CHAPTER 3

RESEARCH DESIGN AND TOOLS

3.0 Introduction

This chapter on research procedures contains a description of the sample, the measures used, and the steps taken in carrying out the research. It starts with a detailed description of the sample followed by the explanation of each measure and explains why the measure was selected and what specific purpose it will serve in the research. In the final section of the chapter, the investigator describes the research procedures starting by identifying the type of experimental design and then showing how it was applied to this study.

3.1 Research Design

This section describes the sample and determines the degree to which the research sample is representative of the population, how the samples were matched, and eventually which method was used to place the subjects into the experimental group or the controlled group.
3.1.1 Description of the subjects or the sample

A detailed description of the sample is needed in order for the reader to assess the generalizability of the research findings i.e., to determine the degree to which the research sample is representative of the population and is comparable to other samples to which one may wish to apply the research results. The investigator used parallel-group design with matching, for which the following steps were taken:

- Administered measures of dependent variable;
- Matched subjects in experimental and controlled groups on the basis of the matching variables;
- Exposed the experimental group to the experimental treatment and administered a placebo to the controlled group;
- Administered measures of the dependent variable to the experimental and controlled groups and compared them in order to determine the effect of the experimental treatment upon the dependent variable.

The question of how to match sample is an important one in most educational research. The investigator used parallel-group design with matching. First we will see what is a parallel-group design and how the investigator used it in her research.
In parallel-group design two or more groups are employed. In its simplest form, experimental group and controlled group are used. The research worker attempts to expose the two groups to as nearly identical treatment as possible with the exception that the experimental group is exposed to the experimental variable. Matching in such designs refers to an attempt to place subjects into experimental and controlled groups in such a manner that they are closely comparable on a pre test that measures the dependent variable correlated with the dependent variable. This reduces the initial differences between the experimental and controlled groups on the dependent variable. Thus it is important to have a matching variable that correlates highly with the dependent variable. Another advantage of matching is that it permits the division of experimental and controlled groups into subgroups on the basis of matching variable in order to determine whether the experimental treatment has had a different effect on subjects at different levels.

Experimental designs are subject to errors arising from the experimental treatments in addition to those arising from subject differences. Linquist (1953) describes three types of errors. These are Type S errors that are associated with the sampling of the research subjects, Type G errors that are associated with the research groups, and Type R errors that are associated with the replications of the experiment.
3.1.2 Problems associated with parallel-matching design

Most difficulties that occur in the application of parallel-group design with matching group design revolve around the matching procedure. The first thing we need to determine is what variable to use for matching. Matching on a number of variables that are correlated with the dependent variable will reduce errors more than matching on a single variable that is less highly correlated. In attempting to match on more than two variables, however, a difficult problem often comes up because of the impossibility of finding individuals who are reasonably well-matched on several variables. Under these conditions, many subjects must be discarded for whom satisfactory matches cannot be obtained. Another related problem is to determine how closely to match the subjects on a matching variable. This close matching, however, although leading to gains in precision, increases the number of subjects who cannot be matched. Many statisticians question the value of any matching procedure that results in losing subjects because such a procedure may seriously change the nature of the sampling and lead to sampling biases.

The question of how to match is an important one in most educational research projects. There are three different kinds of matching procedures. The first one is the person-to-person matching technique in which an attempt is made to locate two persons among those available, and then the examiner assigns randomly one subject to the experimental group
and another subject to the controlled group. This procedure cannot be followed in a situation where established groups have already been formed as in a classroom situation. This procedure leads to non-random groups and, hence, an increase in Type S and Type G errors may be expected.

