CHAPTER V

SPECIFIC PROBLEM AND PROCEDURE IN THE PRESENT INVESTIGATION

Introduction: A personality questionnaire is sensitive to some manipulable variables. Different testers, instructions, social desirability, surroundings, response sets, acquiescence, response categories, language, forms of the questionnaire are found to elicit different test taking attitudes, varying response tendencies and different personality patterns of the subject.

Social desirability: It is extremely easy to falsify the answers in such a way as to convey a picture of 'good personality', and subjects will always tend to do this if there is any incentive. For example, if tests are used for selection, or if they are to be seen by a lecturer upon whom students wish to make a good impression testees will be set to convey much favourable picture. Similarly,
many more unpleasant symptoms' are admitted when tests are answered anonymously, or are to be used for counselling purposes where the subject is motivated to be candid.

More subjects answer 'No' to an unpleasant item than say 'Yes' to the same item reversed. This suggests that most subjects are defensive, and that personality questionnaires generally show quite high reliability or consistency of response, more because of this pervasive attitude rather than because the subjects are really consistent in their neurotic, introverted or other behaviour. This explains too why there is always considerable overlapping between tests of different traits, why the correlations are about as high as between different tests of nominally the same trait. This phenomenon is the same as the halo effect in ratings. Even when the test attempts to describe subjective behaviour symptoms, subjects answer it very largely in terms of their self-halos.

Quite apart from intentional faking, people do not know themselves well enough to answer factually. They are likely to produce rationalizations. However, not everyone aims merely to display a favourable self picture. Some are more suggestible, and they tend to exaggerate their defects e.g. neurotic
patients or highly educated persons such as students are very often self-analytic and introspective, more self depreciatory than non-academics are.

In any test where a range of answer is provided, e.g. from strongly agree through agree and doubtful to strongly disagree, some subjects give many more extreme responses, fewer middle ones, than others. Similarly in interest tests such as strong's, where the responses are like, indifferent, or dislike, some people are universal 'likers', others more guarded.

Some of the sources influencing test performance are noted below:

Influence of test-taking attitudes:

Sarson (1960) suggests that the predictive value of personality (and ability) tests has been disappointing because they have neglected many factors that influence the subject over and above the actual content of the test - the nature of the instructions, his idea of the purpose of the test, time of day, the personality of the tester, what he has learnt from previous testing, etc. etc.

Cronbach makes the useful distinction between 'maximal performance' and 'typical performance' tests, the former being employed in testing aptitudes and achievements, the latter in personality and attitude testing. Neither
The term is really satisfactory because subjects who are trying too hard in ability tests find to perform somewhat erratically. But 'typical' or 'normal' is still more ambiguous since the 'normal' response of any person is to 'put-across' a 'suitable' personality and not to display his 'real' self if he can possibly help it. Cronbach goes on to say that the tester must give the subjects the full possible instructions in ability testing as to what he wants, but in personality testing he must conceal his aims. Thus he should in general avoid performance tests which constitute direct samples of a trait, such as May and Hartshorne used in studying honesty. But the tester who conceals his object is in danger of becoming so deceptive as to transgress ethical principles, or his evasions may merely encourage the subjects to build up more distorted myths.

Response Sets:

This is Cronbach's term for another type of extraneous influence on test scores, namely stylistic consistencies, stimulated by the form of response. Frederiksen and Messick (1959) classify some of the commonly occurring sets as follows:
1. Acquiescence: tendency to accept any personality statement as applying to oneself, or alternatively to reject all items. This may or may not be the same as preference for 'Trues', or for 'Falses', in the true-false achievement test.

2. Evasiveness: giving many doubtful or indifferent responses.

3. Extremeness: giving a preponderance of strongly agree or disagree rather than intermediate responses.

4. Inclusiveness: when the number of responses is unspecified, giving a large proportion of these. Its opposite may be criticalness in accepting words, phrases etc., as adequate or as equivalent.

5. Answering in terms of social desirability.

6. Other tendencies to fake or distort, wittingly or not.

7. Cautiousness: e.g. omitting difficult items in an ability tests VS guessing.

8. Preference for working at speed, or slowly.

9. Tendency to be consistent or inconsistent where two or more responses in the same tests have practically the same content.
Here, however, we are concerned with sets that affect personality test responses and these appear to depend to varying degrees upon (a) significant personality traits, (b) uninteresting stylistic or expressive habits, (c) temporary moods or reactions to the test content or instructions.

