Research Methodology

Research is to see what everybody else has seen, and to think what nobody else has thought. – Albert Szent-Gyorgyi
Research Methodology may also be defined as a process of systematic enquiry into a subject in order to discover or revise facts, theories, applications etc. with the use of method or body of methods (Taylor, et al, 2006). Thus, the discussion on this will focus on the selection of data sources, methods of data collection, methods of data analysis and justification for these selections with the help of existing literature. There are three major methods for conducting research:

1. Quantitative
2. Qualitative
3. Mixed method

Quantitative research or a positivist paradigm has been the foundation stone of social-science research. This guides researchers to “eliminate their biases, remain emotionally detached and uninvolved with the objects of study and test or empirically justify their stated hypotheses” (Johnson & Onwuegbuzie, 2004).

Qualitative purists support a constructivist or interpretivist paradigm and believe that “multiple-constructed realities flourish, and time and context free generalizations are neither desirable nor possible. Research is value bound and it is impossible to differentiate causes and effects fully, logic flows from specific to general and knower and known cannot be separated because the subjective knower is the only source of reality” (Johnson & Onwuegbuzie, 2004).

Mixed Method is a combination of both qualitative and quantitative methods. For this study mixed method has been considered. Thus, it has been explained in detail.
below. Apart from this, both qualitative and quantitative methods have also been discussed.

### 3.1 Mixed Method

The debate of qualitative verses quantitative has been going on for years. This debate brings into play a classic 'paradigm war' with both the sides giving arguments in the support of their method and weakness of the other.

As quoted by Miles & Huberman (1994), Donald Campbell says "All research ultimately has a qualitative grounding", whereas Fred Kerlinger says "There's no such thing as qualitative data. Everything is either 1 or 0".

To understand this debate, it is essential to understand the origin of both the paradigms. The quantitative research emerged in physical sciences like Biology, Chemistry, Physics etc., where it was important to measure and generalize the findings. The observations had to be objective and reproducible.

Later, when fields like Sociology, Psychology, Anthropology etc. emerged, which were concerned with the human beings and their behaviors, it became difficult to analyze the phenomenon with quantitative research only. It also became increasingly complex to explain the ‘why’ of the human behavior. Amidst this, qualitative research came into being. It increased the understanding of the social world. It helped in explaining the ‘whys’ of social phenomenon.

According to Conger (1998) the quantitative method alone cannot explain the phenomenon in totality. The qualitative method along with quantitative method can facilitate the emergence of theory. Smith (1975) also suggests, “qualitative analysis
deals with the forms and antecedent-consequent patterns of form, while quantitative analysis deals with duration and frequency of form”.

The researchers started recognizing the importance of applying both the methods together. Many social-scientists started believing that there is no major problem area that should be studied exclusively with one research method (Terrell, 2012). This thinking gave acceleration to the research technique called ‘Mixed Method’.

Mixed methods are the combination of different research methods. Studies that combine the qualitative and quantitative approaches within different phases of the research process are called mixed method (Tashakkori & Teddlie, 2008).

Everet and Louis (1981) clarify that each of the distinguished two research stances, “inquiry from the outside”, often implemented via quantitative studies and “inquiry from the inside” via qualitative studies, are complementary to each other. For developing better understanding of a topic, it can be studied simultaneously (triangulation) or concurrently with both methods (mixing quantitative and qualitative methods at the same time or in cycles, depending on the problem) (Ospina, 2004). It explores and explains phenomenon holistically and gives both outside & inside view of the object of study.

The logical and intuitive appeal as well as its working as a bridge between the qualitative and quantitative paradigms has increased the number of researchers who are utilizing mixed method research to undertake their studies (Onwuegbuzie & Leech, 2006). There is a surge in the usage of mixed method to study the phenomenon.

