td>25</td>
</tr>
<tr>
<td>7 Analogies</td>
<td>18</td>
</tr>
<tr>
<td>8 Enumeration</td>
<td>31</td>
</tr>
<tr>
<td>9 Problem</td>
<td>15</td>
</tr>
<tr>
<td>10 Diagram</td>
<td>21</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>85</strong></td>
</tr>
</tbody>
</table>

**Scoring:**

Advice of Kenneth Bean\(^1\) was followed in getting the answers noted in the booklet in a vertical column by pro-

---

1. Kenneth L. Bean: "Construction of Educational and Personnel Test".
proTiding a space on the left hand side of each question. This made scoring easy. Each right response was awarded one mark.

Item Analysis.

After a test is administered the pupils' reaction to the test must be found out. Unless there is correct information is got about pupils reaction to an item its inclusion in the final test is not justified. The worth of a test depends upon the worth of the individual items in it. According to Walter Monroe¹ the effectiveness of an item in a general achievement test depends upon three interrelated factors - (1) the validity of the item from the stand point of content, curriculum and educational objectives, (2) the discriminating power, (3) the difficulty of the item. The first of these factors has been dealt with under "Test construction". The other two are discussed here. The purpose of difficulty value and Discriminating index is to find out how hard an item is for the group tested and how well it discriminates between the strong and weak students.

Difficulty Value.

The difficulty value of an item is the percentage of pupils that have answered the item correctly. When we say the difficulty value is 75 per cent it means that 75

per cent of the pupils have answered the item correctly. The greater the difficulty value the easier is the item. The difficulty value of all the items in the different sub-tests was calculated and tabulated as shown in the table.

**Discriminating Indexes.**

The purpose of educational measurement is to grade pupils according to the degree of their achievement. This implies a great discriminating power on the part of the test. If the test as a whole should have this capacity every item in the test must boast of a high discriminating index.

Discriminative power of a test means that a different quality or response may be expected from different pupils. Superior pupils should answer the item correctly better than the inferior pupils. This assumption gives a procedure of finding of the discriminating index.

Some suggest that after scoring the test book lets they must be arranged in a descending order with the highest score at the top and lowest score at the bottom. The number of persons answering an item incorrectly in the lower 27 per cent, must be found out. The number of persons answering the same item incorrectly in the higher group must be found out. If the item really distinguishes between the good and the bad students then $W_p - W_g$ must be significant and positive. From the method both diffi-
culty value and discriminative efficiency can be found out by using the psychometric Research and Service Chart of Davis which is perfected by Sri B. Dasgupta. To use this table the minimum number of examinees must be 370. Since the number of examinees in this try-out is 172 the above method could not be used.

The other method is the whole group is divided into three groups after arranging the scored booklets in descending order of the obtained scores. The number of correct responses for each item by the pupils in the upper 1/3 is found out. This is compared with the responses that item secures from the lower 1/3 and the D.I. is calculated by using the formula.

\[ D.I = \frac{U - L}{N/3} \]

where D.I. is the discriminative index.

\( U \) = correct responses an item secure from the upper 1/3

\( L \) = correct responses - the lower 1/3

\( N = N_o \) of examinees.

Comparison is made between the high and the low group. The item which secures a higher response from the upper 1/3 than the lower one discriminator well between the able and the backward pupils or the strong and weak knowledge. An item which has a zero discrimination between the response of these groups is useless. The item which receives
a higher response from the lower 1/3 than from the upper 1/3 has a negative discriminative. This is a detrimental item. It has to be either modified or discarded.

In this investigation the latter method was used. The scored book-lets were arranged in a descending order of the obtained scores. The response of the upper 1/3 for each item was found out. Similarly the response of each item in the lower 1/3 was found out. The result was arranged in the tabular form as given in Appendix G, with the help of these the difficulty value and the discriminating index were calculated.

Selection of Items for the Final Test.

After finding out the difficulty value and discriminating index of the test items, the next step is the selection of the items for the final test.

