CHAPTER V

ANALYSIS OF THE DATA

The methods of collecting different kinds of data necessary for this study have been described in detail in the previous chapter. The purpose of this chapter is to describe the techniques of analysing the data thus obtained.

In order to prepare the ideal framework of Social Studies teaching the data collected had to be analysed before the rating scale was prepared. Before taking up the detailed analysis of the responses on the rating scales received from the groups of experts and teachers, it is pertinent to review, in brief, the analysis done previously. (For details see Chapter IV).

The vast literature produced by the experts - Indian as well as Western—was screened thoroughly and a compact and common list of major objectives was prepared. Similarly, following the same procedure, another list of common major objectives was prepared by screening the syllabi of the four Southern States. Having compared these two lists, the author was able to select three major objectives of Social Studies at the secondary level found common in both the lists.
Then, with the help of a committee composed of several experts in the field, the analysis of these three objectives in terms of knowledge, skills, attitudes and appreciations was done. Further, the analysis of these objectives with reference to expected outcomes, related topics, learning experiences and evaluation devices was completed. The expert sources were continuously consulted whenever the need was felt. In fact, many items of the prepared model were selected from such authoritative sources (For details see Chapter IV). The same expert committee helped to convert the ideal model into a rating scale which was thought to be the best available way of collecting data. Thus, the first objective of the study was achieved.

Now the task was to describe in detail how the data was tabulated, compiled and the statistical analysis done. As was mentioned before, the rating scales were sent to the samples of experts and teachers selected by the procedures described in the previous chapter.

The author started receiving the scales duly filled in almost after a week from the last date of despatching them. Two months' period was considered good enough to collect data which would suffice to help achieve the purpose of the study. This consideration seemed well based since by the 10th of December, 1966 the author received the following number of scales from both the groups (See Table 3).
TABLE 3 (a, b)

THE NUMBER OF RATING SCALES DESPATCHED AND RECEIVED FROM THE TEACHERS(a) AND THE EXPERTS(b)

(a) Teachers

<table>
<thead>
<tr>
<th>State</th>
<th>No. despatched</th>
<th>No. received Male</th>
<th>No. received Female</th>
<th>Total</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Andhra</td>
<td>200</td>
<td>43</td>
<td>5</td>
<td>48</td>
<td>24</td>
</tr>
<tr>
<td>Kerala</td>
<td>200</td>
<td>23</td>
<td>22</td>
<td>45</td>
<td>22.5</td>
</tr>
<tr>
<td>Madras</td>
<td>200</td>
<td>54</td>
<td>6</td>
<td>60</td>
<td>30</td>
</tr>
<tr>
<td>Mysore</td>
<td>200</td>
<td>64</td>
<td>17</td>
<td>81</td>
<td>40.5</td>
</tr>
<tr>
<td>Total</td>
<td>800</td>
<td>184</td>
<td>50</td>
<td>234</td>
<td>29.25</td>
</tr>
</tbody>
</table>

(b) Experts

No. despatched: 200 / 30%
No. received: 60 /

Analysis of the Responses on the scale

Since the same technique was used in analysing the responses of all the items on the scale, a general description of the procedure is given below.

The analysis of the data was a huge task since there were a large number of statements under each factor, i.e., expected outcomes, broad topics, learning experiences
and evaluation devices which had to be rated separately under different criteria. Thus the multiplicity of responses created a need for developing an elaborate statistical system of assigning scores and treating in a way which was useful to arrive at some meaningful conclusions. Before this system is discussed, it is pertinent to talk about the major design for testing the assumptions hypothesised in Chapter III.

Design

Any statistical analysis is the means to test the hypothetical predictions. It is the analysis which makes the raw data meaningful and helps to bring evidence either in support or against the hypothesis that are made. The need of each problem is different from the other. Most times when the experimenter is not interested in accepting or rejecting the hypothetical statement, comparison by simple percentages of groups can suffice. This method can be used for this research also. Since the attempt has already been made to form the hypotheses (See Chapter III), it is important that a suitable statistical technique is used to accept or reject these hypothetical predictions. Such technique is the technique of Chi-square. This technique is very useful in dealing with variables which are discrete. As well known, there are two kinds of variables considered in statistics--discrete and continuous. Most of the times, it is convenient to treat data on a continuous variable. This is true even when a variable is of discrete nature. But quite a few times, it is more convenient to deal with
data as if it were discrete. This is especially true when there is a large number of subjects involved. Another crucial point to be considered here is the underlying assumption of the distribution of reactions or responses of the subjects. The measurement on a continuous variable assumes that measured responses of a given characteristic will be normally distributed. In other words, when the obtained responses are plotted on a continuum or scale, the graph will take a bell-shape curve. No such assumption implies in the Chi-Square technique. This is the difference which is of the utmost help in this study since the assumption of normal distribution about either teachers' or experts' reactions cannot be made. Lastly, this technique provides an opportunity to compare observed frequencies with those which can be hypothetically predicted in a given set of data. Needless to point out that such statistics or tests prove that the results obtained from the data are beyond the influence of any chance factor. Keeping all these factors in mind the data was analysed.