Acquiescence: Messick and Jackson (1958) have brought out the important influence of acquiescence on such personality inventories as the MMPI. Here the score for the various scales, hysteria, schizophrenia, etc., involve both positive and negative items i.e. the subject has to answer 'Yes' to some, 'No' to others, to score highly on most scales. But it happens that a great many items are keyed 'Yes' in some scales, fewer in others and Messick shows that the same scales that have the largest proportions of yes's, i.e. those most likely to involve acquiescence, are also the scales that load most highly with the first, general factor, obtained by intercorrelating and factorising. In other words, a strongly acquiescent subject who accepts both positively and negatively worded items is likely to score highly on all scales and so to appear generally abnormal or psychotic. The second factor appears to be closely connected with social desirability.
Since it loads most highly these scales with many items keyed 'No', i.e. it involves the rejection of socially unacceptable items.

Further light on the personality correlates of acquiescence is thrown by the common finding that the F scale and other tests of dogmatism and conformity tend to give considerable negative correlations with academic achievement. On the basis of a factorial study with neurotic patients, Eysenck (1962) claims that acquiescence and extremeness vs. indecisiveness affect scores on social attitude tests like the F scale but not those on personality inventories.

This has been a complex discussion, and it has failed to lead a satisfactory outcome. It implies that self-report test scores are seriously distorted in two main directions, and yet at the same time we cannot get rid of these distortions without impovershing the scores. The situation is similar to that with 'halo' in ratings.

The above mentioned points are some of the main problems in personality assessment. Thus, as in any scientific endeavour if something is to be studied and investigated, means must be developed for observing and measuring the thing of interest. Procedures for observing and measuring
personality variables have been described in several books (Eysenck, 1952, Ferguson, 1951, Vernon, 1953) devoted primarily to this subject. Hardly a few of these books give serious consideration as to what the real satisfactory factors affect in personality measurement. It was at this point that the investigator felt the need to study some of the problems that affect, the performance, especially in questionnaire method of personality measurement.

The purpose of this study was to compare the personality questionnaire items of Bernreuter personality inventory when they were presented to high school students in different forms of instructions at different times to study the effect of instructions. Secondly, to study the effect of forms of the questionnaire, this inventory was presented in the question form as well as in the statement form. Thirdly, the effect of different response categories was also studied when response categories were presented in two and three categories. It was aimed to find out whether these changes brought about a predisposing effect on the responses of the subject.
Hypotheses:
Three different hypotheses were studied:

1) To test the effect of two types of instructions.

2) To find out if there is any difference in test performance on different forms of test items, viz. question form and statement form.

3) To test the effect of different categories of responses.

The inclusion of both boys and girls in the sample enabled the investigator to study also the sex differences as far as above mentioned influences were concerned.

Test Materials:

Bernreuter Personality Inventory was used. There are 125 questions which can be answered in yes, no or in ?. This inventory has been validated against other inventories supposedly measuring the same traits and it yields separate scores for:

Bl - N: a measure of neurotic tendency tend to persons scorin high on this scale tend to be emotionally unstable. Those scoring low tend to be very well balanced emotionally.
Persons scoring high on this scale prefer to be alone, rarely ask for sympathy and tend to ignore advice of others. Those scoring low are extravert.

B3 - I: a measure introversion-extraversion.
Persons scoring high on this scale tend to be introverted; that is, they are imaginative and tend to live within themselves. Those scoring low are extraverted; that is, they rarely worry, seldom suffer emotional upsets, and rarely substitute daydreaming for action.

B4 - D: a measure of dominance-submission.
Persons scoring high on this scale tend to dominate others in face to face situations. Those scoring low tend to be submissive.

F1 - C: a measure of confidence in oneself.
Persons scoring high on this tend to be self-conscious and have feelings of inferiority. Those scoring low tend to be wholesomely self-confident and to be very well-adjusted to their environment.

F2 - S: a measure of Sociability.
Persons scoring high on this scale tend to be non-social, solitary, independent; those scoring low tend to be sociable and gregarious.

This inventory has separate norms for male and females at high school, college and adult level.
Administering the Inventory:

1. The inventory was self-administering. No instructions were necessary except those appearing on the blank. To insure the careful reading of instructions, the investigator read them aloud, while the individual being tested read silently.