Collins et al. (2006) have identified the following four major rationales for mixing quantitative and qualitative approaches:
i) Participant enrichment (It optimizes the sample; such as increasing the number of participants or increasing the information from the participants),

ii) Instrument fidelity (i.e., maximizing the appropriateness and/or utility of the instruments used in the study)

iii) Treatment integrity (i.e., mixing quantitative and qualitative techniques in order to assess the fidelity of interventions, treatments, or programs)

iv) Significance enhancement (i.e., mixing quantitative and qualitative techniques in order to maximize researchers’ interpretations of data).

The area of social entrepreneurship is still developing its own theoretical underpinning. Usage of mixed method can enhance the understanding of the area for a lucid interpretation of the data. This idea of deeper understanding of the problem at hand has been instrumental for using mixed method for this research.

As discussed earlier, mixed method has its origin in both the paradigms of research i.e. Quantitative and Qualitative. The combination of these research methods yields various methods of research or types of mixed method research. There are six types of mixed method research identified by Terrell (2012).

1) Sequential Explanatory Strategy – In this strategy, the collection and analysis of quantitative data is followed by the collection and analysis of qualitative data. The main aim is to explain quantitative results by exploring certain results in more detail or explaining unexpected results through qualitative research. This is a method with distinct stages and equal priority to both the phases. However, it is a time consuming process.
2) Sequential Exploratory Strategy – This strategy involves collection and analysis of qualitative data followed by the collection and analysis of quantitative data. This is used for exploring a phenomenon through testing elements of a theory or generalizing the qualitative finding for different samples or developing an instrument of measurement of the phenomenon (e.g., using a small group to create instrumentation and then collecting quantitative data based on the instrumentation) (Terrell, 2012). This is also a method of distinct stages, so it is easier to analyze but it is also time consuming.

3) Sequential Transformative Strategy – According to this strategy, any research method (i.e. qualitative and quantitative) can follow another. The results are assembled at the time of interpretation. Here, the sequence of the research method is decided on the basis of a theoretical perspective or conceptual framework of the research.
4) Concurrent Triangulation Strategy – Both the research methods are applied simultaneously and the results are compared at the stage of data interpretation. This is used for the confirmation and cross-validation in a single study. However, the method with both types of data could be confusing for new researchers.
5) Concurrent Nested Strategy – In this research technique, one method is embedded in the other and it gives a broader perspective. While one method explores the phenomenon, the embedded method may address few of the research questions or gather more information within the topic of research. This is applied concurrently with the other pair of the main and embedded method. Then the results are combined at the interpretation level. This also can be difficult to implement for novice researchers.
6) Concurrent Transformative Strategy – This strategy employs nested approach along with the transformative research method based on a theoretical perspective. The data collected can be combined at the analysis stage. This takes lesser time than sequential techniques but it requires expertise because of the indistinct stages of data collection.

As social entrepreneurship research is still largely phenomenon driven (Mair & Marti, 2006), so for this research, Sequential Exploratory Strategy has been used. Qualitative research phase has been followed by a quantitative phase. The exploratory qualitative study has been employed to get a clear picture of the challenges faced by the social enterprises. The challenges identified by this phase have been converted into a questionnaire and then tested through the quantitative methods.

The first phase of qualitative research phase has been described below.
3.2 Qualitative Research

The qualitative phase had been applied to explore the social entrepreneurship space as this area is emerging and has got the attention of scholars very recently (Alvord, Brown, & Letts, 2004; Dees & Elias, 1998). The research in this area is still in nascent stage and mostly borrows theories from other domains like entrepreneurship, social science, amongst others. The qualitative research offers the opportunity of cross-disciplinary research in the very diverse field of enquiry (Cassell and Symon, 2006) so this is reasonable to apply for this phase. This also provides a good opportunity to explore and establish new theories.

As per Roethlisberger (1977), qualitative research is appropriate in areas in which research and theory are at their early, formative stages. It is also suitable for areas that have “practice-based problems where the experiences of the actors are important and the context of action is critical” (Bonoma, 1983). Thus, above mentioned research justifies the use of qualitative method in this emerging and experience driven area.