The determination of the optimum difficulty of the test item to be used in a standardized test is a problem of controversy. Some are of opinion that there must be roughly equal number of items at all levels from very easy to very difficult. Some other maintain that apart from a few easy and a few difficult items the majority of the items must be of 50 per cent difficulty level.¹

Writing in the Journal of Psychology P.K. Roy also

maintains that though in an ideal test each item must be of 50 per cent difficulty level it is not desirable that all items must be of this level. "It can be demonstrated statistically that an item passed by 50 per cent of a group discriminates between more pairs of persons than does an item passed by say 40 per cent or 60 per cent ........". "But it is not desirable, however, that items shall all appropriate the level of 50 per cent pass. We should try to include both easy and difficult items".¹

The common practice follows the latter suggestion of Sri P.K. Roy. Any item which is answered by all the pupils and any item which is not answered by all the pupils has no place for inclusion in a standard test.

On the above considerations 120 items were selected. Test items mostly lying between 20 per cent and 80 per cent difficulty level and having a discriminating index between - 21 and .7 were selected. However a few items of a higher difficulty level and a lower discriminating index were also selected so as to safeguard the curricular validity.

The following table gives the types of tests and the number of items selected in each type.

### Types of Tests and Items selected for the Final Test.

<table>
<thead>
<tr>
<th>Type</th>
<th>No. of items in the Try-out</th>
<th>No. of items selected for the final test</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 True false</td>
<td>50</td>
<td>20</td>
</tr>
<tr>
<td>2 Modified true false</td>
<td>30</td>
<td>10</td>
</tr>
<tr>
<td>3 Multiple choice</td>
<td>30</td>
<td>10</td>
</tr>
<tr>
<td>4 Completion</td>
<td>25</td>
<td>10</td>
</tr>
<tr>
<td>5 Matching</td>
<td>21</td>
<td>10</td>
</tr>
<tr>
<td>6 Classification</td>
<td>25</td>
<td>10</td>
</tr>
<tr>
<td>7 Analogies</td>
<td>18</td>
<td>10</td>
</tr>
<tr>
<td>8 Enumeration</td>
<td>31</td>
<td>20</td>
</tr>
<tr>
<td>9 Diagram</td>
<td>21</td>
<td>20</td>
</tr>
</tbody>
</table>

**Final Test form.**

120 items were selected on the basis of item analysis, satisfying curricular validity, difficulty value, and discriminative index and they were arranged in the ascending order of difficulty. The next step was to fix the time limit for answering this test. The time taken to answer the try-out was found out for each sub-test and the test as a whole. It worked out to be 85 minutes for 266 items. Lindquist and Hawkes and Mann suggest that "Try to adjust time allowance, except in a rate, or speed test so that at least 75 per cent of the pupils will have time at least the to consider all items in each section."

Rush is more liberal; he favours time limits "so that 90 per cent can attempt all items within their power." A time allowance of 45 minutes was fixed for the final test in accordance with the above two suggestions. The time fixed was purely meant for answering the test items after all the directions and instructions were given for each sub-test. Each item was subjected to the scrutiny of the science staff of the Teachers' College and the Practising High School at a series of meetings.

The English version of the test was translated into Kannada with the help of experienced science teachers. Two Kannada Pandits of the Practising High School scrutinised the Kannada version of the test with regards to spelling, grammar language and punctuation. Needed improvements and modification as per the suggestions of the Science teachers and language experts were effected before the items were drafted in the final form.

The Get up or the Format of the final Test.

According to Thorndike the following considerations must be kept in view in the get up of a test.

1) Legibility, (2) Convenience in taking the test, (3) Convenience in scoring the test, and (4) attractiveness.

The test was got printed legibly both in Kannada and

The important directions were got printed in the cover page of the book let. Instructions given at the beginning of each statement were got printed in bold types. The items in each sub-test were arranged in the order of difficulty value.

The test was got up in the form of a book let. Each type of test should have run over only one page. But due to the practical difficulties this could not be adhered to. Any how one and the same item was not allowed to run over two pages. To prevent fatigue and monotony the number of the items in the test was restricted to 120 and the working time of the test was 45 minutes. Except in the case of Enumeration and diagram tests, in all other cases arrangement was made to get the pupil responses only on the left side of each item. Two examples were worked out at the beginning of each test and the method of working was explained. With the help of these examples and explanation the pupils were in a positive position to answer the test comfortably. Thorndike feels that this kind of working out examples is "one of the most effective techniques for guaranteeing understanding of the test task, or for discovering and correcting misunderstanding, if it is present".\[1\] A scoring key was prepared so that the score may remain the same whoever scores the test. The key was pre-

---

pared in such a way that it fits in the position of the students responses on the printed test page. With the help of the key, the answers could be scored easily. The test answer booklets were got printed on a good paper. The test booklets gave an attractive look.