2. Each person was left to interpret questions for himself. The investigator did not explain how he thinks a question should be interpreted.

3. There was no time-limit. Very few subjects required more than 40 minutes to complete the inventory.

4. The investigator was careful to point out to the subjects that the value of the results was dependent upon their own sincerity and further the examiner had guaranteed the confidential treatment of the findings.

5. The exact nature of the traits being measured was not revealed before the subjects had finished test.

Sample:

For the present investigation, a sample of 100 students studying in grade 'X' was taken. The 'purposive incidental sample' i.e. the unselected groups of different high school students in Baroda City as mentioned below were chosen for this study.
1. Uni Exp. High School
2. Sayaji Boys' High School
3. Sayaji Girls' High School
4. Jeevan Sadhana
5. New Era High School
6. Shikshan Sadhana
7. Wadi High School
8. Bharati High School etc.

Most of them belonged to the middle class status. The students were between the ages of 14 to 16 with a mean of 15. The method of random sampling was adopted.

Experimental Design

These subjects were administered different forms of the test (condition) in different orders or sequences which formed a sort of counterbalanced design. Whenever the subjects are exposed to two or more conditions, there is a possibility that order effects will influence the results, that is, the impact of a particular condition may depend on whether it is administered first or second. In view of this, a counter-balancing procedure was used to ensure certain additional information from the data.
Reliability of test Material:

For the present investigation, this inventory was translated in Gujarati version. To see the reliability of inventory, it was administered on the same individuals on two different occasions at the time interval of 20 days. Retest reliability was found to be 0.9. To see whether the subjects understand the same thing in the same sense, this inventory was given in original form (English version). After an interval of 20 days it was again administered in Gujarati version on the same population. Reliability was found to be 0.91.

Procedure:

Three types of forms or conditions were investigated:
1. The effect of two types of instructions
2. The effect of Question form and Statement form
3. The effect of two response categories and three response categories.

To examine the first condition i.e. to study the effect of two types of instructions, the Bernreuter Personality Inventory in Gujarati version was administered to students studying in X grade. This inventory was applied in question form both the times, but quite different instructions were given at different times. First, all the subjects were required to answer the questions very honestly and frankly and after a time-interval of 20 days instructions were
changed. At the second time, they were required to answer the questions in such a way as to convey a picture of good personality, i.e. they were required to answer the questions strictly in a socially desirable way. (as understood from the forms given in Appendix 4 (a) and (b))

The counter-balanced design stood thus:

One Group : Honest response : 20 days Social Desirable (time response interval) 

Second Group : Socially desirable response : 20 days Honest response (time interval)

To study the second condition i.e. to study the effect of question form and statement form the same method was used. First of all this inventory was applied in question form and after an interval of 20 days the same inventory was administered in the statement form (as understood from the forms given in appendix 4 (c) and (d)).

To give an example:

Do you daydream frequently? (Question form)

I daydream frequently. (Statement form)

A counter-balanced design was taken to control the effect of instructions or form. Two groups were taken. To one group, the inventory was administered first in question form and after a time interval of 20 days, it was administered in the statement form. To another group, this inventory was administered in first statement form and
after a time interval of 20 days it was then administered in question form. This counter-balanced design can be represented thus:

One Group : Questions form 20 days (Time interval) Statement form
Second Group : Statement form 20 days (Time interval) Question form

To study the third condition i.e. to examine the effects of different categories of responses the same method was used. The same inventory was applied in question form first with two response categories and after a time-interval of 20 days, they were required to answer questions in three response categories. (As understood from the form given in appendix 4 (e) and (f).

To give an example:

Do you daydream frequently? Yes No (Two category responses)

Do you daydream frequently? Yes No Doubtful (Three category responses)

Again, the counter-balanced design was used as shown below:

One Group : Two Categories : 20 days time : Three interval categories
Second Group : Three Categories : 20 days time : Two categories interval

Scoring:

Six separate scoring keys as given in standardized procedure, by test authors were used in scoring, one for each of the
traits tested. In the preparations of these keys, the diagnostic value of each response to each question was determined for each of the traits. Weights from -7 to -7 were assigned in accordance with these diagnostic values. The total score for a trait was the algebraic sum of the weights which corresponded to the responses made by the individual, as given on the key for that particular trait. If an individual failed to answer a question, it was scored as though he had encircled the question mark (doubtful). A summation was done of each subject's score of 125 questions.