Chapter 4 Research Methodology

It is clear from the literature review, as given in Chapter 2, that well formulated studies on impact of PBL approach to teaching and learning, OR and Statistics in particular, evaluated in comparison with traditional teaching are rare. At school levels there seems to be a huge number of recorded attempts. Research attempts on choices of techniques (Sections 2.8, 2.10) also are seen to leave a lot of ground yet to be covered. Use of data mining analysis is one area not explored to its full potential. Hence the present study of Data mining approach to analyze impact of PBL on learning of Statistics and Operations Research is undertaken.

In this chapter the research methods adopted are briefly described. Section 4.1 discusses overview and the research questions. The research hypothesis and objectives are presented in section 4.2. Research method adopted and experimental set up for present research is explained in section 4.3. Data description, collection and analysis plan is presented in section 4.4. Section 4.5 presents the summary. The research questions related to the research problem are discussed in the next section.

4.1 Overview

Several research questions were posed for analyzing students’ learning outcomes with respect to Problem Based Learning (PBL), their background, engagement and attitude towards statistics & operations research. The major questions were

Research question 1: Does PBL affect Knowledge in Statistics and Operations research?

Research question 2: Does PBL affect Skill in Statistics and Operations research?

Research question 3: Does PBL affect Attitude in Statistics and Operations research?

In this regard the most fundamental question that arose was, ‘how can one estimate the Knowledge, Skill and Attitude of a student’. This question has been addressed to reasonable extent as details in later sections reveal.

Research question 4: Are the Students in the class really heterogeneous with respect to state, year of birth, social category, board at 12th standard, marks obtained in 12th standard and mathematics, and location within a country?
Research question 5: How learning of students is related to background of student like gender, category, state, discipline, board and math background, semester wise.

Tracking the cause-effect relationships from students’ pre-admission attributes, and then tracking the dynamic changes/ transitions that take place, how the system helps student evolve into a strong learned-human being coming out of the system, is another aspect that is addressed.

The researchers are basically interested in which, amongst the various indicators are related to student background such as math and no math students, science-no science students and student engagement such as attendance, faculty interaction are significant as the learning proceeds from sem1 to sem3. The researchers also track the learning outcomes in terms of KSA as and when students’ are exposed to PBL.

Laying down proper and meaningful hypothesis was naturally an important step.

4.2 Proposed Hypotheses

Following research hypotheses have been proposed and tested:

Regarding curriculum:

1. $H_0$syllbi: Syllabi of OR and Statistics as taught to BCA and BBA in Pune University, SNDT University and Symbiosis International (Deemed) University do not differ significantly.

$H_1$syllbi: Syllabi of OR and Statistics as taught to BCA and BBA in Pune University, SNDT University and Symbiosis International (Deemed) University differ significantly.

Regarding the important construct Student Background (SB) from the conceptual model:

2. $H_0$SB: Student demography factors do not have significant effect on performance of Students in Statistics and Operations research

$H_1$SB: Student demography factors have a positive significant effect on Performance of Students in Statistics and Operations research;

Regarding the important construct SE from the conceptual model:

3. $H_0$SE: Student engagements do not have significant effect on performance of Students in Statistics and Operations research

$H_1$SE: Student engagements under the treatment of PBL have a positive significant effect on Performance of Students in Statistics and Operations research;
Regarding the pedagogical treatment

4. $H_0^{PBL}$: PBL treatment does not have significant impact on overall Performance of Students related to Statistics and Operations research

$H_1^{PBL}$: PBL treatment has significant positive impact on Performance of Students related to Statistics and Operations research;

Regarding the most important aspect of measurements mentioned in conceptual model (Fig 3.11) that PBL helps in building Knowledge, Skill and Attitude:

5. $H_0^K$: PBL does not affect Knowledge gain related to Statistics and Operations research significantly

$H_1^K$: PBL affects positively Knowledge gain in Statistics and Operations research

6. $H_0^S$: PBL does not affect significantly Skills related to Statistics and Operations research

$H_1^S$: PBL affects Skills related to Statistics and Operations research

7. $H_0^A$: PBL does not affect significantly the Attitude towards Statistics and Operations research

$H_1^A$: PBL affects Attitude positively towards Statistics and Operations research

4.3 Research Methods

The research methodology used for present research is essentially a combination of different approaches selected as per objectives and context.

A broadly modified Experimental approach was adopted which as applicable to data from in-situ situation. In Situ is a situation where researcher has to take all data as it comes from the system, and try to draw inferences from it. It is a most natural and no external intervention type of approach.