Another matching technique is an attempt to match the entire group in terms of mean and standard deviation rather than to match individuals. In Group matching, as it is called, wherein if the groups are not comparable in mean and standard deviation, the investigator juggles a few subjects from one group to the other in order to obtain a closer match. The principal disadvantages of this approach are the loss of randomization caused by the usual juggling, the lack of random groups available at the start, and the smaller gains in precision as compared with individual matching. The third method of matching that many statisticians consider superior to the first two is to place all subjects in rank order on the basis of their scores on the matching variable. Two subjects are chosen based on their rank order and using random means (flipping a coin) one subject is assigned to the experimental group and another subject is assigned to the controlled group until all the subjects have been assigned. The principal advantage of this procedure over the previous two is that it provides a more random group and no cases are lost because of inability to match. Matching, of course, will be less precise than would be the case in close person-to-person matching, but the advantages of not losing subjects outweigh the disadvantages of the less precise matching.
The investigator used all three methods of matching. Group matching was used when one E5B ESL English class was selected as the experimental group and another E5B ESL English class was selected as the controlled group. Person-to-person matching was used when the investigator matched students based on their social status, scores in previous exams, and their overall scores on the report card from the previous term. Then the investigator put members of the experimental group and the controlled group in the rank order to bring the matching process to a fruition.

3.2 Methods of data analysis

Statistics, along with tests of other measurement techniques, provide the basic tools used by the research worker. Without an understanding of the tools available, it is impossible to design and carry through a sound educational research project.

Most statistical tools used in educational research serve one of three functions. Descriptive statistics involve tools that tell us about the nature of the groups we are studying. The mean, medium, and standard deviation are examples of devices that describe groups. Inferential statistics provide us with the means of measuring differences, and relationships. Such tools as the t-test, analysis of variance, and chi-square are employed to determine the significance of differences that we obtain in educational experiments.
The various correlational techniques are used for measuring the degree of relationship between educational variables and are employed in research where relationships are sought out and applied. Studies involving educational prediction and selection employ correlational techniques. Statistical tools are used during the analysis of the research, but these tools are also important in other aspects of the total research picture.

We will analyze four kinds of information about statistics used in the research. It could be summarized as follows:

1. What statistical tools are available?

   It is not uncommon for a research worker to overlook and fail to make maximum use of statistical data collected. Mastering the mathematics involved is much less important than knowing what tools exist and how to apply them to research problems.

2. Under what conditions is each tool used?

   To develop a sound research plan, the investigator must be able to examine the hypotheses and procedures and decide what statistical tools will be most appropriate for carrying out his analysis. This decision should always be made prior to the collection of data because different statistical tools often require that the data be collected in different forms.
3. What do the statistical results mean?

After the investigator has decided what statistical tools are most appropriate for the analysis of his/her research data and has applied these tools, he/she is faced with the problem of interpreting the research results.

4. How are the statistical calculations made?

It is important to direct the investigator's attention to learning what statistical tools are available, when they are used, and what the results mean after they are used.

We shall now examine the statistical tools that are most commonly used in educational research and we shall try to fit each of the statistical tools into our "mental tool box." We shall consider each tool available and think of the conditions under which one might use this tool and what one must remember to interpret accurately the findings provided by the tool.

3.3 Descriptive statistics

Calculating mean or median are considered the best measure of central tendency. Among various measures of variability, standard deviation is the most stable measurement of variability. The bell shaped curve is known as the normal probability curve.
3.3.1 Measures of central tendency

Though a number of measures of central tendency can be found, the only ones used to any extent in educational research are the mean and median. These measures are sometimes called descriptive statistics because they give a very brief description of the sample that has been measured. Calculating the mean or median gives a rough description of the group in terms of average score attained. The mean is generally considered the best measure of central tendency.

Computing the mean is one of the initial steps in applying many of the more advanced statistical tools such as standard deviation, analysis of variance, and correlation. One advantage of the mean over median is that it is more stable. This means that if one or two of one's subjects are lost, as often occurs in educational research, this change in your sample would have a more predictable effect on the mean than on the median.

The mean, calculated by dividing the sum of scores by the number of scores, emphasizes each score in direct proportion to its distance from the mean. Scores that are extreme are weighed more heavily in calculating the mean than scores that are less extreme. The median, by comparison, is merely the middle score in distribution and is not affected by the magnitude of scores at either end of the distribution. When a distribution of scores is symmetrical, the mean and median are located at the same place. When the
distribution has more extreme scores at one end than the other, that is, when it is skewed, the mean will always be in the direction of the greater number of extreme scores. Because of the difference between the mean and median in skewed distributions, the two measures can have considerably different implications. In research, the purpose of statistics is to provide the most accurate picture of data being studied. In most cases, the median gives a more accurate description of the group when the distribution is considerably skewed.