Before proceeding to the steps used for implementing qualitative research in this study, it’s indispensable to define Qualitative research.

Denzin and Lincoln (2000) define qualitative research as means to study things in their natural settings, attempting to make sense of, or to interpret, phenomena in terms of the meanings people bring to them”. In other words, it does not disturb the context of the phenomenon and brings out the holistic interpretation. It also analyses objects in such a way that produces findings or concepts and hypotheses that can not be arrived at by statistical methods (Glaser, 1992).
According to Shank (2002) it is “a form of systematic empirical inquiry into meaning”. He emphasizes the fact that it is a systematic inquiry which is empirical in nature. Thus, he is putting an argument in support of the rigor and validity of the qualitative research, which is now getting better acceptance than in the past. There has been widespread acknowledgement of qualitative research as a valuable and valid research approach (Eisenhardt, 1989; Morgan and Smiricich, 1980; Carroll and Swatman, 2000; Weber, 2004; Cassell and Redman, 2001).

To conduct qualitative research, there are various tools available which can be applied for data collection and data analysis. The most frequently used ones are:

1. **Ethnography** - Ethnography is the study of social interactions, behaviours, and perceptions that occur within groups, teams, organizations, and communities (Reeves et. al., 2008). In this, the researcher becomes a part of the investigating entity and brings out the insider perspective.

2. **Phenomenology** – It is a branch of philosophy which studies human phenomenon without considering questions of their causes, objective reality or their appearances (Husserl, 1983).

3. **Hermeneutics** – It focuses on defining shared linguistic meaning for a representation or symbol. It also posits that linguistic meaning is likely to open to infinite interpretation and reinterpretation due to the interpretative ambiguity coming from pre-suppositions, to the conditions of usage different from authorial intention, and to the evolution of words (Marshall et al. 2001).

4. **Observation** – It is a social research technique that involves the direct observation of phenomena in their natural setting. In contrast to experimental research, it does not change the context of the entity to be studied.

5. **Grounded Theory** - Grounded theory is regarded as an inductive methodology for generating new theory from data (Goulding, 2002; Locke, 1996; Chenitz and
Swanson, 1986). It was developed to generate substantive theory on scholarly basis by Glaser and Strauss (1967).

6. Case Studies – In a case study, an entire organization or entity can be investigated in-depth with meticulous attention to detail. This is highly focused, so it enables the researcher to carefully study the order of events as they occur or to concentrate on identifying the relationships among functions, individuals, or entities.

As suggested by Benbasat et al. (1987) Case studies are more suitable for the exploration, classification and hypothesis development stages of the knowledge building process, the data collection for this research has been done through the Case Study method. The details of the selected organizations have been gathered chronologically with special attention to the challenges faced by them in the process of growth. The data was collected from Employees, Decision Makers and Beneficiaries. It was gathered through semi-structured interviews, micro ethnography and focus group interviews.

3.2.1 Case Study Research

As discussed earlier, case study holistically studies an entity and gathers detailed data. As per Merriam (1998) case is a unit, entity, or phenomenon with defined boundaries that the researcher can demarcate or “fence in”. Moving on from case to case study Benbasat et al. (1987) has defined it as “a method of examination of a phenomenon in its natural setting, employing multiple methods of data collection to gather information from one or few entities (people, groups or organizations)”.

The organizations studied under this method were chosen on two criteria.

