CHAPTER 3
RESEARCH METHODOLOGY

Ethics in Research

The primary goal of any research must be to pursue knowledge and truth. Therefore fabrication and false use of data must be prevented. For this ethics must be followed while conducting research. This study follows the ethical code of the research. All the respondents were well informed about the purpose of the study. Informed consents of all respondents were duly taken before the collection of data. Due credits were given to past studies by appropriately citing them in this report. The entire thesis has also undergone for plagiarism check.

Research Process

The research process commences when the researcher identifies Management dilemma. This management dilemma is transformed into a research problem and this problem is then converted into research objectives. These research objectives then guide the entire research. Once the objectives are framed, the research design is finalized which then guides for data collection process and instrument development process. Then actual data is collected from identified samples and later such collected data is analysed. Interpretations are drawn from the result of analysis and based on these interpretations recommendations may be given to the management.

This research study follows this scientific process for arriving at the framed objectives.
Research Design

Descriptive research design has been used for this research because it was found to be most appropriate. Descriptive research design is a scientific way of observing and describing the relationship of variables without influencing them in any way. In this study variables such as demographic variables, employability and socio-economic status were observed for their relationship. None of the variable under the study were influenced or manipulated.

1. Data Collection

Both primary and secondary data are collected in this study.

Sources of Primary data

Students, their parents or local guardians and HR managers of organizations wherein the Learn and Earn Scheme is run are the sources of primary data. Primary data is collected through a structured questionnaire using an interview technique. Three different sets of the questionnaire are used for different sample units.

Besides the interview, observation is also used as a tool for collecting primary data. Research assistants helped in collecting data from samples after receiving training from the researcher.

Sources of Secondary data

Secondary data is collected from existing literature relevant to the subject in the form of research articles, papers, journals, magazines, newspapers etc.
2. **Sampling Frame**

Learn and Earn program is no more active course for students in Maharashtra. Due to some reasons, no new admissions will be possible to the Polytechnic course under Learn and Earn program from the academic year 2017. However, so far 2 batches have successfully passed this program and 2 more batches would be undertaking this course. Out of these 2 batches one batch (4th Year) will complete its program in 2018 and the other batch (3rd Year) will complete this program in 2019.

This study focuses on employees who have either completed or pursuing Polytechnic education and working in manufacturing companies.

3. **Population**

As of now Yashaswi Education Society’s Yashaswi Institute of Technology is the only institute which in association with Government of Maharashtra runs the Learn and Earn Scheme (Vocational Education) for 10th passed and 12th failed students who for one or the other reasons could not continue higher studies. As the data received from Yashaswi Institute of Technology Pune, around 17000 students are enrolled in this learn and earn program. Some of these students have successfully passed the program and continued working in manufacturing companies. In addition to these students, many other people have completed their Polytechnic education from the traditional education system and work in a manufacturing company. Out of these 17000 students, those who reside in and around Pune are only 5000. For these 5000 students there would be around 10000 parents/ local guardians. As of now, there are around 60 manufacturing companies in and around Pune which employ Learn and Earn enrolled candidates.
To summarize the total population size for this study is as follows-

i. Number of employees who either enrolled in Learn and Earn or completed the traditional polytechnic program- >17000

ii. Number of Local Parents- 10000 (From the authorities of Yashaswi Institute of Technology it was understood that about 29% of all enrolled students are from Pune (local) and live with their parents/ guardians. 29% of 17000= 4930. This figure is rounded off to 5000, thus sample size of local parents= 5000*2= 10000.

iii. Number of Human Resource Professionals- 60

4. Sample Unit

There are three different sample units in this study.

i. Employees who have either completed or pursuing polytechnic education

ii. Students’ Parent

iii. Manager of Human Resource Department of organization which employs students under learn and earn scheme.

5. Sample size

The sample size should be large enough to make the findings of the study generalizable. The sampling error, however, will always be there irrespective of the size of the sample. Sampling error represents the power of samples in representing the entire population. Larger the sampling error smaller will be the power and vice versa. The sampling error can be reduced by increasing the sample size. To make sampling error zero the sample size must be equal to population. However, this is neither feasible nor needed in many cases. Therefore samples are drawn from the population.
To generalize the findings of the study it is very important to calculate sample size with precaution and based on scientific techniques. Therefore care was taken while framing the sample size for this study. Three components are very crucial while calculating the sample size. These components are described below.

6. **Level of Precision**

Level of precision is the measure of how close the estimate is to the actual characteristic of the population. It may be termed as sampling error. Sampling error depends on the amount of the risk researcher is willing to take while making inferences from the data. It is expressed in percentage.