The Arithmetic Mean

To approximate the arithmetic mean of data organized into a frequency distribution, recall that the observations in each class are represented by the midpoint of the class. The mean of a sample of data organized in a frequency distribution is computed by:

\[
\bar{X} = \frac{\sum fX}{n}
\]

where:

- \( \bar{X} \) is the designation for the arithmetic mean.
- \( X \) is the mid-value, or midpoint, of each class.
- \( f \) is the frequency in each class.
- \( fX \) is the frequency in each class times the midpoint of the class.
- \( \sum fX \) is the sum of these products.
- \( n \) is the total number of frequencies.
Another characteristic that can be measured is the degree of skewness of a distribution. Recall that if a frequency distribution is symmetrical, it has no skewness—that is, the skewness is zero. If one or more observations are extremely large, the mean of the distribution becomes greater than the median or mode. In such cases the distribution is said to be positively skewed. Conversely, if one or more extremely small observations are present, the mean is the smallest of the three measures of central tendency, and the distribution is said to be negatively skewed.

Karl Pearson also developed a measure to describe the degree of skewness, called the coefficient of skewness:

\[
S_k = \frac{3(\text{mean} - \text{median})}{\text{standard deviation}}
\]

If two averages of a moderately skewed frequency distribution are known, the third can be approximated. The formulas are:

- Mode = Mean – 3(Mean – Median)
- Mean = \frac{3(\text{Median}) – \text{Mode}}{2}
- Median = \frac{2(\text{Mean}) – \text{Mode}}{3}

(Mason and Lind 1996: 144)
3.3.2 Measures of variability

There are several measures of variability, including the quartile deviation, average deviation, and standard deviation. In the educational research very little use is made of any measure of variability except the standard deviation. Standard deviation is the most useful and stable measure of variability. It is an important tool as it is a necessary step for applying more advanced statistical tools. The standard deviation, like the mean, provides a way of describing the scores of the group on the basis of a single measure. Like the mean, the standard deviation gives a rough picture of the composition of a group. The mean and standard deviation taken together give us a reasonably good description of the nature of the group being studied. Variability is important in educational research because the significance of many of the findings is dependent to a degree upon the variability of the groups being studied.

3.3.3 Standard deviation and the normal curve

Many characteristics of human behavior, including most of those measured by educational tests, have been found to be distributed along a bell-shaped curve similar to that shown in figure. This curve is a frequency polygon—that is, the height of the curve at a given point indicates the proportion of cases at that point. This curve shows that the majority of individuals whom we have measured are clustered close to the mean, and as
we move further and further from the mean we find fewer and fewer cases. The bell-shaped curve is known as the normal curve or normal probability curve. The exact normal curve has a mathematical formula. The normal curve can be developed by any technique that involves chance occurrences.

If the data of interest are in grouped form (in a frequency distribution), the sample standard deviation can be approximated by substituting \( \sum X^2 \) for \( \sum X \) and \( \sum fX \) for \( \sum X \). The formula for the sample standard deviation then converts to:

\[
s = \sqrt{\frac{\sum fX^2 - (\sum fX)^2}{n - 1}}
\]

where:

- \( s \) is the symbol for the sample standard deviation.
- \( X \) is the midpoint of a class.
- \( f \) is the class frequency.
- \( n \) is the total number of sample observations.
Most of the measures used in educational research are such that their distributions approximate the normal curve. Because our measures are not precise and the number of cases we work with is usually small, the distributions we obtain sometimes deviate considerably from the normal curve formula. If we divide the base line of the normal curve into a number of equal intervals such as shown in figure, we find that the percentage of cases that occur between two points on the base line is always the same if the interval we are concerned with is at the same distance from the mean and of the same length. The standard deviation divides the normal curve into a number of equal units. Because most of the data in educational research only approximates the normal curve, these percentages are also only approximate but are still of great value in statistics.

