CHAPTER IV

METHODS OF INVESTIGATION

This chapter describes the design of the study, various procedures implemented in the construction and finalisation of data gathering instruments to measure different variables which are used in the study, the methods adopted in the selection of sample, collection of data, mode of scoring and the different statistical techniques employed in the analysis of data.

4.1 THE DESIGN OF THE STUDY

The research method adopted by the investigator in the present investigation is the descriptive survey method, because it usually covers a large number of traits or characteristics of the group. The survey has undoubtedly been the most popular and the most widely used research method in education. Trow (1967) makes the following observations regarding survey research in general, "The errors and inadequacies of survey research in education appear at many points from the way problems are initially chosen and defined to the choice of the subject population, the selection of the sample, the design of the individual questions and the questionnaire as a whole, and the analysis of the resulting body of data"

The survey is an attempt to analyze, interpret and report the status of an institution, group, or area in order
to guide practice in the immediate future Tukeman (1978) expressed about survey research "a potentially useful technique in education as it is in public opinion polling and the social sciences. And the value of the survey as a means of gathering data is not to be denied."

Mouley (1969) also was the same opinion that "No category of educational research is more widely used than the surveying type. Educational surveys are particularly versatile and practical in that they identify present conditions and point to present needs. Descriptive surveys are oriented towards the description of the present status of a given phenomenon."

Survey studies are mainly of the "what exists" type; that is to say, they are designed to determine the nature of an existing state of affairs. They may be considered to be research in that they result in the accumulation of a certain type of knowledge, although their scientific status may be questioned, because it does not aspire to develop an organised body of scientific laws but provides information useful in the solution of local problems.

The survey is very useful in doing what it is designed to do, that is, in getting descriptive data which people can provide from their own experiences. The survey is often the only means through which opinions, attitudes, suggestions for improvement of instruction and other such data can be obtained.
Information can be obtained in either of two ways in the survey method. It can be obtained personally, through face-to-face contact or through some kind of paper-and-pencil instrument. The personal approach is handled through the interview and the paper-and-pencil approach is handled through the questionnaire. The questionnaire is probably the most used and the most abused survey instrument.

4.2 MEASUREMENT OF THE VARIABLES

Out of 37 variables (both dependent and independent) included in the study 8 are personal and demographic variables, for which information was gathered through a personal data sheet. The remaining 29 variables were measured with suitable instruments. Among these instruments an attitude scale towards teaching profession, and a rating scale to measure teaching effectiveness were developed by the investigator. Teaching Aptitude Test (TAT) developed by Jai Prakash and Srivastava was translated and adopted for the purpose of measuring TAT characteristics of science teachers. 16 Personality Factor Questionnaire developed by Cattell was translated and adopted for the purpose of measuring personality of science teachers. The different techniques and procedures followed in the development of the instruments are discussed below.

4.2.1 Teaching Effectiveness : A Controversy Regarding Criteria

On the question of utilizing value judgements for studying effectiveness, Robinowitz and Travers (1953)
pointed out that "there is no way to discover the characteristics which distinguish effective and ineffective teachers unless one had made or is prepared to make a value judgement. The effective teacher does not exist pure and serene, available for scientific scrutiny, but is instead of fiction in the mind of men. No teacher is more effective than another except as some one so decides and designates." Hence, various approaches as well as techniques came into existence to measure teaching effectiveness. To have a glance of different methods used in the past research on teaching effectiveness, a few studies are referred to below.

Anderson (1954) employed eight different types of criteria of teaching effectiveness on 30 teachers and found that their standing on the eight criteria were correlated moderately. On the basis of the results, the author suggested that the evaluation of teachers was to some extent influenced by the choice of the criterion in general. No adequate basis for validation of teacher evaluation has existed at present. Johnson Jr. (1955) reported that there were three primary approaches to measure teaching effectiveness, viz., (1) evaluation of qualities assumed to function in the teaching act (presage variables), (2) appraisal of teaching activity (process variables), and evaluation of pupils growth (product variable).

Knox (1956) obtained a composite score of teaching success based on (1) acceptability ratings by principals and superintendents, (2) a short form of Wisconsin adoption of
the M-Blank used to obtained principals estimate of the teacher's performance, (3) peer ratings and (4) the judgement of a supervisory official of the state department of public instruction.

Beecher (1961) summarised the different types of criteria of effectiveness and the data-gathering devices employed in Wisconsin studies. The number of studies used in different types of criteria and devices were as follows: (1) In-service ratings by Superintendents-9, principals-4, other supervisory officials-5, teacher educators-20, departmental personnel in the areas of specialisation-4, state departmental personnel-11, and self-rating-11; (2) peer ratings-6, (3) pupil gain scores-70; (4) Pupil rating-16; (5) composite of test scores from tests thought to measure teaching effectiveness-14; (6) practice teaching grades-13; (7) combination or composites of some or all of the above criteria-83.