1) Age of the organization – The exploratory study suggested that the age of the organization may change the challenges faced by the organization. Thus, one of the
criteria for selecting the organizations for study is the age of the organization. As per the definition given by National Entrepreneurship Network (Sinha, 2010) the age of organization has been divided into

a. 7 years & less as Start-up

b. more than 7 years as mature organizations,

During the course of the research, it was evident that the organizations in the mature category also needs to be divided further into two categories i.e. growth organization - more than 7 years till 15 years and Mature - more than 15 years. Thus, the organizations have three categories based on age as given below:

Figure 3.8
Venture Life Cycle

2) Geographical spread – The second criteria selected is the geographical spread of the organization. It suggests whether the organization is working in one or two geographical region(s)/ beneficiary groups or they decide to move to more than two geographical regions or beneficiary groups. The assumption behind this criterion was to test whether geographical spread affects the challenges faced by social enterprises.
The organizations selected on these criteria are:

**Table 3.1**

<table>
<thead>
<tr>
<th>Age of Organization</th>
<th>Less than 7 years</th>
<th>7+ to 15 years</th>
<th>More than 15 years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geographical Spread</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limited Spread</td>
<td>SammaaN</td>
<td>NIDAN</td>
<td>Development</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Alternatives</td>
</tr>
<tr>
<td>Wide Spread</td>
<td>KNIDS Green Pvt. Ltd</td>
<td>Goonj</td>
<td>SEWA</td>
</tr>
</tbody>
</table>

KNIDS Green Pvt. Ltd. and SammaaN were the start-up organizations where systems were not expected to be formed yet. At the time of data collection, SammaaN was just working in Bihar, whereas KNIDS Green had presence in Bihar, Uttar Pradesh, Chhattisgarh and Jharkhand.

In the next category, Goonj is working pan-India with its partner institution and NIDAN is working in Bihar & has office in Delhi. They were expected to have some systems in place and some in formative stage.

The third category had Development Alternatives (DA) and SEWA. DA is working in Bundelkhand region with its head office in Delhi whereas SEWA has offices in northern India. These organizations were supposed to have most of the processes in place.

Research says that qualitative sampling is concerned with information richness (Kuzel, 1992), and two key considerations should guide the sampling methods. One
of the considerations is appropriateness and other one is adequacy (Morse, 1995). Taking care of these two conditions, the data has been collected from these organizations till the data saturation is reached. Data saturation means that the data has been collected till the time responses had nothing new to add to the findings.

Many researchers have given strategy to conduct the case study. Yin’s case study protocol (1984) and Eisenhardt’s eight-step roadmap (1989) are amongst those which give guidelines for conducting case study research. According to Yin (2003), case study strategy has five components: the study’s questions, its propositions which reflect on a theoretical issue, its unit(s) of analysis (the event, entity, or individuals noted in the research questions), the logic - linking the data to the propositions, and the criteria for interpreting the findings.

The field research done with these organizations enabled a deeper understanding of the ‘complex interaction of people, processes and technology’ (Cepeda et al., 2004) within organizations. As for conducting case study cooperation from the organization of research is needed, so the data may have chances of being biased. To remove the chances of biases, triangulation of methodologies has been conducted in this study.

The triangulation also provided validity of the findings and proved that the gathering of data using multiple techniques strengthens the study.

3.2.2 Triangulation

Triangulation refers to the use of more than one approach to the investigation of a research question in order to enhance confidence in the ensuing findings (Bryman, 2001). It is associated with measurement practices in social and behavioral
research. According to Webb et al. (1966), triangulation gives most convincing evidence in the support of measurement processes.

Denzin (1970) extended the idea of triangulation beyond its conventional association with research methods and designs. He distinguished four forms of triangulation:

1. Data triangulation, which entails gathering data through several sampling strategies, so that slices of data at different times and social situations, as well as on a variety of people, are gathered.
2. Investigator triangulation, which refers to the use of more than one researcher in the field to gather and interpret data.
3. Theoretical triangulation, which refers to the use of more than one theoretical position in interpreting data.
4. Methodological triangulation, which refers to the use of more than one method for gathering data.