In analysis of research data, the standard deviation is an intermediate step that is computed as a part of the statistical calculations. The standard error, product-moment correlation, and many other statistical tools are based primarily on standard deviation. The standard deviation also forms the basis for various types of standard scores such as Z-scores and T-scores.

3.4 Inferential statistics - measuring differences

Much educational experimentation is concerned with measuring the different results that occur under different experimental conditions and determining the level of significance of these differences.
3.4.1 Standard error of measurement

Every measure used by man is subject to error. The measures used in the more natural sciences, such as physics, often are accurate with very small tolerances. The measurement devices used in educational research are subject to much larger errors than those that we find in the physical sciences. Relatively small errors occur with considerable frequency, but, as the size of the error increases, the frequency decreases. The standard deviation of the error curve is called the standard error of measurement. Standard error of measurement can also be regarded as a means of indicating how much a test score may be relied upon. It is closely related to the reliability coefficient although expressed in a different form.

3.4.2 Standard error in testing significance

The standard error of the mean tells us the extent to which we might expect the group mean to fluctuate if we repeated our study with other random samples. Standard error of the mean is an important concept in educational research because it forms one basis for determining whether the groups we compare in our research are significantly different. The comparison most frequently made is between mean scores of the same measure administered to two or more groups.

To determine whether the difference between the two means is significant, we compute a critical ratio or t-test. This is a ratio comparing
the difference that we have found between the two means in the experiment and the difference, based on the standard errors of the two means, that might be expected due to chance. If the difference between the means obtained in the experiment is larger than the difference we could expect to occur by chance, we report that the difference is statistically significant. This indication of significance tells the reader that the differences are large enough so that there is only a slight possibility that the difference obtained between the two mean scores in the experiment is due to chance fluctuations.

The standard error of the sample means is computed by:

\[
\sigma_x = \frac{\sigma}{\sqrt{n}}
\]

where:

- \( \sigma_x \) is the symbol for the standard error of the sample means.
- \( \sigma \) is the standard deviation of the population.
- \( n \) is the size of the sample.

Formula (8-1) for the standard error of the sample means assumes that the population standard deviation, \( \sigma \), is known. If it is not known and \( n \geq 30 \) or more (considered a large sample), the standard deviation of the sample, designated by \( s \), is used to approximate the population standard deviation, \( \sigma \). The formula for the standard error then becomes \( s_x \) is substituted for \( \sigma_x \) to indicate that the standard error is based on sample statistics:

\[
s_x = \frac{s}{\sqrt{n}}
\]

Note that the standard error of the sample means will vary according to the size of the sample (the denominator). As the sample size, \( n \), gets larger and larger, the variability of the sample means gets smaller and smaller. Logically, an estimate of the population mean based on a large sample is more reliable than an estimate made using a small sample. To put it another way, the error in estimating the population mean decreases as the sample size increases. If the sample size kept getting larger and larger and finally equaled the size of the population, there would be no error in predicting the population mean because the sample size and the size of the population would be the same!

(Mason and Lind 1996: 318)
3.4.3 "One-tailed" and "two-tailed" tests of significance

The ends of a normal curve where it approaches the base line are called the "tails" of the distribution. In effect, when we compare two means to determine whether they are significantly different, we are checking the degree of overlap between the tails of the standard error curves of these two means.

In most educational research, such as a study of comparing two teaching methods, we may expect the experimental method to produce greater achievement gains than the traditional method but one must not ignore the possibility that the reverse will be true. If it is highly unlikely that the experimental method would produce lower achievement gains, the "one-tailed" test of significance is appropriate to use. If the possibility is otherwise, then the "two-tailed" test of significance is appropriate to use. The main advantage of the "one-tailed" test of significance is that a smaller critical ratio (t-test) is needed to be statistically significant.
3.4.4 Standard error of other statistical measures

In addition to the standard error of the mean, standard errors can be calculated for the standard deviation, correlation coefficient, and other statistical tools. These standard errors can be employed to determine whether the two statistics compared are significantly different from one another. Hence, the critical ratio based on the t-test is the most used device for determining the level of significance of differences.