Recent researchers have taken pupil gain or pupil growth as a criterion of teacher competence. Burkhart (1969), while reviewing studies, has shown that there are twenty seven studies which make use of the pupils growth as the criterion of effective teaching. But Cohen and Trent (1973), have shown methodological difficulties inherent in measuring pupil gain. McCall (1952) in his year statewide study of North Carolina, attempted to find measures usable in merit ratings. The criterion of teacher merit was the proven ability to produce educational growth in pupils as
measured by a wide variety of tests. Hartley and Hogan (1968), who compared self-development ratings of students with their evaluations of course work found small relations between the response patterns to the respective scales.

Many problems are identified in developing a criterion based upon pupil growth viz., (a) which pupil growth shall be measured? (b) how can effects other than teacher effects be controlled? (c) how shall grade differences be controlled? (d) how shall school effects be controlled? (e) how can effects based on pupil characteristics be controlled? (f) how should the reliability of the teacher effectiveness index be estimated? etc. Similarly the ratings provided by different agencies on the basis of close observation also have many deficiencies and the methodological observations made by specialists to predict the actual behaviour pattern of an effective or ineffective teacher too suffers from many difficulties.

4.2.1.1 Why is rating method preferred to?

It is clear from the existing literature that the problem of measurement of teaching effectiveness may be approached in a number of ways. One way with which many studies were carried out, is trying to find out to what extent a teacher possesses the various personal, professional and social qualities which make him a successful teacher. But this alone would not suffice, since mere possession of these qualities is no guarantee that a particular teacher is successful in this work, one should
necessarily study the other factors, too, that go to contribute to teaching effectiveness and this is not an easy task, for a list of these factors is truly a formidable one.

The total personality development of the pupil may also be studied and measured, but how can it be said that it is just the result of the work of a single teacher or even a group of teachers? As explained earlier, a variety of factors pertaining to the pupil's homes and environment would influence their growth.

Observation of a given number of lessons by a single person or even a team of observers with the help of a rating scale, however perfect this instrument may be, would not ensure accurate measurement. The very presence of the observers in the classroom would tend to influence the teaching of the teachers concerned and also there is no guarantee that throughout the year their teaching would be of the same standard. It is quite possible that really unsuccessful teacher may just impress his observers in a few lessons when they are presented to him and may consequently get classified as successful or average and the opposite is also possible, i.e., successful teachers may get classified as unsuccessful or average on the basis of a few lessons observed during the course of an investigation.

In view of the practical difficulties anticipated in the adoption of the above mentioned procedures, it was
decided that the rating of the subjects on the degree or level of teaching effectiveness achieved by them should be done not on the basis of a few lessons observed by him or even a team of persons that might have been enlisted to assist him, or by considering the pupil growth, but by the heads of the schools concerned that would be quite conservant with teachers over a long period, and in addition, by their students.

Apart from the above reason, the most frequent method of measuring teacher effectiveness was the rating (Gage, 1963). A few studies which have employed the rating method are referred to in the following paragraphs.

Domas and Tideman (1970) in their 1006 item annotated bibliography on teacher competence reported the wide-spread use of rating methods in research on teaching. But Medley and Mitzel (1959) reviewed studies of the relations between ratings and other criterion measures of teacher effectiveness and found consistently no relation between ratings of teacher effectiveness and pupil learning in the classroom. The only significant result found was that between the ratings and emotional climate in the classroom.

Remmers (1963) pointed out that in general, teachers and their professional representatives in the National Education Association were strongly opposed the use of ratings as there was no generally acceptable method of measuring teacher's merit. Rating scales, however, popular, carry serious limitations as evaluation, feed back and
grading tools. Their appearance of objectivity is an illusion, expressed by Verble (1979). Still there were many situations where ratings proved more reliable than other methods. For example, Druker and Remmers (1950) compared ratings given by alumni 10 years after graduation with campus students on the relative importance of 10 teacher characteristics and found that the rank correlation coefficient \( r = 0.92 \) was very high. Remmers (1960) pointed that if 25 or more students' ratings were averaged, they were as reliable as the better educational and mental tests available at present. The above fact is expressed by Noerrlinger (1987), that the student ratings of their teachers significantly correlated with the ratings of their principals on more effective and less effective teachers.

In view of the above evidences, the investigator preferred a rating scale to measure teaching effectiveness in the present study.

4.2.1.2 Preliminary study to identify the characteristics

The various characteristics which influence the teaching effectiveness were gathered from thorough review of research. To supplement this list of teacher characteristics, 30 experienced science teachers and 15 headmasters and some students were asked to write down some other characteristics which are most important for an effective science teacher. Their responses were sorted out and added to the list of characteristics collected from
review of research. All the items were rewritten avoiding ambiguity and overlapping. A preliminary form containing 36 characteristics thus prepared is given in Appendix-A1.