To identify the impact of PBL, strategies are designed using data mining approach, which are more powerful and can extract hidden patterns that are not otherwise possible using normal statistical analysis.

A most important component of pre-post evaluation based research, mainly in the context of experience of students was incorporated. The pre post evaluation, in the form of a planned
tests before and after in each PBL implemented semester was a special part implemented to enable learning outcome measurements.

All this research has a strong Conceptual modeling support deduced from learning theories. Details of each of these components are given below:

4.3.1 Details of Experimental part

Experimental part is all creating teaching learning environment with PBL and without PBL, so that learning outcomes could be gathered, analyzed and inferred upon. Researcher being a part of the teaching-learning system, introduction of PBL as a treatment, was the first step.

Table 4.1 shows the total number of participating students considered in the experimental plan.

<table>
<thead>
<tr>
<th>Course</th>
<th>Admissions</th>
<th>No. of Records</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre admission</td>
<td>Sem1</td>
</tr>
<tr>
<td>BBA(IT)</td>
<td>508</td>
<td>302</td>
</tr>
<tr>
<td>BCA</td>
<td>343</td>
<td>213</td>
</tr>
<tr>
<td>Total</td>
<td>851</td>
<td>515</td>
</tr>
</tbody>
</table>

The whole experimental plan was In-Situ. There are reasons why we consider it so, and why it is considered adequate for our objectives.

Experiments can generally be classified as designed experiments where the experimenter introduces treatments, replications and local control specifically to suit the objectives he is tackling, and in-situ experiments (include quasi experiment), where the system would be running in its natural manner, and the researcher is expected to extract metrics and measurements without destruction or externally influencing it. In randomized experiments the experimenter conducts randomized control trials. However this experiment being an “In Situ” situation, experimenter could not introduce exclusive control cases. Randomization was not introduced but was natural system generated and accepted as it is. A quasi-
experimental design by definition lacks random assignment (White & Sabarwal, 2014). The present study lacks randomness because it was conducted on students’ as they were assigned to researcher. Hence it is closer to quasi experiment.

The learning outcomes of students were noted down, from the existing system, for all students of six batches. Freedom to introduce PBL was again a part of natural policy of teaching provided by the system. Hence the whole batch allocated to the teacher goes through the process. The researcher has collected data regarding students behavioral aspects, numerical skills and related metrics feeling motivated in learning, expectations and interests through pre-posts evaluations using tests, which included questions on behavioral aspects in addition to subject related questions (See Appendix F1-F5, for a sample test instrument)

4.3.2 PBL Treatment Process

For the present study, PBL as pedagogical treatment is formulated with incorporation of definition by Schmidt (1983) and assessments discussed in literature review.

1. study of real or hypothetical problem from domains like education, service industry, manufacturing, transportation, events happening in campus,
2. small groups (4-5 students),
3. collaborative study,
4. hypothetico-deductive reasoning,
5. faculty direction on group process than imparting information
6. presentation of problem undertaken
7. filling evaluation form after presentation (1-5, as discussed by Schmidt, 1983)

The facilitator’s strategy/ methodology for PBL employed, was as per five stage plan, out of which stages 1, 2 and 5 are used for data collection from experimental sample. Stage 3 and 4 are used for execution of the planned pedagogy. The five stages are

Stage 1: Collect Students’ demographic data at the beginning of semester
Stage 2: Evaluate students’ knowledge, skill and attitude using a questionnaire
Stage 3: Define Semester assessment plan
Stage 4: Execute PBL through following steps by giving guidelines to students:
Step 1: to form groups with their choice
Step 2: to select real life scenario and to identify the problem
Step 3: to define problem and define objective
Step 4: to gather information, relevant data and formulate problem
Step 5: for analysis and interpretation of data
Step 6: for report writing and presentation

**Stage 5:** Evaluate Learning outcomes of students’ regarding knowledge, skill and attitude using a questionnaire post PBL, PBL report and presentation.

Table 4.2 and 4.3 give details of the batches of students who were subjected to PBL treatments.