3.4.5 Relative dispersion

It is a direct comparison of two or more measures of dispersion. Karl Pearson (1857-1936), who contributed significantly to the science of statistics, developed a relative measure called the coefficient of variation. It is used in measuring when the data are in different units and also when that data are in the same units, but the mean are far apart.

Coefficient of variation is the ratio of standard deviation to the arithmetic mean, expressed as a percent.

In terms of a formula for a sample:

\[ CV = \left( \frac{S}{X} \right) \times 100 \]

Multiplying by 100 converts the decimal to a percent.
3.5 Testing the hypothesis

Hypothesis is a statement about the value of a population parameter developed for the purpose of testing. The hypothesis for the research study in question was as follows:

"The new strategies and methods are effective and as a result the experimental group will score substantially higher than the controlled group in a test which specifically measures the learners achievement at the end of the experiment."

Null hypothesis is a statement about the value of a population parameter. The null hypothesis for the research study in question was as follows:

"There is no difference in the traditional methods and the new strategies and as a result there will be no significant difference between the scores received by the experimental group and the controlled group."

Alternate hypothesis is a statement that is accepted if the sample data provide evidence that the null hypothesis is false. It describes what you will conclude if you reject the null hypothesis. It is referred to as the research hypothesis. The alternate hypothesis is accepted if the sample data
Step #2  The level of significance

It is the probability of rejecting the null hypothesis when it is actually true. It is designated with the Greek letter alpha $\alpha$. It is referred to as the level of risk i.e., the risk one takes of rejecting the null hypothesis when it is really true.

The investigator might make one of the two errors. Type I error is rejecting the null hypothesis when it is true. Type II error is accepting the null hypothesis when it is actually false. We often refer to these two errors as the alpha error, and the beta error $\beta$. Alpha ($\alpha$) is the probability of making a Type I error, and beta ($\beta$) is the probability of making a Type II error. The following table summarizes the decisions the researcher could make and the possible consequences.

<table>
<thead>
<tr>
<th>Null hypothesis</th>
<th>Accepts $H_0$</th>
<th>Rejects $H_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>If $H_0$ is true and</td>
<td>Correct decision</td>
<td>Type I error</td>
</tr>
<tr>
<td>If $H_0$ is false and</td>
<td>Type II error</td>
<td>Correct decision</td>
</tr>
</tbody>
</table>
Step #3 The test statistic

The test statistic is a value, determined from sample information, used to determine whether or not to reject the null hypothesis. There are many test statistics namely t, z, f, and χ², called chi-square. Regarding the sample size, Best says the following:

It is often stated that samples of 30 or more are to be considered large samples. It is approximately at this sample size of 30 that the magnitude of student's 't' critical values for small samples approach the 'z' critical values of the normal probability table for large samples (Best, 1983: 14).

In most cases, it is highly unlikely that the population standard deviation would be known, so the sample standard deviation must be used to estimate . As long as the sample size 'n' is fewer than 30, 's' can be substituted for as illustrated in the following formula. For the research study in question, the investigator used the t-test statistic because the sample size is under 30 students; to be exact there are 20 students in each of the two groups namely the experimental group and the controlled group. The value of 't' is determined by using the following formula:
where:

\[ \bar{X}_1 \] is the mean of the first sample.

\[ \bar{X}_2 \] is the mean of the second sample.

\[ n_1 \] is the number in the first sample.

\[ n_2 \] is the number in the second sample.

\[ s_p^2 \] is the pooled estimate of the population variance.

The number of degrees of freedom in the test is equal to the total number of items sampled minus the number of samples. Since there are two samples, there are \( n_1 + n_2 - 2 \) degrees of freedom.

The following formula is used to pool the sample variances. The two factors make up the weights: the number of observations in each sample and the sample variances themselves:

\[
s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}
\]

where:

\[ s_1^2 \] is the variance in the first sample.

\[ s_2^2 \] is the variance in the second sample.