The preliminary instrument was administered to a sample of 75 science teachers of both sexes, requesting them to select 10 most important characteristics which influence the science teachers' effectiveness. They were also instructed to rank the characteristics 1, 2, 3,...10 according to their importance in contributing to teaching effectiveness. Their responses were tabulated and analysed. On the basis of frequency, 10 characteristics were picked up as the most important. On the basis of both frequency and weightage of ranking, again the top 10 characteristics were selected. In both procedures, the top 10 characteristics were the same, except a minor variation in their order. Thus the most significant 10 characteristics which are suitable to science teachers were identified for the purpose of developing a rating scale to measure the teaching effectiveness of science teachers.

4.2.1.3 Description of graphic rating scale

The graphic type of rating scale is probably the most popular and the mostly widely used. It is nothing but a kind of a numerical rating scale and belongs to the non-analytical type. In this type of rating scale, each trait is represented by a segmented line. One end indicates one extreme of the trait to be rated; the other end represents the other extreme. The rater places a check mark at an
appropriate place on the line to represent the degree to which the subject possesses the trait, while describing the different techniques of rating scales, Remmers (1963), explained that the graphic rating scales should provide a continuous straight line with cues or categories along the line to guide the rater.

4.2.1.4 Why is the graphic rating scale preferred to?

The virtues of graphic rating scales are many; their faults are relatively few. They are simple and easily administered. As pointed out by Symonds (1932), the graphic rating method was probably the most serviceable and widely adopted method. In the graphic rating method, the continuity is maintained and while the respondent is asked to rate, he is free to place a mark of indication of the subject's level of proficiency in the particular trait on which the rating is asked. Even a minor variation can be obtained by using graphic rating method, which is not possible in other types of rating methods. They are interesting and require little added motivation. They are quickly filled out and they do not require the rater to bother with numbers. These features the rater finds attractive. No doubt, the scoring of graphic ratings is a little bit tedious but there is a possibility of obtaining wider discriminations both among raters and ratees by adopting graphic rating method. Keeping all the above mentioned points in mind, a graphic rating scale was constructed to measure the teaching effectiveness of science teachers in secondary schools.
4.2.1.5 Pilot form of the rating scale

The pilot form of rating scale consisting 11 sub-scales with one for each of the 10 traits or characteristics identified as the most important and one for overall efficiency of the science teacher.

A pilot form of the graphic rating scale was developed with eleven characteristics, eleven segmented straight lines of 10 cms length each. One centi-metre gap was maintained between any two consecutive segmented straight lines. Each segmented straight line was meant for each characteristic or trait. Every page consisting of eleven segmented straight lines which were meant for eleven characteristics or traits and each straight line was divided into ten segments - 0, 1, 2, ..... 10. The raters were given clear instructions about how to rate. The researcher demonstrated with the help of an example, the procedure of rating to the respondents before they were asked to rate. This form of rating scale along with instructions is given in Appendix-A2.

Pilot Study: A pilot study was conducted on a sample of 100 science teachers of both sexes working in 43 schools under different managements selected at random. All the forty three heads of the institutions were asked to rate their teachers in the manner explained to them personally. The same rating scale was given to five tenth class students who were taught by the same science teacher. Thus each science teacher was rated by five tenth class students and
the head of the institution. The students of 10th class alone were considered for obtaining student ratings because they might have a chance of being his student for a longer period than the students of lower classes. The rating scale was administered to students by demonstrating the procedures of rating step by step with suitable illustrations following the same pattern of demonstration throughout the schools. The investigator was very keen in obtaining accurate ratings from the students.

4.2.1.6 Scoring of the scale

For this purpose an ordinary ruler was used. Numerical values were assigned to each teacher's standing in respect of each sub trait, depending upon the distance of the cross-mark, made by the rater from the base line. As the length of the segmented straight line was 10 cms or 100 mm, if the cross mark was at a distance of 8.5 cm from base line, the score for the individual concerned was 85. The scores thus arrived at for each individual teacher for all the ten sub-traita were added and then averaged to get a single score (termed as 'R₅') for each subject. Then the reading on the overall teaching effectiveness (R₂) was noted and added to R₅ and then divided by two to get a score of teaching effectiveness (R₃) for that teacher as rated by one particular student rater. Thus, a teacher would obtain five effective scores as rated by five students. The average student rating (R₄) was obtained from the average of five R₃ scores. Similarly R₅ was obtained from the ratings given by
H.M./Principal. The final effectiveness score (TE) which was termed as a "composite index" of teaching effectiveness was obtained by averaging $R_4$ and $R_5$.

4.2.1.7 Validity of the scale

Dowell and Neal (1982) stated that a large number of studies have attempted to link student ratings to student learning in order to validate student rating as a measure of teaching ability. Despite the large number of studies carried out, few were interpretable because of three methodological issues largely neglected in the literature reviewed. They were: (1) section means, (2) controls for student ability, and (3) criterion inadequacy with all the limitations. The scores (composite index) of teaching efficiency obtained from the graphic rating scale were correlated with the results obtained by the students of corresponding teachers in their subjects concerned at both internal and external examinations. The percentage of passes and the number of marks obtained by the students were considered as an external criterion in the teachers' subject concerned to see whether more efficient teachers were able to produce better results in the public examination at the end of S.S.C.. It was found that the correlation coefficients were significant at 0.01 level. Thus the rating scale was valid on the basis of student gain or achievement as an external criterion.
There are various methods of estimating the validity of measuring the instrument. The following types of validity were established for the rating scale that was developed for this investigation.