The second major objective of this research was to know the degree of agreement of subjects, on the whole, with the model as well as with a group of items under each factor on the scale. This degree of agreement was obtained in terms of A (high agreement), B (moderate agreement) and C (low agreement). That means the responses were in three discrete categories of A, B or C which were mutually exclusive. In simple words, a response
of a subject was either, on any item circling A or B or C, but not in-between anywhere. To be specific, under the factor 'expected outcomes' under objective I, there were 12 items. Each one of them was rated by every subject. So there were 12 reactions of agreement - high, moderate or low. As the total sample of experts and teachers consisted of 60 and 234 respondents respectively, there were $60 \times 12 = 720$ reactions from experts and $234 \times 12 = 2808$ reactions from teachers to be dealt with. Similarly, under the factors 'broad topics,' 'learning experiences', and 'evaluation devices', the number of reactions was 14, 40 and 16 respectively for each subject. Again, the items in 'learning experiences' and 'evaluation devices' were rated on two more criteria, i.e., practicability and provided or not or used or not in addition to the criteria of suitability and desirability. This resulted in the multiplicity of responses as follows:

<table>
<thead>
<tr>
<th>Objective I</th>
<th>Expert's responses</th>
<th>Teachers' responses</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Factor to be analysed</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Broad Topics</td>
<td>$60 \times 14 = 840$</td>
<td>$234 \times 14 = 3276$</td>
</tr>
<tr>
<td>Learning Experiences:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Desirability</td>
<td>$60 \times 40 = 2400$</td>
<td>$234 \times 40 = 9360$</td>
</tr>
<tr>
<td>Practicability</td>
<td>$60 \times 40 = 2400$</td>
<td>$234 \times 40 = 9360$</td>
</tr>
<tr>
<td>Provided or not</td>
<td>$60 \times 40 = 2400$</td>
<td>$234 \times 40 = 9360$</td>
</tr>
<tr>
<td>Evaluation Devices:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Desirability</td>
<td>$60 \times 16 = 960$</td>
<td>$234 \times 16 = 3744$</td>
</tr>
<tr>
<td>Practicability</td>
<td>$60 \times 16 = 960$</td>
<td>$234 \times 16 = 3744$</td>
</tr>
<tr>
<td>Used or not</td>
<td>$60 \times 16 = 960$</td>
<td>$234 \times 16 = 3744$</td>
</tr>
</tbody>
</table>

It can be easily seen that the permutation and
combination of reactions under different factors must have become extremely complicated. For two major reasons this had to be reduced. One, it was not economical and meaningful to do analysis by single items. Secondly, the items under each factor could not be considered exclusive i.e., an assumption could not have been made that the reactions on items under each factor would be independent from a reaction to the preceding or following items. In addition to this, the stated hypotheses also limited the kind and direction of the statistical analysis. For example, it was necessary to derive support from the analysis for the understanding of the objective as a whole. This 'whole' was composed of different number of items interrelated with each other. Consequently, analysis of the responses by single items was considered fruitless.

Tabulation Form

Although item-by-item analysis was not needed, recording of the responses under different factors, on different criteria and items under them, was necessary. For this purpose a scoring sheet was devised. (See Appendix D). A quick glance at this form would give an idea of its usefulness for treating the multiple responses in several ways for accepting or rejecting the hypotheses. Another advantage of this form was that it could present a compact and complete picture of a person's total reaction to each objective.
The reaction of each subject was recorded on this sheet and frequencies of agreement on each were tabulated.