Higher the level of precision higher will be the chances of results of the study being closer to the actual populations’ characteristics. For this, the sample size has to be larger.

7. **Confidence level**

The confidence or risk level is ascertained through the well-established probability model called the normal distribution. Generally, normal distribution occurs whenever a large number of independent factors influencing the outcome are collected. In other words, if the sample size is sufficiently large then the distribution of such samples is normal. Usually, 95% and 99% of probability are taken as the two known degrees of confidence for specifying the interval within which one may ascertain the existence of population parameter (e.g. mean). 95% of confidence level means that if the researcher takes 100 samples from the same population then 95 samples out of 100 will meet the estimated value of a variable within the precision set by him. While calculating the sample size, the desired confidence level is specified by the z value. The z value is the value of abscissa along with the standard normal distribution. It can be known from the table of the normal curve.
8. **Degree of variability**

The degree of variability in the attributes being measured refers to the distribution of attributes in the population. The more heterogeneous a population, the larger the confidence is 99%, then it means out of 100 samples 99 cases will be within the error of tolerances specified by the precision. For less variable (more homogeneous) population, smaller sample sizes are appropriate. It should be noted that a proportion of 50% indicates a greater level of variability than that of 20% or 80%. This is because 20% and 80% indicate that a large majority do not or do, respectively, have the attribute of interest. Because a proportion of 0.5 indicates the maximum variability in a population, it is often used in determining a more conservative sample size.

9. **Sample size calculations**

Formula developed by Cochran (1977) to calculate a sample size is

\[
 n = \frac{n_0}{1 + \left(\frac{n_0 - 1}{N}\right)} \quad (3.1)
\]

Where \( n_0 \) the sample size is calculated using formula developed by Cochran (1977) for infinite population size. It can be calculated using following formula-

\[
 n_0 = \frac{z^2 pq}{e^2}
\]

Where

\( n_0 \)= sample size

\( z \)= critical value of the desired confidence level

\( p \)= the estimated proportion of an attribute that is present in the population

\( q \)= 1-\( p \) and
e= desired level of precision

Since the degree of variability is not known the maximum variability is assumed for this study and thus p=0.5 (50% variability). Taking 95% confidence level with ±5 precision the sample size can be calculated as-

\[ p=0.5, \quad q=1-p=0.5, \quad e= 0.05 \quad z= 1.96 \text{ (at 95% confidence level)} \]

So \[ n_0 = \frac{(1.96)^2 (0.5)(0.5)}{(0.05)^2} \]

\[ n_0 = 384 \]

Since the total number of employees working in manufacturing industry that either enrolled in Learn and Earn Scheme or completed polytechnic education through traditional education system is not known with certainty the sample size for such employees can be fixed to 384. To round up this odd number researcher has taken 400 samples.

Now using this value final sample sizes for parents and HR managers are calculated using equation 3.1 as follows

Sample size for parents

\[ n_2 = \frac{384}{1 + \frac{(384 - 1)}{10000}} \]

\[ n_2 = 369 \]

Sample size for HR professional

\[ n_3 = \frac{384}{1 + \frac{(384 - 1)}{60}} \]

\[ n_3 = 52 \]
Finally sample sizes for all sample units under the study are given below

\[ n_1 = 400 \text{ Students} \]

\[ n_2 = 400 \text{ Parents (Rounded up)} \]

\[ n_3 = 52 \text{ HR professionals} \]

10. **Sampling technique**

Non-probabilistic Purposive sampling technique was used for selecting samples from populations. Students from different levels of the course were selected to make the sample size a good representative of the entire population. Parents or local guardians of only already selected students were chosen as samples. HR managers of organizations of which students had been selected were considered.

11. **Development of research instrument**

Three different questionnaires are developed for data collection. One is used to collect data from students enrolled in Learn and Earn Scheme. Other two is used to collect data from HR managers and Parents of enrolled students. With a consultation with a research guide and the experts in the field. Existing relevant literature is also reviewed to prepare a valid questionnaire. The validity of the questionnaire was checked by conducting a pilot study as described in point number 3.5.4.
**Instrument for students/employees**

Questionnaire for students/employees includes 9 items. All of these items were rated on 5 points Likert’s scale. Where 1 means ‘Strongly Disagree’ and 5 means ‘Strongly Agree’. In addition to these items, demographic information of students is also collected through forced-choice questions.

**Instrument for parents**

For parents also 9 items were included in the subscale. Parents were asked to report their agreement about the statement on 5 points Likert’s scale. On this scale 1 meant ‘Strongly Disagree’ and 5 meant ‘Strongly Agree’.