(a) Intra-rater consistency of the raters

Each ratee was rated by five students and a headmaster and ten main characteristics which effects teacher's efficiency.

At the end of the rating scale an overall teaching efficiency item was given in which the raters rate the overall efficiency of the science teacher was asked to see whether the raters are having consistency within themselves. Correlation was established between the overall rating and the average rating on 10 traits for all 500 students raters and 43 headmasters. It was found that the coefficient was in the range of 0.32 to 0.91. Therefore, it was concluded that the instrument was possessing intra-rater consistency of the raters.

(b) Inter-rater consistency

Another measure of validity that was established for this scale was inter-rater consistency. The average of five student ratings were considered on one side and the headmaster ratings on the other side for the group of science teachers. The correlation was established between these two types of raters and found very high \((r = 0.72)\)
Thus the inter-rater consistency of the graphic rating scale was statistically significant and hence valid.

4.2.1.8 Reliability

It is true that the reliability of ratings, in general, is found to be variable and disappointingly low, and as Symonds (1932) remarks, "so low to cast grave doubts on the value of rating as a method for getting trust-worthy data". But, it should be, at the same time, pointed out that the reliability may be increased by taking a number of precautions. Firstly, as many raters as possible who are quite familiar with the subjects who are to be rated should be involved in the investigation. Authorities in this field are of the view that "human character can be rated accurately enough for practical purposes in education only where the rating is the average of three independent ratings" (Symonds, 1932). Secondly, the trait to be rated, should split into a number of sub-traits each of which is to be rated separately. To quote Symonds (1932) again "indeed if a trait is analysed into sub-traits which have been found by analysis to correlate with the trait, and ratings on the sub-traits by an individual are combined into one rating, the results may have a reliability so much higher that is only slightly under what might be predicted by the Spearman Brown Prophecy formula from a single rating". As pointed out earlier, the above two precautions were implemented in the present investigation.
The reliability was established using the test-retest method. After 3 months, the same scale was readministered to the same sample of 100 science teachers (sample for pilot study) of both sexes and the correlation coefficient between the first and the second sets of data was computed. The correlation thus found was 0.93 which was high. Therefore the instrument is considered to have high reliability.

4.2.2 Construction of Attitude Scale

To measure the attitudes of science teachers towards teaching profession a Likert (1932) type of summated rating technique was employed.

(i) Likert type of summated rating technique scale:

Compared to either the Thurstone or the Remmer's scales, the Likert type scales are fairly easy to construct. If too, begins with a considerable number of statements that reflect favourable and unfavourable attitudes about an attitude object. Then the subjects are asked to respond to them on a five point scale - SA (strongly agree)\'; A (Agree); U (Undecided); D (Disagree), SD (Strongly Disagree). The responses have weights of 5, 4, 3, 2, and 1 for favourable statements, and 1, 2, 3, 4, and 5 for unfavourable statements. The subjects' score is the total of the values indicated. A high score indicates a high favourable attitude, a low score indicates a low favourable attitude.

Likert type scales can be constructed in a relatively short time, require no judges, and can be scored
rapidly. The results obtained with this type of scale as far as reliability and validity are concerned, are quite comparable to those obtained by Thurstone and Remmers.

(ii) Preliminary form of attitude scale:

Bearing in mind the important dimensions of teaching like students, teacher's status, merits and demerits of the teaching profession, etc. a number of items related to attitude of teachers towards teaching profession were collected from a number of sources. The item pool thus prepared was placed before 10 selected experienced high school science teachers who are asked to suggest changes, if any, to avoid ambiguity, repetition, etc., among the items. The suggestions made by them were incorporated and 81 statements were selected. Each of the items was arranged on a five point scale with the following alternatives: Strongly agree (SA); Agree (A); Undecided (UD); Disagree (D), and Strongly Disagree (SD). Thus a preliminary form of the questionnaire was prepared, using which the pilot study was conducted. To avoid taking of responses, the questionnaire was made anonymous.

Pilot Form of Attitude Scale:

The pilot form of attitude scale consisting of 81 statements which are collected from various sources and refined with the help of experts. This attitude scale in English version was translated with the help of some experienced high school headmasters and university research
scholars. The translated version was given to university lecturers those who were well-versed with both Telugu and English languages for review and suggestions. Their suggestions were incorporated and the questionnaire was given to 5 high school teachers and 3 research scholars to see whether they could understand the statements. Such of those words, the meaning of which was not clear for them, were changed and finally Telugu form was prepared. Both English and Telugu versions were given under Appendices-B₁ and B₂.

Pilot Study

The pilot study was conducted on 100 science teachers of both sexes, selected by a random sampling procedure from 43 high schools belonging to Chittoor district in Andhra Pradesh. The sample includes only the L.T. assistants or B.Ed. assistants. The pilot form of the attitude scale which is self-constructing was administered to the sample of science teachers individually in their respective schools during leisure hours.

