C H A P T E R - III

M E T H O D O F S T U D Y

3.1 Historical Orientation to testing and Evaluation
3.2 A Criticism of Measurement and Testing.
3.3 General Method of Study
3.4 Variables Used
3.5 Idea Behind Sampling
3.6 Data Collection And Credibility of tools Used.
3.7 Method of Analysis and Statistical Treatment
3.1 Historical Orientation to Testing and Evaluation

People have always been interested in the assessment of human attributes, but testing, as we know it today, is a phenomenon of the twentieth century. In the western world, the oral examination reigned supreme for centuries, although the Chinese had a well-developed system of written examination as early as 1000 B.C. Qubis (1965) writes that achievement or proficiency tests were uniformly applied to all men in the empire who aspired to public office.

It was not until the mid 1800s that American educators began seriously to question the validity of the oral method of appraising student progress. The first recorded instances of written examinations being substituted on a large scale for oral questioning of students occurred in the Boston Public Schools in 1845 (Caldwell and Courtis, 1924).

In order to understand the development of educational testing and evaluation in the twentieth century, we must also consider the growth of measurement in Psychology. The two paralleled each other, and developments in both education and psychology were
spurred by events in the latter half of the nineteenth century which changed man's way of thinking about himself. In 1850, psychology was still largely a philosophical discipline and had not yet developed the research methods essential for scientific enquiry into human behaviour. One of the most promising milestones in the growth of psychology was the work of Gustav Fechner, during the middle decades of the nineteenth century. At the University of Leipzig, he laid the logical foundation of an area of research known as "Psychophysics". Which is the precise and quantitative study of how human judgements are made.

Another challenge to man's traditional way of thinking arose with the publication in 1859 of Darwin's "On the origin of Species," Not only did the theory of evolution have a great impact on the philosophical world, but it also had a profound influence on the direction of research in the scientific community. Prior to this time, the Predominant view was that man is a static being, having possessed since the day of creation a Uniform and unchanging set of physical and mental attributes. Any individual differences were considered "nature's mistakes" in producing the
average man. There had been systematic attempt to study the part which individual differences play in every day life.

Sir Frances Galton, an adherent of Darwin's theories began the first large scale testing programme to measure many different human attributes at his Anthropometric Laboratory in 1884. Each Visitor was charged three pence for having his measurements taken on a variety of physical and sensory tests. In order to analyze the data obtained, Galton made use of statistical methods. His particular need was for a measure of association, or correlation to detect the amount of resemblance between the individual characteristics of fathers and their sons. Karl Pearson, a colleague of Galton's and a genius in mathematical statistics, derived the statistics of correlation, the procedures which are now widely used in the study of individual differences.

The last two decades of the nineteenth century saw a considerable expansion of studies of individual differences in England & America. The advent of modern intelligence tests paralleled important changes in social philosophy. The Binet–Simon scale published in 1905 was radically different from the tests of Galton and his
followers in that it concerned the child's ability to understand and reason with material in cultural environment. By 1915, the basic principles and techniques of educational and psychological measurement were becoming established. Fundamental statistical methods were known.

In the period between 1900 and 1915, standardized achievement tests in school subjects were being developed as well as intelligence tests.

The entrance of the United States into world war I spurred widespread applications of the new principles of testing. As a result the Army Alpha and Beta tests of intelligence evolved. Nearly two million men were tested with these instruments in 1917 and 1918. In 1923, the first standardized achievement battery was published. The period from 1930 to 1945 was a time of critical appraisal and of taking stock of the tests published. The emphasis shifted from measurement of a limited range of skills to evaluation of the whole range of educational objective, including interests, attitudes, and appreciations.

During world war II, psychologists were again called upon to devise tests to and in the screening and
classification of large numbers of men and women
in the armed forces.

To-day standardized commercial instruments are
available for many different purposes. The expanded
range of concern with educational and psychological
measurement fit in with the shibboleth of the times-
concern for the "Wholechild" or "Whole human being".

3.2 A Criticism of Measurement and testing

After a brief idea of the history and evolution
of testing and evaluation, the investigator considers it
very important to study the significance and hence the
philosophy of testing and evaluation. Since the study
involves the use of various tests, it is wise to consider
the proper attitudes to hold towards the use of tests in
making educational decisions.

Although standardized tests are becoming widely
accepted by both educational experts and the Public at
large, there still are many who criticize their use.
The first is that the wide use of tests leads to
"unfair" educational practices, and the second, tests
are poor measurement devices. Regarding the first point,
critics will say that tests serve to brand" students or
subjects, unfairly segregate them into ability groupings restrict the ranges within which students are allowed to grow and change, and encourage unhealthy feelings of superiority and inferiority in human beings. These criticism are potentially correct, but it must be firmly kept in mind that it is not the test, perse, that bring about such unfortunate consequences but rather the improper use of test results.