**Derivation of Scores**

The items under each factor were interrelated. For example, there were 12 reactions under expected outcomes of Objective I. The subjects' reactions on these 12 items were different, i.e., on one item it was A, second item C and so on. Although on each item A, B or C reactions were exclusive, responses on 12 items were pretty inclusive. The problem was that of arriving at some measure which could again help the author to categorise the S as exclusively in high, moderate or low agreement with the factors under Objective I as presented in the model.

Two different ways were appropriate in deriving a total category for each subject on each of the several factors on the scale. One, a standard technique used by most researchers, was to find out the mean of the whole group and then divide the subjects as high, moderate or low in terms of one or two standard deviations away from that mean. The first technique was obviously of no use to the investigator, as it assumes the normal probability distribution of responses. The second method was to prepare the ranges or intervals of the theoretical scores, group them into three equal step intervals and assign the highest interval as category A, the second highest as B.
and the lowest as C. It is described in detail as follows. The objective I has been taken here for illustration.

The first step was to arbitrarily assign values of 3, 2, and 1 to A, B and C ratings respectively. So, a high degree of agreement was assigned 3 points, moderate 2 points and low 1 point. There were 12 separate reactions on 12 outcomes under objective I. It was easy now to derive a theoretical range of scores. It was assumed that the probabilities of a subject circling A, B or C was equally the same. It was a statistical probability that one might circle either all As, or all B’s or all C's, if given to random selection. But, if the assumption of the author was correct, the subjects with understanding would more often than not circle A, sometimes circle B and rarely C. This would reverse with those subjects who would tend to show lower degree of understanding with the model. Since it was observed that reactions were different for 12 different items, the theoretical range had to be calculated as follows:

The highest range was derived by multiplying the number of items under each factory by three, e.g., \(12 \times 3 = 36\) (expected outcomes, objective I). This implied that a subject circled all A’s on all items given. Similarly, the lowest score was derived by multiplying twelve by one, e.g., \(12 \times 1 = 12\). The moderate score was calculated by multiplying the number of items by two, i.e., \(12 \times 2 = 24\) (objective I).
Then a standard step of finding step intervals of ungrouped data was followed. The range for the factor was found by subtracting the lowest value from the highest one, e.g., $36 - 12 = 24$. As the need was only for three class intervals, i.e., for categories A, B and C, this range was divided by 3 which gave 8 as a size of the interval. Using 8 as a size of the interval, the same range of scores was grouped again. The result for objective I (outcomes) was as follows:

**Theoretical range**

\[
\begin{align*}
12 \times 3 &= 36 & \text{Range} &= 36 - 12 &= 24 \\
12 \times 2 &= 24 & \text{Size Interval} &= \frac{24}{3} &= 8 \\
12 \times 1 &= 12
\end{align*}
\]

<table>
<thead>
<tr>
<th>No.</th>
<th>Step Interval</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>29 to 36</td>
<td>A (High Agreement)</td>
</tr>
<tr>
<td>2</td>
<td>21 to 28</td>
<td>B (Moderate Agreement)</td>
</tr>
<tr>
<td>3</td>
<td>12 to 20</td>
<td>C (Low Agreement)</td>
</tr>
</tbody>
</table>

A subject scoring between 29 and 36, was categorised as in A, one scoring between 21 and 28 as in B and that between 12 and 20 as in C. This procedure was followed for every factor (For further details, see Appendix E).

In this way, the individual subjects were categorised as in high, moderate or low agreement with all the items under each one of the factors, i.e., expected outcomes, broad topics, learning experiences and evaluation devices. In other words,
the total reactions of each individual experts and Social Studies teachers for various groups of items on various criteria under each factor were categorised as in A, B or C in accordance with their high, moderate or low agreement respectively in the model. This was done for the three objectives in the rating scale. The next problem was to tabulate the total frequencies of the samples of experts and teachers belonging to each category. This necessitated the devising of another scoring sheet (See Appendix F). Just as the Tabulation Form (refer Appendix D) was devised to secure a compact and complete picture of a subject's total reaction to each objective, a 'Final Scoring Sheet' was devised to record the total categories of both the groups of experts and Teachers.

The total categories of each subject for various factors were recorded and frequencies of agreement tabulated. The form used for this purpose has been given in Appendix. This form was the modified version of the Tabulation Form which was meant for one subject. The 'Final Scoring Sheet' was meant for the whole group. In the 'Tabulation Form' each row referred to a single item whereas in the latter it referred to a single expert or teacher. Each column in the 'Tabulation Form' referred to a certain number of items or a group of items under a factor while in the 'Final Scoring Sheet' it referred to the total number of subjects. In addition, the 'Final Scoring Sheet' included information such as rating scale number, sex and experience. The analysis of data for
sex and experience was necessary for the teachers' group and consequently these columns were added only in their forms.