Even though the scale is self-administering as a matter of motivating the science teachers, they were explained the purpose of the research, and the way they had to answer the items. They were also assured that the data would be used only for research work and would be kept confidential.
For the purpose of scoring, numerical weights were assigned each of the 5 categories of responses, from 1 to 5 in the case of negative items and 5 to 1 in the case of positive items as suggested by Likert (1932).

Table 1: Numerical Weights given to the five alternative responses

<table>
<thead>
<tr>
<th>Type of statement</th>
<th>Strongly agree</th>
<th>Agree</th>
<th>Undecided</th>
<th>Disagree</th>
<th>Strongly disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Negative</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

Item Analysis

Item analysis of the responses given by the teachers was carried out by the method of criterion of internal consistency suggested by Likert (1932). This method was used because the results obtained by this procedure agree very well with the results of the traditional method of item analysis and at the same time it is less labourious than the latter, and is advocated by many investigators like Likert (1932), Murphy and Likert (1937), and Edwards (1969).

This method consists of rank ordering the subjects with respect to their total scores and taking the highest scoring 27 per cent and lowest scoring 27 per cent of subjects and calculating the discrimination value of each statement using the formula of Kelly (1939).
Discrimination value of any statement = The Difference between the mean scores of the high and low groups on the statement.

Apart from the discrimination index the 't' values are also calculated for each item given in Appendix-B₃.

Selection of the Items for final form:

Those items whose discrimination value is greater than 0.30 and whose 't' value is greater than 1.96 were selected to be included in the final form of the attitude scale. Thus a total number of 52 items were included in the final form of the scale. The Telugu and English versions of the final forms are shown in Appendices-B₄ and B₅.

Validity:

There are different types of validity viz., content validity, intrinsic validity, criterion validity, predictive validity, factorial validity, concurrent validity, item validity and construct validity. But for the present tool, content, item, criterion and intrinsic validities are established.

Content Validity: This form of validity is estimated by evaluating the relevance of the test items individually and as a whole. Each item should be a sampling of that aspect which the test purports to measure and taken collectively the items that should constitute a representative sample of the variable that is measured.
In the construction of the attitude scale for the present investigation the items were collected from various sources. The preliminary form of questionnaire was given to many senior teachers and university lecturers to see that the items cover all the aspects of teaching profession. Therefore it is believed that the instrument possesses content validity.

Item Validity: There are numerous procedures in which the item validity can be determined one of which stresses the number of discriminations of the desired sort that the item is capable of making. The discriminative power of each item was established and it was tested for significance at 0.01 level, before including them in the final form. Thus each item of the attitude scale was valid which ensured the validity of the attitude scale itself.

Criterion Validity At the end of the attitude scale a specific question was asked: "If you are given the same salary and other facilities equal to teaching profession do you prefer any other profession or teaching profession itself". The mean scores of those who favoured to take teaching profession and those who wanted to take other profession were tested for significance.
Table-2: Means and SDs of science teachers who favoured teaching profession and those who did not and the results of 't' test.

<table>
<thead>
<tr>
<th>Number (N)</th>
<th>Means (M)</th>
<th>SDs</th>
<th>'t' value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Science teachers who favoured teaching profession</td>
<td>72</td>
<td>268.00</td>
<td>17.24</td>
</tr>
<tr>
<td>Science teachers who did not favour teaching profession</td>
<td>28</td>
<td>234.00</td>
<td>13.41</td>
</tr>
</tbody>
</table>

Note @ indicates not significant
* indicates significant at 0.05 level
** indicates significant at 0.01 level

These notations are followed throughout the report.

From the above table, it is seen that the 't' value is significant at 0.01 level. Therefore, it could be concluded that the scale is measuring the attitudes of science teachers towards teaching profession validly.

Intrinsic Validity Guilford (1954) defines "intrinsic validity as the degree to which a test measures what it measures". This can also be stated in terms of how well the obtained scores measure the tests true score components. This validity is given by the square root of the proportion of true variance i.e., the square root of its reliability.

The intrinsic validity of the scale was $\sqrt{0.8166} = 0.9036$. 
Reliability:

The reliability of the questionnaire was measured by the split-half method. The attitude questionnaire was split into two equal halves and the scores on the odd and even items of 100 subjects were calculated using Karel Pearson's formula for product movement correlation, this gave, the reliability of half test. The reliability of the half-test was 0.69. This was corrected for full length of the test by Spearman Brown Prophecy formula. The reliability of the full test thus obtained was 0.8166.

4.2.3 Description and Adoption of Teaching Aptitude Test

(i) Description of the Teaching Aptitude Test

To measure teaching aptitude of science teachers a test developed by Jayaprakash and Srivastava is adopted in the present investigation.