The criticism that tests are poor measurement devices take several forms. Thus the indictment of tests is that they are not very effective measures of aptitude, achievement, personality and other characteristics which they attempt to measure. It is easy to find some poor items on even the most carefully constructed test, and it is easy to find whole tests which are not good. Also, there is still much that we do not know about educational measurement, and in some instances admittedly crude measures must be used for want of more valid instruments. However, these potential and real faults should not be used as arguments for doing away with tests, rather, they should act as spurs towards the construction of better instruments.
Whenever research comparisons are made between the differential effectiveness of tests and subjective evaluations, test prove to be cheaper, quicker and more valid. Undoubtedly tests provide a useful source of information that is helpful in making educational decisions. The purpose of tests is to help take the personal element and the guess work out of decisions. To disown the use of tests and ignore the facts only postpones the day of reckoning and keeps all disturbed. Taking a frank look at our abilities and personality characteristics is not always pleasant at the moment, but it is necessary for long-range achievement and happiness. Because there is nothing better to employ, it is not sensible to discuss whether or not tests should be used. Rather, the important point to discuss are how tests should be used and what procedures can be employed to make tests more effective.

3.3 General Method of Study

The main thrust of the present research work is concerned with analysis and description of Principal effectiveness and also to evaluate its relationship with certain other variables like Professional Attainment etc. It is the study of establishing relationships. Such a study which may include present facts or current conditions
concerning the nature of a group of persons, a number of objects, or a class of events and may involve the procedures of induction, analysis, classification, enumeration or measurement fall under Description survey methods of research. This usually covers the following aspects of the School system: -

(1) Aims, outcomes, pupils achievement, curriculum method and instructional aids.

(2) Administrative problems and procedures of the School.

(3) Financial Policies and procedures

(4) Operation and maintenance of the physical plant

(5) Pupil transportation

(6) Staff personnel

(7) School plant and related factors

(8) Instructional programs and the related policies and procedures that effect the educational program.

This investigation too follows survey method. It is a correlation study in which it seeks to empirically study correlation between variables selected for the study. The questionnaire technique was employed to collect the data.

3.4 Variables Used

A research question asks about the relationship
between two or more variables. A research hypothesis states the expected answer to the research question.

Both, a research question and a research hypothesis contain at a minimum an independent variable. The research question asks about the relationship between the independent and the dependent variable, the research hypothesis states an expected relationship between the independent and the dependent variable. There are several other type of variables. The variables used in the study have been described below:

(a) **Dependent Variable**: It is the outcome which is expected to arise from some treatment. Such variables are dependent in the sense that they depend on the treatment. For example, if the treatment has one effect, the dependent variable may have a low value; whereas if the treatment had a different effect the dependent variable may have a high value.

Since the investigator intends to find out the effect of professional attainment, socio-economic status, attitude towards teaching and values of life of Principals on the effectiveness of School Principals, Principals effectiveness is the dependent variable.
(b) **Independent Variable**: It is the treatment that is expected to produce an outcome. Independent variable and treatment are used synonymously in this text. The independent variable is independent in the sense that it does not depend on the outcome (dependent) variable.

The treatment given thus, professional attainment, socio-economic status, attitude towards teaching and values of life of Principals are the independent variables.

(c) **Moderate Variables**: This is the variable that moderates the impact of the independent variable upon the dependent variable. In the present study the sex and age of the Principal are moderate variables.

(d) **Intervening Variable**: This is a hypothetical variable that is assumed to be created by the operationally defined independent variable and that in turn is assumed to have an impact upon the dependent variable. The treatment produces the intervening variable, and the intervening variable in turn produces the outcome. This variable intervenes in the sense that the treatment does not produce the outcome directly but rather through the mediation (intervention) of this invisible, hypothetical, internalized process. The investigator assumes that there is no intervening variable.
(e) **Control Variable**: It is designed merely to eliminate (rather than measure and describe) its impact upon the relationship between the independent and dependent variables.

Since the study is confined to only High School Principals or Junior High School Principals, this forms the control variable.

3.5 **Idea Behind Sampling**:

Two important factors have to be taken into account while identifying the sample to be studied. They are

1. **Method of Sampling**
2. **Size of the Sampling**

The term sampling refers to strategies which enable us to pick a subgroup from a large group and then use this subgroup as a basis for making judgements about the large group. In order to use such a subgroup to make decisions about the large group, the sub group has to resemble the larger group as closely as possible.