When the sample size is greater than 30, ‘z’ test is used.

\[
Z = \frac{\bar{X} - \mu}{s/\sqrt{n}}
\]
Step #4 Formulate the decision rule

The critical values of t-test are given in a standardized table which also indicates the degrees of freedom (df). Degrees of Freedom is the number of items in a sample that are free to vary. Suppose there are two items in a sample, and we know the mean. We are free to specify only one of the two values because the other value is automatically determined (Since the two values total twice the mean). If one wants to show that the experimental group progressed much more than the controlled group, we need the standard deviation of the difference of gain and means of both the groups. That is why it is essential to use the 't' test.

Step #5 Arrive at a decision

On the basis of all the steps described above, it becomes possible for the investigator to arrive at a conclusion regarding the acceptance or rejection of the hypothesis.

In the research study in question, the null hypothesis was rejected. The investigator came to a conclusion that the traditional methods as well as experimental methods both improve student performance. However, between the two, the experimental method generated better results.
This section deals with the specific design of the experiment. It goes over and reviews briefly the curriculum used for the experiment and how it was implemented by using the various teaching strategies. It discusses the tests used to measure the significance of the data, and the scoring procedure used to procure the raw data. The data was then turned into a formula to be used to measure the standardized score.

3.6 Design of the study

The following is the breakdown of the objectives, procedures, samples, and methods.

A. Objectives:
   1. To understand the problem
   2. To develop techniques and strategies for teaching ESL
   3. To try new techniques on two groups of students to collect the statistical data
   4. To analyze the data
   5. To report the findings

B. Procedures:
   1. Survey the immigrant scene in the USA with the specific interest in the academic problems of high school ESL students of Asian origin
   2. Review the literature, study the syllabus and textbooks, and arrive at an appropriate technique.
3. Conduct teaching sessions and administer pre test and post test

4. Analyze the data statistically using the services of a professional statistician in the field

5. Detailed presentation in the form of a dissertation

C. Sample:
   1. E5B ESL students at Cardozo High School, Queens, New York City

D. Method:
   1. Survey, interviews, formal and informal meetings
   2. Developing the techniques
   3. Field Trial of the Techniques
   4. Statistical analysis and testing of the hypothesis
   5. Reporting

The aim of the project was to understand the problems of high school ESL immigrant students to develop teaching strategies to achieve higher proficiency in English. This would eventually help them in other content area subjects. The strategies were experimented on a representative sample of students. The methods of experimental research were employed for the study. The data for the pre test and the post test were collected to carry out the research further.
3.6.1 Strategies used for the experiment

The following strategies were evolved with an aim to teach *Diary Of Anne Frank*, *Nectar In A Sieve* and *Blue Level Anthology*. Each strategy was incorporated into the study to accommodate the student population in question. Hence, each strategy was uniquely administered to suit the needs of the sample namely the experimental group.

The investigator used a different strategy for each book. She used the strategy of writing an original newspaper for *Diary Of Anne Frank*, daily journal for *Nectar In A Sieve*, and making up an original short story to teach the *Literature Blue Level* anthology book. As mentioned in chapter 4, there is no single strategy to yield optimum results for enhancing the vocabulary. More than one approach was used to attack this problem which will be discussed later on in this chapter. Over and beyond these strategies, the investigator used the following strategies. They are self-inventory, student data bank, graphic organizers, conflict web, and word splash. We shall now discuss the various strategies used in executing the research project.

A. Original newspaper

Students working in groups of two to four each, were assigned the section each student was supposed to work on, and eventually come up with a newspaper from World War II era or the futuristic newspaper with the
stories about the characters from *Diary Of Anne Frank*. Students also had an option to make an original textbook based on *Diary Of Anne Frank*. Students would then make up an audio tape of the process they went through while working on the project in groups.

This assignment served several purposes: It encouraged communication among the group members and each student had the option to volunteer to work on the section of the newspaper that he/she was interested in. Students were encouraged to use the library as part of their assignment required some research work. While audio taping the process they went through, they were able to hear themselves talk which gave them additional practice in speaking and listening in English without going through the embarrassment of speaking in front of their peers and the teacher. The main purpose served by this strategy was to get students involved into the research process, freely express their views within the group, and feel comfortable about audio taping themselves.