This study uses the methodological triangulation which validates the information gathered from different sources. The methods used for data collection are:

1) Interview – Interviews are an uncomplicated source of data collection. In this research, open ended interviews with the social entrepreneurs, other decision makers, and the employees revealed perception of these stakeholders. The data gathered through this method may have been influenced by the biasness of the respondents. Therefore, the ability to read between the lines is very much required for its analysis.
Table 3.2
Interview details

<table>
<thead>
<tr>
<th>Organization</th>
<th>Number of In-depth interviews = 31</th>
</tr>
</thead>
<tbody>
<tr>
<td>Development Alternatives</td>
<td>8</td>
</tr>
<tr>
<td>SEWA</td>
<td>7</td>
</tr>
<tr>
<td>NIDAN</td>
<td>3</td>
</tr>
<tr>
<td>Goonj</td>
<td>6</td>
</tr>
<tr>
<td>KNIDS Green Pvt. Ltd.</td>
<td>3</td>
</tr>
<tr>
<td>SammaaN Foundation</td>
<td>4</td>
</tr>
</tbody>
</table>

2) Focus Group – It is a method of group interviewing in which the interaction between the moderator and the group, as well as the interaction between group members, serves to elicit information and insights in response to carefully designed questions. As discussed by Morgan (1993), security provided by the group allows members who are lower in the 'power hierarchy' within an organization to express feelings and experiences that they would not otherwise share. Thus, this method was used with the beneficiary groups of the social enterprises studied. The findings were compared with the findings received from the interviews.

Table 3.3
Focus Group Discussion

<table>
<thead>
<tr>
<th>Organization</th>
<th>Number of FGD = 8, avg. group size = 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Development Alternatives</td>
<td>2</td>
</tr>
<tr>
<td>SEWA</td>
<td>2</td>
</tr>
<tr>
<td>NIDAN</td>
<td>1</td>
</tr>
<tr>
<td>Goonj</td>
<td>1</td>
</tr>
<tr>
<td>KNIDS Green Pvt. Ltd.</td>
<td>1</td>
</tr>
<tr>
<td>SammaaN Foundation</td>
<td>1</td>
</tr>
</tbody>
</table>
3) Micro-Ethnography – Ethnography is a study of cultures and members of a group to understand their behavior first hand in their natural setting. Spradley (1980) identified different forms of Ethnography on a continuum with both extremes as Micro and Macro ethnography. Microethnography is a scale down version of ethnography and it employs ethnographic methods to understand practices and problems of a smaller group better in shorter duration (Gerrish & Lacey, 2010).

<table>
<thead>
<tr>
<th>Micro-ethnography details</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Organization</strong></td>
</tr>
<tr>
<td>Development Alternatives</td>
</tr>
<tr>
<td>SEWA</td>
</tr>
<tr>
<td>NIDAN</td>
</tr>
<tr>
<td>Goonj</td>
</tr>
<tr>
<td>KNIDS Green Pvt. Ltd.</td>
</tr>
<tr>
<td>SammaaN Foundation</td>
</tr>
</tbody>
</table>

The data collected through the above methods has been analyzed by applying Grounded Theory (Glaser and Strauss, 1967).

3.2.3 Grounded Theory

Grounded Theory (GT) is an inductive and interpretive research method, which was discovered by Glaser and Strauss in 1967. This has been described by them as “a general methodology for developing theory that is grounded in data which is systematically gathered and analyzed. Theory evolves during actual research, and it does this through continuous interplay between analysis and data collection” (Strauss and Corbin, 1998).
Rooted in medical sociology, GT has spread to wider practitioner fields such as nursing, education, psychology, accounting, business management, public health and social work over the last 40 years (Tan, 2010).

As stated over time, GT has diversified and evolved in different disciplines. But there exists practical difficulties for many new GT researchers in analyzing data and generating an integrated grounded theory. The reason behind this is the different versions evolved from Glaser and Strauss’ original idea, later their own disagreed discourse on coding techniques and some misinterpretations of their ideas.