In order to test the stated hypotheses, the data in terms of discrete variable was ready for the Chi-square tables.

**Chi-Square Test**

The Chi-Square values were obtained by calculating the hypothetical frequencies in a given number of categories and comparing them with the observed frequencies of subjects in the same categories. The hypothetical frequencies were obtained by dividing the total number of subjects by the number of categories. For example, in the present study, the total number of experts was 60 and the number of categories was 3. This gave 20 as the hypothetical frequency for each category. These hypothetical frequencies were compared with the observed frequencies, i.e., the number of experts falling in categories A, B or C. The Chi-square statistic gave information whether the observed frequencies were in line with the expected ones or they were different. If there was a variation, was it due to chance factor or not? If such variations were observed, were they small in number or as large as to make some significance beyond the factor of chance.

In this study, an assumption was made that the
subjects (experts and teachers) will be more often than not in agreement with the theoretical model. That meant that the number of frequencies in B would be larger than in C and in A would be larger than both in B and C. In other words, the number of observed frequencies would be the largest in category A and the smallest in category C. When such a trend was confirmed by the value obtained from the Chi-Square test, the hypothesis was considered highly confirmed. When a trend was found in the reverse direction, i.e., the number of observed frequencies were the largest in category C and the smallest in category A, the hypothesis was considered lowly confirmed. In case the largest number of frequencies was found in category B and smaller in both A and C, the hypothesis was considered moderately supported.

Level of Significance

Statistics are considered as estimates of the values of the parent population from which a sample is drawn. These values may be means, medians or just numbers (frequencies). The values obtained for drawn samples have errors. In other words, they are wrong. The occurrence of errors may be due to chance or extraneous factors. Every data has errors which may contribute to the differences of two groups of scores or frequencies. Statisticians want to reduce the errors which are the results of chance factors. How far the occurrence of an error has to be accepted in a given data or how much wrong a statistic
can be accepted? Various techniques are used to calculate statistical values which tell how much variation is due to the chance factor and whether it is tolerable for drawing valid conclusions. Such tolerance levels can be decided from the obtained value of Chi-square tests. It must be made clear here that these values only tell whether an error will occur a certain number of times (5 or 1) in a hundred cases. This tolerance reference is known as level of significance. In simple words, at what level an error in a certain data will be accepted. The selection of a level of significance depends upon how much a worker is willing to accept such variations or error in his data. Usually the two levels are accepted by the statisticians for accepting or rejecting a hypothesis. They are .05 and .01 levels. These are, of course, arbitrary standards of tolerance of an error in the data.

The results of analysis of the data are given in a number of tables in the following pages. These tables contain the Chi-Square values calculated from the original frequency tables and their significance at a certain probability level. The Original Chi-Square tables contained two observations: One, hypothetical frequencies and two, actual frequencies. The derivation of hypothetical frequencies in terms of the chance factor was easy. The total number in a group was divided by the number of categories, i.e., 3. For example, the total number of experts was 60. This divided by 3 gave 20 frequencies in each category. The
mathematical calculations gave the values which were compared with values given in the tables for Chi-Square that could be obtained only due to chance factor.

The number of such values was very large. There were three major objectives. Each one was analysed into four factors, i.e., expected outcomes, broad topics, learning experiences and evaluation devices. Again, the factors 3 & 4 were rated on three criteria when the former two were rated on a single criterion. Consequently, under each objective, in all, there were eight criteria. So there were eight Chi-Square values in each group—experts as well as teachers—for each objective. This gave, in total, 24 values for each group and 48 for the whole group. In other words, 24 sub-hypotheses were drawn from the major two hypotheses and tested both for the experts as well as teachers.

As the comparison between the responses of experts and teachers was necessary, the percentages of both the groups in each category were calculated and presented graphically (see Graphs 4.1 to 4.8, 5.1 to 5.8 and 6.1 to 6.8). In addition to this, the care was taken to see that apart from the chance factor some other extraneous factors might not have contaminated the data. The factors state, sex and experience of the teachers group were investigated to check their influence. The results and discussion follow in the next chapter.