Random Sampling is generally the best way to draw from a population with random sampling, every member of the population has an equal opportunity to be in the
sample, and pure chance is the only factor that determines who actually goes into the sample. In the present study sampling is done randomly. For the preparation of the scale measuring Principals effectiveness, teachers from various Schools, Principals and Educationists have been consulted. Further for the evaluation of Principals effectiveness, teachers of varying experience have been studied. In order to study the Principals professional attainment, socio-economic status, attitude towards teaching and values of life both private Schools and Government Schools at various places in Aligarh, Yamuna Vihar, Delhi, Shahdra and Gonda districts have been surveyed.

The investigator predicts that a sample study of 50 School Principals is sufficiently large to generalize characteristics of effective and Ineffective Principals.

A detailed discussion of the distribution and characteristic of the sample is furnished in the following chapters.

3.6 Data Needed For the Study

The hypotheses sought to be tested and the
objectives to be fulfilled required the following data recording:

Principals & Educationists.

(a) Teachers' perception of Principals Effectiveness for preparing the scale, "Measure of Principal Effectiveness.

(b) Teachers' evaluation of School Principals

(c) Professional attainment of School Principals

(d) Socio-Economic Background of School Principals

(e) Attitude towards teaching of School Principals

(f) Values of life of School Principals

3.7 Methods of Analysis and Statistical Treatment

The data recorded was categorized into two samples, one indicating the Principals' Effectiveness as shown by the teachers and the other indicating the Principals' Socio-Economic Background, Professional Attainment, Attitude towards teaching and values of life shown by the Principals' themselves.

From the first Sample an average score was calculated from the five scores obtained on the scale, "Measure of Principal Effectiveness", as given by five teachers of each School. The scoring of this scale has been mentioned in chapter IV. This score
contributed to the Principals' Effectiveness of each School. The Second Sample giving scores of Socio-Economic Background, Professional Attainment, Attitude towards teaching and values of life of Principals were tabulated against the average scores of Principal Effectiveness respectively.

The Age and Sex of the Principals were also noted. This table of raw scores were statistically treated as follows:

**Stage (1):** At this stage the data was used in "Construction of tools".

The tool "Measure of Principal Effectiveness" prepared by the investigator herself had to be critically treated for determining its reliability and validity. The reliability of the tool, "Measurement of study of values" was also found out because of the translation made by the investigator from Hindi to English.

(i) *Split-half method* was used for determining the reliability

(ii) *Item validity* was found out by correlating each item score with the total score.
Stage (2) : The data was statistically treated in three ways:

(i) 't' values were found out

(ii) Pearson's Product moment correlation 'r' was calculated

(iii) Regression analysis was carried out

To determine the variability or dispersion of a group, Standard Deviation (S.D.) are considered. 't' ratios show whether the difference between means of two samples truly exist or it is merely due to sampling fluctuations. If we are able to say with considerable confidence that difference between the means of two samples is significant, we reject the null hypothesis. If there is no difference then the null hypothesis is retained.

Product moment coefficient of correlation is also termed as Pearson's 'r' or linear correlation. This is one of the standard techniques used for describing the degree of association existing between two variables. There are certain assumptions to be met in employing the Product moment coefficient. First, there must be a linear relationship between the variables. Secondly, each of the variables must be normally distributed and finally the relationship must
be homoscedastic that is, the spread about the best fitting straight line must be the approximately the same at all levels of two variables. But, in practical situations perfect linearity and normality are not possible. There has been a considerable controversy as to whether the above three characteristics should be considered "assumptions" in a correlational analysis (Nunnally, 1967). When these characteristics are present, the relationship is said to be bivariate normal. Other than for the assumptions required for the employment of the inferential statistics, bivariate is an assumption not too much for using the Product moment correlation, as for interpreting the results. Thus there is nothing to prevent the use of Product moment coefficient of correlation even if one of the distributions is markedly different from the other in shape.

In the present study, the variables involved are continuous and with an eye over the above mentioned discussion it was decided that Product moment coefficient of correlation would be an appropriate measure to carry out the analysis.

Multiple regression analysis and Multiple
Correlations were also employed. The multiple regression equation is used for two purposes (1) Analysis (2) Prediction. In analysis, the purpose is to determine the "weight" of each variable in contributing to some final result. While in prediction some methods are followed to predict the criterion with maximum degree of efficiency, (Garrette, 1966). The present study covers only the analysis part.

Regression analysis was performed in order to test the hypothesis regarding the effect of the variables, Professional Attainment, Socio-Economic Background, Attitude towards teaching and values on Principal Effectiveness. The analysis was done step wise and a F ratio for the addition of each variable to the regression equation was calculated. This was done to see whether the addition of the variable just added makes significant contribution (Thomson, 1951) to multiple correlation.