Amidst the conceptual difference between the original 1967 Glaser and Strauss version, the 1990 Strauss and Corbin rendition, the 1992 Glaser interpretation, or other GT versions, the researcher has to identify one of these versions of GT to be followed for the study. But before deciding about the relevant version of GT for this study, it is essential to identify cases where it is applicable. GT is applicable mainly in the following cases:

- The researcher seeks to create a theory about issues of importance in people’s lives and specifically focuses on human interaction or aims to explore new territory (Denscombe, 2003).
• The study of new socio-technical phenomena (Fernaández, 2004).
• The area of interest is a new developing one and does not have a long, firm and empirically based literature yet (Goulding, 1999).

The area of Social Entrepreneurship is new and emerging. It has got the attention of scholars very recently (Alvord, Brown, & Letts, 2004). Thus, the third case proposed by Goulding (1999) makes this study suitable for applying GT.

Amongst different versions of GT, this research has employed the original 1967 Glaser and Strauss version.

To start the research, they indicated that researchers should start with an open mind (Glaser and Strauss, 1967). Many researchers interpret that one should go to the field without literature review as the prior knowledge and pre-conceived notions may affect the theory generation process.
On the other hand, Seidel and Kelle (1995) explained, “an open mind does not mean an empty head”. The literature review may help in defining the research focus, identifying problems and gaps, as well as choose a suitable methodology for the study. The researcher’s personal experience in this regard also advocates a thorough literature review.

Strauss and Corbin (1998) suggested several ways of using the literature in GT. Some of them are:
• Helping researchers make comparisons to data at the dimensional level;
• Enhancing sensitivity;
• Extending a theory under certain circumstances;
• As a secondary source of data;
• For formulating question;
Suggesting theoretical sampling;
- Helping researchers extend, validate and refine knowledge in the field.

The next steps in GT are data collection, coding and categorizing, identifying the relationships or connection between the categories. The theory of GT suggests that these steps are not sequential. The researcher needs to continually move back and forth between data collection, coding and categorizing, refining, and interconnecting through additional data collection. This process goes on until the study reaches to theoretical saturation. Theoretical saturation means “a stage where any additional analysis no longer contributes to discovering anything new about a category” (Mason, 2002).

3.2.4 Coding

Coding is an important step in this process. It is a fundamental analytical process, which plays a vital role in analyzing, organizing and making sense of textual data (Tan, 2010). As Seidel and Kelle (1995, p. 58) stated, “Coding is heuristic devices for discovery”. In the literature, there are different coding strategies depending on the version of GT being used. The original Glaser and Strauss (1967) version vouch for ‘explicit coding’ and ‘constant comparative method’ to be used for this process. But due to clarity of the procedure, Strauss and Corbin (1990) recommended dynamic and fluid coding procedure, i.e. open coding, axial coding, and selective coding, which has been used for this research.

The details of coding and categorizing process followed for this research have been discussed in the next chapter.

3.2.5 Theory Generation

Glaser and Strauss (1967) mentioned that theory generation from the data is the main aim of GT. But the question arises that what is the meaning of theory.
Denscombe (2003) defines it as “A theory is a proposition about the relationship between things”

Suddaby (2006) expresses in other words that the purpose of grounded theory is not to make truth statements about reality, but, rather, to elicit fresh understandings about patterned relationships between social actors and how these relationships and interactions actively construct reality.

These definitions put emphasis on patterns of relationship. Becker (1993) added to it by saying that a GT study should answer “what is going on there and how” along with giving the proposed relationships among the generated concepts and categories.

This study also exhibits relationship amongst the categories (challenges) and the concepts (influencing factor). This relationship has been tested through quantitative method.