In the development of the test, the job psycholgraphic method of Otis and Smith was followed for the purpose of collecting and analysing information on various aspects of teaching profession. This yielded a list of abilities and traits which are considered important in the teaching job. The traits mentioned by Witty (1947), Barr (1948), Adaval (1952), and Menon (1949) were also considered. Finally, twenty common traits, which were essential for the successful teacher, were selected for the first experimental draft. The findings of foreign and Indian
researchers were also considered, and due attention was paid to the comments of some experienced teachers on the selection of the traits. These twenty traits were rated by a group of twenty educationists on a five-point scale - Highly agreeable (HA), Agreeable (A), Indifferent (I), Disagreeable (D), and Highly Disagreeable (HD). Their ratings were tabulated and rankings made out of twenty, ten higher ones were retained and the other ten, the lower ones, were deleted. The ten finally selected traits in order of their frequency were: (i) Moral character, (ii) Discipline, (iii) Patience, (iv) Cooperative attitude, (v) Kindliness, (vi) Impartiality, (vii) Scholarly taste, (viii) Enthusiasm, (ix) Optimism, and (x) Wide interest. Initially, 240 Likert type items were pooled and were reviewed by 20 sophisticated judges. After reviewing the item pool, the 40 items were deleted and 200 items retained in the test-form. This form was tried out on a sample of 630 undergraduate pupil teachers of six basic teacher's training institutions of Vindya region of Madhya Pradesh. Their age range was 20 to 30 years and they belonged to rural areas and were of more or less of same socio-economic group. As a result of item-analysis, out of 200 items, only 150 items, which were highly scored, were taken out for the final draft. The rest of 50 items, which were poorly scored, were dropped from the test.

Weightage to Items: For every item in the test, the weights were determined by a variance and its covariance with other items. In this test, the weights of +3, +2, +1 were assigned to the right responses of 'HA', 'A' and 'I' or 'HD', 'D' and
'I', respectively. In the same way, -3, -2, and -1 were given to the wrong responses of 'HA', 'A' and 'I' or 'HD', 'D' and 'I' respectively. After this, a scoring formula (Lindquist, 1961) rights minus wrongs ($S = R - W$) was adopted to obtain the correct scores. The coefficient of correlation between the raw-scores of the test and the rated-scores was found to be +0.579 by product moment method and this justified the given weightages to a greater extent.

Standardisation Sample: The final test consisting of 150 statements was given to large sample of 1050 pupil teachers, under training in eleven government teachers' training institutions of Madhya Pradesh, spread over in some eight districts. These teacher's training institutions were divided into three grades - above average, average and below average - on the basis of their results in public and home examinations, and from each of these grades 336, 300 and 414 teachers were drawn, respectively. The raw scores indicated that we were dealing with a normally distributed population. The frequencies revealed the scores ranging from 51 (lowest) to 400 (highest). The highest frequency was 201 which belonged to the class interval of 201-250 scores, nearly the middle step of the distribution. As a result, the mean being 212 (with a standard deviation of 61.6), the median = 213.6 and the mode - 217. The $SEm$ (1.9) was also very small. The skewness obtained was only -0.09; and the kurtosis being 0.253, which were all insignificant. Such a nice normal distribution induces to expect that the norms of this test
would be applicable to a wider universe of teachers and the test can safely be used to assess the teaching aptitudes of all sorts of teachers working in the junior and senior higher secondary schools even the very superior and the defective ones.

The validity of test was secured by computing a coefficient of correlation between scores on the test and the assessment marks obtained in the final examinations. The coefficient of correlation between the total marks of theory, practice teaching and craft, and the test score on 200 pupil teachers, was +.5. The obtained validity coefficient is quite satisfactory.

The reliability of the test was calculated by split half method using Guttaman and Spearman Brown Prophecy formulas which yielded the coefficient of correlations as +.891 and +.91, respectively, on a sample of 100 cases. The test-retest method on a group of 50 teachers yielded a correlation of +94. All these coefficients are high and therefore the test has a good reliability. The test has also a higher degree of internal consistency and all the ten traits measure independently their respective qualities irrespective of a moderate overlapping among four traits. The inter-element correlations between the 10 components mostly centre around +.5. Hardly any one of them is less than +.5 and none more than +.547.
(ii) Adoption of the Instrument:

The instrument was in Hindi version. With the help of the two research scholars and a lecturer of Hindi Department of S.V.University it was translated into Telugu and English versions. The copies of both English and Telugu versions were given to language experts and two staff members of English and Telugu Departments for their comments on translation. Their suggestions were carried out and the final form in English and Telugu versions was prepared. The forms are shown in Appendices \(- C_1 \text{ and } C_2 \).