**Table 4.2: Design for PBL Evaluation Data Collection of BBA(IT) Students**

<table>
<thead>
<tr>
<th>Semester</th>
<th>Pre Evaluations</th>
<th>Post Evaluations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre Test Survey</td>
<td>Post Test Survey</td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>B4 B5 B6</td>
<td>B4 B5 B6</td>
</tr>
<tr>
<td>2</td>
<td>B1</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>B4 B5 B6</td>
<td>B4 B5 B6</td>
</tr>
</tbody>
</table>

**Table 4.3: Design for PBL Evaluation Data Collection of BCA Students**

<table>
<thead>
<tr>
<th>Semester</th>
<th>Pre Evaluations</th>
<th>Post Evaluations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre Test Survey</td>
<td>Post Test Survey</td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>B4 B5 B6</td>
<td>B4 B5 B6</td>
</tr>
<tr>
<td>3</td>
<td>B1</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>B1 B3</td>
<td>B1 B3</td>
</tr>
<tr>
<td></td>
<td>B4 B5 B6</td>
<td>B4 B5 B6</td>
</tr>
</tbody>
</table>

The objective of implementing PBL along with traditional teaching is that student should be able to apply the concepts learn in the class to real life situations. This refers to the third level of Bloom’s taxonomy. Pune University and Symbiosis International University have used Bloom’s taxonomy to define learning objectives and level of syllabus hence authors have adopted Bloom’s taxonomy as framework to define students’ learning outcome.
The evaluation methods as adopted by the institute had lot of freedom to choose from various combinations as a standard policy. Hence, researcher could implement three evaluation methods proposed by Major et al. (2001) out of six, in the present study. Outside Evaluation by Experts, Content Analysis of Projects and Personal Reflections form important methods in the current implementation.

Outside Evaluation by Experts was done for 4 batches out of 6 batches for students’ presentation. Content Analysis of PBL reports was done by facilitator on the basis of different resources, writing skills, analysis, discussion and conclusion for project. The personal reflections were used to assess the metacognitive skills of students’ learnt through PBL. Students were judged using pre and post evaluations form containing questions which help in measuring the change in the outcomes such as interest, critical thinking etc.

4.4 Data Description and Collection Plan

Two types of data, structured and unstructured, were used. The unstructured data was in form of syllabus, project records as maintained by the instructor. The structured data comprised of student details from two sources, official records of the students and pre-post evaluations conducted of PBL participants. Following figure represents data collection plan where all Marks, GPAs, Attendance are gathered as it is through from Sem1 to Sem3.
Figure 4.1: Complete Data Collection Plan Sem1 to Sem3
4.4.1 Collection of Data

Syllabi were collected from three universities via internet and in person. The data is collected for period 2010 to 2016 for six batches of BBA(IT) and BCA each. The data consists of results of students, their background and demographic information from the questionnaire and records of institution database, if not available from feedback. The list of variables is defined in Chapter 5.

All the students of BBA(with different specialization like HR,IT etc.) and BCA, who are taught Statistics and/or Operations Research in colleges of Pune offered by various universities is the population under study. There are five universities in Pune namely Pune university, SNDT University, Dr. D. Y. Patil University, BharatiVidyapeeth Deemed University and Symbiosis International (Deemed) University (SIU) offer BBA and BCA course. SNDT University offers only BCA course whereas remaining universities offer, both BBA and BCA courses. This forms the population who are the beneficiaries of the study (Appendix B1). The syllabus of Statistics & Operations Research at these colleges will be compared. The syllabus for BharatiVidyapeeth Deemed University and Dr. D. Y. Patil University were not available. Hence could not be considered for similarity analysis.

The research is framed within regular course of “Statistics” and “Operations Research” which is offered to undergraduate techno-management first year students, in face to face learning environment. All the students of the BBA(IT) & BCA sem1 at Symbiosis Institute of Computer Studies & Research, (SICSRS) Pune will be a part of sample under the treatment. The learning outcomes as per literature review for PBL suggests that students should be able to think critically, analyze and solve complex, real-world problems, to find, evaluate, and use appropriate learning resources, to demonstrate effective communication skills, to work cooperatively and to use subject knowledge. Hence as part of PBL implementation evaluations were also conducted using questionnaires to collect learning outcomes in addition to semester end performance. The questionnaires were designed on the basis of the learning outcomes as mentioned above. Hence questionnaire is used as an instrument for evaluation of learning outcomes.
Three evaluations were conducted in three phases. First evaluation was conducted to get student background information, their attitude (Appendix F1-F6) and knowledge, based questions on statistics and operations research by designing single questionnaire. The knowledge based questions on statistics and operations research of this questionnaire was used to conduct posttest to judge the change in knowledge level of students. The third questionnaire was designed to know the learning of students from the problem under taken for PBL.

The pre admission variable such as gender, state, category, discipline at 12th, board at 12th, math back ground, nationality, age, are classification variables with conditions that existed prior to the study. The 10th marks, 12th marks, 12th math marks are indicators of students’ academic background. The researcher cannot change the characteristic of the variable, so they study the phenomenon being explored under several conditions of the various aspects of the variable. Each independent variable may have two or more levels. Levels or classifications of independent variables are sub categories of the independent variables used by researcher in the experimental design. A dependent variable is the response to the different levels of the independent variables. It is the measurement taken under the conditions of the experimental design that reflects the effects of the independent variable(s).