3.3 Quantitative Research

As per Creswell (1994), the quantitative research is a type of research that `tries to explain phenomena by collecting numerical data which is analyzed using mathematically based methods (in particular statistics).’ This concise definition captures the essence of research, by mentioning the explanation of the phenomenon in the first part and the second part refers to the process involved with quantitative methods of data collection and analysis. Quantitative research is good at providing information in breadth from a large number of units. It makes its findings applicable to a large population. These types of research are well-suited for the testing of theories and hypotheses. In the ‘sequential mix method’ setting, the findings of the qualitative phase could be tested with the help of quantitative research.
This study may benefit from the quantitative research in fulfilling its third object i.e. to study the effect of identified factors on challenges of social enterprises in livelihood sector. The result of qualitative phase i.e. the challenges faced by social enterprise and the factors affecting them, has been fitted into a causal relationship. The intensity of this cause & effect relationship is being tested with the help of quantitative research.

3.3.1 Data Collection
To execute the quantitative part, the constructs identified through qualitative research were broken into measuring indicators. The reason to identify measuring indicators for the constructs is that there are only a handful of variables that can be considered as observable (Haenlein and Kaplan, 2004). Most of the variables identified in the qualitative research could be manifested through other variables. These measuring indicators or manifest variables were then converted into a questionnaire.

The respondents of the questionnaire were the decision makers of the Social enterprises which belong to Ashoka Fellows (discussed in the previous chapter), who are working in livelihood sector. The data gathered from them was analyzed through structural equation modeling. The section below explores the reasons of this selection.

3.3.2 Traditional quantitative data analysis
Traditionally, the data analysis techniques which examine the liner relationship of one dependent and multiple independent variates include factor analysis, regression, discriminant analysis, cluster analysis etc. These belong to the core set of statistical instruments which can be used to either identify or confirm theoretical hypothesis based on the analysis of empirical data. The brief definition of these techniques is given below:
Regression - It is a statistical tool for the investigation of relationship between variables. It includes many techniques for modeling and analyzing several variables, with the focus on the relationship between a dependent variable and one or more independent variables.

Factor Analysis – It is a statistical method used to study the dimensionality of a set of variables. It investigates whether a number of variables of interest are linearly related to a smaller number of unobservable factors. As the factors are not observable, it disqualifies regression and other methods.

Discriminant analysis – It undertakes the same task as multiple linear regression by predicting an outcome, but unlike regression it can be applied to the categorical data.

Cluster analysis – It divides the data into groups so that the data within a group is homogeneous and between the groups there is heterogeneity. This way it summarizes the data for meaningful interpretations.

These traditional methods of quantitative data analysis have three limitations (Haenlein & Kaplan, 2004) namely,

(a) The postulation of a simple model structure (only one relationship can be tested);
(b) The assumption that all variables can be considered as observable; and
(c) The conjecture that all variables are measured without error

These limitations may bind their applicability in some of the research situations, especially where the model is complex with multiple relationships or the variables involved cannot be measured directly. The structural equation modeling helps in
overcoming these limitations. It allows the simultaneous modeling of relationships among multiple independent and dependent constructs (Gefen, Straub, & Boudreau, 2000). It is an extension of the factor analysis and regression. This study has similar constraints as given above. Before presenting the detailed argument for the usage of structural equation modeling, it is important to note its basic properties.

3.3.3 Structural Equation Modeling

Structural equation modeling (SEM) is a tool for analyzing multivariate data, which has been long known to be especially appropriate for theory testing (e.g., Bagozzi & Yi 1988). It goes beyond ordinary regression models to incorporate multiple independent and dependent variables as well as hypothetical latent constructs, that clusters of observed variables might represent. These independent and dependent variables are called endogenous and exogenous variables. Endogenous represents variables which have the influence of the variables present in the model (this acts as dependent variable), whereas exogenous variables are influenced by the factors beyond the scope of the model, they are considered as independent variables. The endogenous variables are explained by others while the exogenous variables do not have any arrow pointing towards them in an Inner model (explained in coming paragraphs). These variables are also called latent variables. This term is applied when the variables of interest cannot be measured perfectly and some observable indicators are used for measuring these variables. These indicators also provide a way to test the specified set of relationships among latent variables as a whole, and allow theory testing even when experiments are not possible.