B. Journal writing

This strategy is commonly used in Science education. However, I decided to modify the technique and used it for language teaching purposes.

Journal writing was used while teaching *Nectar In A Sieve*. It was used in two different ways; one was a technique of free writing during the first five minutes of the period, another one was the journal entry of the
various characters in the book using a different format for each journal entry which was assigned as homework. Free writing by students was read in class to motivate students to focus on the lesson of the day.

Free writing encouraged students to write without thinking about the grammatical errors. It helped them focus on their thought processes. Students expressed and exchanged their ideas freely among their peers, and the class became student centered rather than teacher centered. It gave them a valuable practice in expressing their views in front of the class without feeling embarrassed. Journal entry of the various characters in the book helped students to visualize the same incident from the point of view of a different character. It also trained them in different formats of writing namely a letter, a report, an essay or a dialogue which was actually a practice for the RCT writing test. The purpose of this activity was to enhance the student's writing skills and oral skills.

C. Original short story

This particular strategy was employed while teaching a literature anthology called Literature Blue Level Book. After reading a few short stories and poems based on the theme of relationship, students were assigned to write an original short story using the relationship theme and in their assignment try to mix and match the characters and situations from the stories they had studied. The investigator intended to publish the short
stories in a class magazine. This strategy also enhanced students' originality and creativity writing.

D. Using supplementary material to enhance vocabulary skills

The vocabulary lessons were structured (for both the groups) and a day was set aside (Friday) for a vocabulary lesson once a week. Supplementary books like 30 Days To A More Powerful Vocabulary, English For The College Boards, Webster's Dictionary were used. The study of words was theme-centered or root-centered. More details of this strategy are given in Chapter-4.

3.6.2 The tests

This sub section reports on the tests administered as a part of the study. They were used for the purpose of collecting data regarding the effectiveness of teaching techniques. The two tests that were used in the study are pre test and post test. The scores on pre test and post test were to be used for data collection and, as a result care was taken to establish reliability and validity of these tests.

The investigator's first choice was the LAB (Language Assessment Battery), but she had to abandon the idea of using the LAB test for pre test and post test because it is a closed test and can never be given any time
besides when the ESL student actually takes it to be placed in an ESL program or in a main stream English class.

The investigator's next choice was the TOEFL test which is a standardized test which measures various aspects of language proficiency and thus an overall proficiency test which measures a foreign student's English proficiency objectively. There are three sections of the test. Section A tests student's ability to understand spoken English. Section B tests his ability to recognize standard written English. Section C measures his comprehension of standard Written English. These three sections deal with the basic components of language: listening, reading, vocabulary, and structure.

TOEFL is a significant outcome of a conference sponsored in Washington by The Center for Applied Linguistics at a meeting in January, 1962, that led to the formation of the Advisory Council on the Testing of English as a Foreign Language. What evolved as a result of this was the TOEFL test and testing programme, now operated on a world basis by the Educational Testing Service and the College Entrance Examination Board.

TOEFL is an example of discrete item test. Perhaps the largest gain made in turning to discrete point testing was specific identifications of the categories and an objectivity in scoring which was impossible with the earlier types of tests. Multiple choice items can be statistically analyzed for difficult scores, discrimination and correlation with external or internal criteria (Jadeja 1988, 132).
3.6.3 Scoring procedure

Each correct answer has an equal weightage, one point each. Total number of correct answers is called the "raw score." The raw score is converted into a scaled score by adding the sum of three section scores multiplied by ten thirds to arrive at the total score. Standard error of measurement is approximately 14 scaled score points. The test was administered according to the ETS guidelines for administering the test.

This chapter described the sample as well as the tools of measurement which included the testing of the hypothesis, and research design. As a sub section of the research design, actual teaching strategies were discussed, describing how they were tried out on Cardozo high school ESL immigrant students in New York City. The next chapter will focus on the actual experiment and how it was carried out.