The pre and post evaluations were conducted at the beginning and at the end of semester only to judge the basic knowledge of subjects Statistics & OR and to study the efficacy of PBL. Questions based on guidelines given by NSSE (NSSE, 2003) are also considered to evaluate students. The questionnaire was prepared to test above said characteristics and knowledge based questions to cover the syllabus of Statistics and Operations research at SIU. The other learning outcomes (internal and external evaluation results) throughout the semesters were collected.

4.4.2 Data Preprocessing

Data sources were predominantly form the system records, since the experiment was In_Situ in nature. Records which came and were collected, required preprocessing to some extent due to cases of missing records, imbalances created due to uneven promotions of students from semesters, dropouts and many similar reasons. Change of measurements like change from 4-point to 10 point scale was one of the major aspects.
As this research consists of students’ data for six batches of two courses BBA(IT) & BCA, out of which performances of students’ of first five batches are measured on 4 point CGPA scale and performance of students’ for last batch (Batch6) for both courses is measured on 10 point CGPA. Hence performances of Batch6 students were converted to 4 point scale from 10 point scale by multiplying 10 point GPA by 0.4. ANOVA were carried out for absolute scores and z-scores of performances (Entrance score, GPA SEM1, SEM2 & SEM3). Z-scores are calculated batch wise for each course separately. ANOVA for all batches of BBA(IT) and BCA were carried out by considering 4-point GPA. Standard deviations are calculated using sample standard deviation. The policy to deal with missing observations is discussed wherever it is used.

4.4.3 Data Analysis Plan

Syllabi are documents containing textual matters. Hence Text Mining is the appropriate method to analyze. The syllabus contents were subjected to similarity analysis. The results of similarity analysis are presented in Chapter 5. These results are used before performing ANOVA.

For this study the syllabus of Statistics and Operations research for BBA course of Pune University, SNDT University and Symbiosis International University are compared. The Statistics and Operations research syllabus of Bharati Vidyapeeth and D Y Patil (Deemed University) could not be availed.

The data analysis plan for structured data is as follows:

1. Descriptive Statistical Analysis
2. ANOVA and Factor analysis
3. Clustering
4. Classification
5. Statistical analysis of feedback data conducted

The analysis of demographic data carried out is given in Chapter 5 and ANOVA is discussed in Chapter 6. Clustering and Classification are discussed in Chapter 7. Analysis of questionnaires to extract knowledge, skill and attitude is discussed in Chapter 8. It is assumed that essentially there are only 3 variables (indices of K, S, A) which define the status of a student. and places him in the three dimensional space [0,1]x[0,1]x[0,1], so that the points in the cube (0,0,0) corresponding to a hopeless case to [1,1,1] corresponding to the best ever possible case. Each student has an initial (x,y,z) at the time of entry and which follows
a path \((x(t), y(t), z(t))\) as time progresses. Researchers feel that \(d(S_1, S_2)\) = distance between math and no math students progressively decreases as we move towards Sem 3, some due to basic pedagogy of the institute and some (additional) due to PBL. Data mining techniques are used to extract these changes.

### 4.4.5 Software tools and environment used.

Several tools were used for data analysis. Advance tool Pack of Excel (2010) is used for descriptive statistics, SPSS (18) was used to run most of the ANOVA, AMOS(22) for Path analysis, R(3.3.2) for processing text mining and WEKA(3.8.1) was used for Classification and Clustering. The text mining in R is done using R-studio. The packages used for preprocessing documents and generating Dendrogram, are TM, stringr and hclust along with functions cosine distance and entropy.

### 4.5 Summary

The conceptual model (Chapter 3, Eqns(1.4), Figs 3.1-3.8) forms the basis of our investigations. The methodology adopted was not a Designed Experiment using principles of DoE, but was an analysis of In Situ situation, which still had all provisions for comparing treatments. Tracking of students performance from entry to Sem1 to exit to Sem 3 was done through two Cause-effect sequential models one using descriptive statistics and other using Data mining approach. Initially students’ performances are analyzed batch wise and then by classifying them according to academic learning outcome CGPA with respect to student background and student engagement. Data mining techniques would be used to extract hidden patterns from data apart from basic descriptive statistics.

Exact definitions of learning outcomes, measures of which formed the data are not detailed in this chapter and many would be defined at appropriate places in the chapters to follow.