(iii) Scoring of the test:

As may be noted, the test has two sets of scoring keys: One, for the right (R) and the other for Wrong (W) score. Accordingly, the weights of 3, 2, 1, 2, 3 are given to right or wrong responses of HA, A, I, D, and HD respectively, which is correct as is visible through the blank circles of the right and wrong keys. Thus, the scoring keys give two sets of scores separately, i.e., right and wrong scores. The correct score is obtained by subtraction the wrong scores from the right scores (R-W) and this remainder is an individual's raw score. The individual raw scores of all the 10 sub-tests were added that will give the total aptitude test score. The total aptitude score and the 10 sub-tests scores were numerically obtained separately to suit the computer analysis.
4.2.4 Description of Adoption of 16 PF Questionnaire

The investigator searched for the theories of personality and the means of measuring it which account for the totality of behaviour. Cattell's theory, of all the numerous theories, is the only theory based on the principle of totality of behaviour of the individual. According to Cattell (1950) "Personality is that which permits a prediction of what a person will do in a given situation". His theory is based on personality sphere concept (Cattell, 1946, 1957, 1964) a design to ensure initial item coverage for all the behaviour that commonly enters rating and the dictionary description of personality. It focussed heavily on source traits. Cattell defines source traits as the spring of human behaviour. Much is becoming human about the nature of these dimensions through studies with ratings, with laboratory measures, and with real life situations. According to him, a trait of any variety is a mental structure which is relatively fixed characteristic of the individual, functioning from time to time in behaviour.

(1) Selection of Tool and its Uses:

It is obvious that selection of a tool for measuring personality poses serious problems. In this connection it may also be noted that the problem of justification of the choice looms large. One may cut a sorry figure in explaining for choice. This problem can be solved if we study the theory behind a particular tool and the rationale with which it was prepared. The selection of
Cattell's 16 PF test in the present research was also not arbitrary and has been made after a lot of deliberations and study of theory which has been supported amply by Stern (1921) and Allport (1937) in the following works.

Stern observes as under "We have the right and obligation to develop a concept of trait as a definitive doctrine, for in all activity of the person, there is besides a variable portion, like wise a constant purposive, portion, and this latter we isolate as the concept of trait."

Allport's contention is equally forceful. He asserts that: "traits are discovered not by deductive reasoning, not by faith, not by meaning, and are themselves never directly observed. They are discovered only through an inference made necessary by the demonstrable consistency of the separate observable acts of behaviour."

Vernon (1963) says that a person's behaviour in any situation depends, of course, on specific features of that situation and on his temporary feelings or state of mind, but it depends also on his more enduring characteristics, abilities, habits and more general dispositions which may be called traits.

Cattell (1961), says that the source traits, as measured by the 16 PF test, are the spring of human behaviour. He defines personality as "that which permits a prediction of what the person will do in a given situation"
(Cattell, 1950). This definition is consistent with the contention of Marxi and Hillix (1973), that "the theory of personality is really identical with general theory of behaviour, for Cattell's definition would fit theories of behaviours".

The above theoretical as well as practical considerations Cattell's 16 PF questionnaire was selected. The 16-PF questionnaire is an objectively acceptable test devised by basic research in psychology to give the most complete coverage of personality possible in a brief time. Coverage of personality is insured by the sixteen functionally independent and psychologically meaningful dimensions isolated by over twenty years of factor analytical research on normal and clinical groups. Therefore, having a certain position on one does not prevent the persons having any position whatever on any other. Thus, each of the sixteen scales brings an entirely new piece of information about the person a condition not found in many alleged multi-dimensional scales.

Moreover, a scoring system is provided whereby, from the sixteen factors, one can extract and work with only four broader (and less specific) traits - anxiety, extraversion, alert poise and independence. Experience with the 16 PF in clinical, educational, and industrial psychology shows that the use of the 16 traits gives actual prediction.
Adoption of the Instrument and Reliability

Form-C of the 16 PF questionnaire was adopted for the purpose of the present study. The instrument was translated into Telugu. The copies of both English and Telugu versions were given to language experts and two staff members of psychology department of the university for their comments on translation. Their suggestions were carried out and the final form in Telugu version was prepared. The forms are shown in Appendices $D_1$ and $D_2$. The 16 PF questionnaire, as its name implies, provides the multi-dimensional measurement of personality through normative scores on 16 bi-polar factors. The 20 years of factor analytical work sponsoring this instrument lends empirical credence to its construct validity as a personality measure and predictor. The test-retest (four weeks interval between the two administrations) measures of reliability coefficients for the 16 Factors were as follows: Factor $A = 0.42$, $B = 0.39$; $C = 0.36$; $E = 0.34$, $F = 0.41$, $G = 0.32$; $H = 0.49$; $I = 0.42$; $L = 0.37$; $M = 0.31$; $N = 0.32$, $O = 0.29$, $Q_1 = 0.52$; $Q_2 = 0.34$; $Q_3 = 0.37$ and $Q_4 = 0.40$.

4.2.5 Development of Personal Data Sheet

The information with regard to sex, experience, marital status, qualifications, caste, family size, management of the school, working place etc., of the science teachers were obtained from a carefully worded personal data sheet. Both Telugu and English versions of the sheet are given in the Appendix.
4.3 SELECTION OF THE SAMPLE

The problem is aimed at measuring teaching effectiveness of science teachers working in secondary schools in Chittoor district of Andhra Pradesh State, India. Among the secondary school science teachers those who are handling 8th, 9th and 10th classes are included in the study. Therefore the population of the study constitutes all those who teach science to the higher classes in all the secondary educational institutions in Chittoor district.