**Instrument for HR professionals**

Human resource professionals of sample organizations’ were asked to respond to the questionnaire which was rated on 5 points Likert’s scale. On this scale 1 meant ‘Strongly disagree’ and 5 Meant ‘Strongly Agree’. There are a total of 6 items in this subscale. The questionnaire was designed to understand the HR managers’ perception of employees from vocational education and traditional education. To make the balance HR managers/professionals were asked to presume that they were comparing these students/employees have the same amount of on the job training.
12. Pilot study

The questionnaires were tested for their reliability by conducting a pilot study. For the pilot study, 55 samples were selected from students/employees, 50 samples were selected from parents of polytechnic students. Further, 20 HR professional was selected to test the reliability of the questionnaire designed for HR professionals. For testing the reliability of the questionnaires, all collected data were coded in SPSS and Cronbach’s alpha coefficient was investigated.

Table 4

Reliability statistics for all three questionnaire

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Cronbach's Alpha</th>
<th>N of Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>For HR Professionals</td>
<td>.743</td>
<td>11</td>
</tr>
<tr>
<td>For Students</td>
<td>.819</td>
<td>17</td>
</tr>
<tr>
<td>For Parents</td>
<td>.640</td>
<td>15</td>
</tr>
</tbody>
</table>

In social science, Cronbach’s alpha coefficient of equal to or more than .7 is considered as indication of a reliable scale (questionnaire) (Nunnally (1979). The value of Cronbach’s Alpha of the questionnaire used for parents is closed to 0.7 (.640). The values of the other 2 scales’ Cronbach’s alpha are greater than .7. This indicates that the questionnaires are reliable and can be used further for data collection. Therefore the questionnaires aroused for final data collection.
13. Hypotheses formulation

Based on the pilot study and literature review the hypotheses framed during the proposal of this study are confirmed and considered for further analysis. The hypotheses of this research study are-

$H_1$: There is a significant correlation between vocational training and employability of students.

$H_2$: There is a significant correlation between vocational training and the socio-economic status of students.

$H_3$: There is a significant correlation between vocational education and training cost.

14. Tools for data analysis

Primary data so collected, was analysed in Statistical Package for Social Science (SPSS). The data is coded in SPSS for analysis purpose. Appropriate statistical tests were used for testing hypotheses.

*Statistical Tests used for data analysis*

*Mean*

Mean is an average which is being used to understand the central tendency of any data. In this study mean was used for deriving the central tendency of respondents on a particular scale. To be specific mean was used in the descriptive analysis of various subscales used in this study.

*Standard Deviation*

In any data set standard deviation is used to examine the amount of variation in the data. In other words, it shows the dispersion of data set values. Higher the standard deviation value higher is the variation or dispersion of data around the mean. Mean of the data set can be
significant if standard deviation value is less. Generally, the standard deviation value close to zero is desirable in any data set. This study used the standard deviation values to determine the significance of Mean values of subscales.

**One-way ANOVA Test**

One-way ANOVA is a parametric test. One way ANOVA tests if the means of two variables differ significantly or so. ANOVA makes use of the F test. The null hypothesis of the F test assumes that the means of two variables do not differ from each other significantly. If the significance value of the F test is less than 0.05, then the relationship between two variables can be considered significant as the F tests fail to accept the Null Hypothesis. As the variables under investigation in this study were parametric, to test if their means differ significantly one way ANOVA was used.

**Pearson’s Correlation Test**

Pearson’s correlation is also a parametric test. Pearson’s correlation test shows if two or more variables are correlated and such correlation is statistically significant. To test the hypotheses of this study Pearson’s correlation test used.

**Tests used for hypotheses testing**

Testing hypotheses of a study is a crucial part of any research. Therefore it must be done with precaution and with the help of science. Table 5 gives a justification of tools used for hypotheses testing.
Table 5

*Justification of tests used for hypotheses testing*

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Test</th>
<th>Justification</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>H₁</strong> - There is a significant correlation between vocational training and</td>
<td>Pearson’s Correlation</td>
<td>First, all the variables in the hypotheses are parametric therefore Choosing parametric test was essential.</td>
</tr>
<tr>
<td>Employability of students.</td>
<td>test</td>
<td>Second, to test if these variables are significantly correlated or so this test was thought to be appropriate.</td>
</tr>
<tr>
<td><strong>H₂</strong> - There is a significant correlation between vocational training and</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Socio-economic status of students.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>H₃</strong> - There is a significant correlation between the Salaries of</td>
<td></td>
<td></td>
</tr>
<tr>
<td>technicians paid by organizations and number of technicians who undertook</td>
<td></td>
<td></td>
</tr>
<tr>
<td>vocational Training at those organizations.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>