As the sample for the study has to be selected from a large population spread all over the district, a two stage stratified random sampling technique was adopted. The locality and management to the school were considered for the purpose of stratification.

At the first stage the list of secondary schools in the Chittoor district was obtained from the District Educational Officer. The total number of schools were then divided into two groups as rural and urban schools. Again the stratification was made on the basis of management viz., Government, Local bodies and Private. From each stratum a representative sample of schools were selected at random. The total number of schools thus selected were 106. In the second stage all the science teachers handling higher classes (classes VIII to X) from all the selected schools were included in the study by adopting cluster sampling technique. Thus a total of 200 subjects working in 106
schools was selected by a two-stage stratified random sampling technique.

For obtaining ratings on teaching effectiveness of 200 science teachers, the supplementary samples viz., all 106 heads of institutions (junior college principals, headmasters and headmistresses) and 5 to 25 students of 10th class in each one of the respective schools were selected at random.

Table-3 : The detailed sample frame is given below.

<table>
<thead>
<tr>
<th>1. Locality</th>
<th>Rural</th>
<th>Urban</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. Management</td>
<td>Govt. Local Bodies</td>
<td>Pri-</td>
<td>Govt. Local Bodies</td>
</tr>
<tr>
<td>3. Schools</td>
<td>22</td>
<td>52</td>
<td>3</td>
</tr>
<tr>
<td>4. Science Teachers</td>
<td>50</td>
<td>59</td>
<td>8</td>
</tr>
<tr>
<td>5. Heads of Institutions</td>
<td>22</td>
<td>52</td>
<td>3</td>
</tr>
<tr>
<td>6. X class students</td>
<td>250</td>
<td>295</td>
<td>40</td>
</tr>
</tbody>
</table>

4.4 COLLECTION OF DATA

As explained earlier, data from three sources, viz., teachers, heads of institutions and students were necessary for the study. Hence, the investigator personally visited all the schools included in the sample (106 schools) a good rapport was developed with the head of the institutions. The rating scale meant for him/her was given with a request to rate the teaching effectiveness of science
teachers handling higher classes in his/her institution. The procedure of rating was explained in the instructions with the help of an example to make him/her understand clearly what he/she should do.

The science teachers, who were rated by their head, were given the copies of all the instruments meant for them, viz., teaching aptitude test, attitude scale, 16 personality factor questionnaire and personal data sheet and requested them to respond to all parts of the booklet of instruments and to all questions without leaving any part or any question.

With the help of the class teachers of different sections of 10th class, 5 students of both sexes, were selected from each of the sections to obtain the ratings of teaching effectiveness of their science teachers. The students were selected randomly from among those who were attending the school regularly and those who were studying in the same school for the past three years. The rating scale was administered to those students after making them to sit in a separate room where no teacher would observe what they do. The investigator demonstrated the procedure of rating step by step for obtaining the same from the students. Thus, the data were collected from science teachers, heads of institutions and students.
4.5 SCORING OF THE RESPONSES

Attitude scale was scored on a five-point scale by giving weights 5 to 1 in the case of positive items and 1 to 5 in the case of negative items to the five alternatives, viz., strongly agree to strongly disagree respectively. The grand total on the entire scale was obtained by adding the weights on all the statements.

The composite index of teaching effectiveness was obtained as described earlier by calculating the average of the average student ratings and the ratings given by the corresponding heads of the institution for each of the science teachers.

For every item in the teaching aptitude test, the weights were given to 3, 2, 1, 2, 3 for Highly agree (HA), Agree (A), Indifferent (I), Disagree (D), Highly Disagree (HD). Of both the right and wrong items respectively, subtract the total wrong score from the total right score which gives the individual correct raw score in a particular sub-test. The grand total of the teaching aptitude test was obtained by adding the scores on all the 10 sub-tests.

16 PF questionnaire was scored as per the weights provided by the authors. Except the items in Factor B, all the other items were placed on a 3-point scale with weights, 0, 1, 2, for the items in factor B the weights are 0 and 1.
The information provided by the respondents in the personal data sheet was also numerically coded to suit the computer analysis.

4.6 STATISTICAL METHODS EMPLOYED IN THE ANALYSIS OF DATA

The total scores obtained by each of 200 subjects on all the variables were computed. The data were carefully analysed by employing appropriate statistical techniques.

To estimate the general level of teaching effectiveness of science teachers all the descriptive statistics such as mean, median, mode, range, Q.D., S.D., skewness and kurtosis were calculated. To study the influence of independent variables on teaching effectiveness inferential statistics such as 't' test and 'F' test were employed appropriately. The profile similarity coefficient 'r_p' was also computed to study the personality profiles of less and more effective science teachers. Finally, step-wise regression analysis was carried out to identify the contribution made by each one of the independent variables on teaching effectiveness of secondary school science teachers.

The detailed analysis of the data and the discussions made on the results are presented in the succeeding chapter.