These observed indicators can be split into two groups: (a) reflective indicators which are construct dependent and (b) formative ones which cause the formation of or changes in an unobservable variable (Bollen & Lennox, 1991). The formative indicators are highly correlated with each other and a change in one may create
change in others and the construct, whereas the reflective indicators may have positive, zero or negative correlation with each other (Hulland, 1999). The study is considering only reflective indicators for the latent variables.

There are two types of approaches to apply SEM

- **Covariance-based techniques** (CB-SEM; Jöreskog 1978, 1993)

**Covariance-based (CB-SEM)**

It attempts to minimize the difference between the sample co-variances and those predicted by the theoretical model. It can map the path for multiple latent and observed constructs and can analyze all these paths simultaneously. It is usually used with an objective of model validation and needs a large sample (minimum 100 and preferably more than 200). It supports confirmatory research. CB-SEM has a history of long usage. CB-SEM estimates parameters which reduce the discrepancy between covariance of the estimated value and sample. It puts constraints regarding the number of observations and small sample sizes, because that may often lead to biased test statistics, inadmissible solutions, and identification of problems - especially in complex model set-ups (Hair, J. et al, 2012). On the other hand, PLS-SEM can analyze small sample size as well and generally it achieves ‘high levels of statistical power’ (Reinartz et al. 2009).

**Partial Least Square (PLS-SEM)**

PLS is a component-based alternative for estimating Structural Equation Model. It is a method for analyzing a system of linear relationships between multiple blocks of variables. Wold (1975) originally developed PLS-SEM under the name NIPALS (Nonlinear Iterative Partial Least Squares), and Lohmöller (1989) extended it. It has been developed as an alternative to CB-SEM that would emphasize prediction
while simultaneously relaxing the demands on data and specification of relationships (e.g., Dijkstra 2010; Jöreskog and Wold 1982).

The process of PLS SEM has three basic steps. At first it starts with outside estimation of the weighted relationship between indicators and their respective latent construct or unobserved variables. Then, on the basis of these weighted relationships, weight of each construct is calculated by averaging the weights of its indicators. The weights, used to calculate this aggregation, are determined in a manner similar to a principal-components analysis for reflective or regression analysis for formative indicators (Cassel, Hackl, & Westlund, 1999). These weights are used in a set of regression equations to determine the parameters for the structural relations (Fornell & Bookstein, 1982).

According to S. Wold (1993) PLS is applied where the number of ‘significant’ latent variables is small, much smaller than the number of measured variables and the number of observations. Though, in this study the number of significant variables is large but there are several other reasons which justify the usage of PLS for this research. Some of them are stated below.

_Reasons for Using PLS SEM_

According to Chin (1988), in the exploratory settings where new measures are used and less substantial knowledge is available in the area of study, use of PLS SEM is more appropriate. The area of social entrepreneurship is nascent and the scales are not fully developed, so this argument in support of usage of PLS SEM is important. As discussed earlier, the indicators of the unobservable construct are of two types i.e. reflective & formative. A study by Jarvis, MacKenzie, and Podsakoff (2003) gives an indication that managerial constructs might be reflected better by formative than by reflective indicators. The formative indicators may give misleading result when they are analyzed through CB-SEM whereas PLS does not
have any such discrepancy. It can therefore be used for models with either reflective, formative, or both types of indicators (Fornell & Bookstein, 1982). The unobservable constructs of this study have reflective indicators.

The research by Marsh, Hau, Balla, and Grayson (1998) suggests that more indicators per latent variable lead to fewer improper solutions and more stable results. But as per Baumgartner & Homburg (1996), in some areas such as management research, researchers rarely have more than a few indicators per unobservable variable to their disposition. Researchers have also argued that if a construct’s scope is narrow, unidimensional, and unambiguous for the respondents, using single-item measures is the best approach (Haenlein & Kaplan, 2004). This research consists of few construct which has single item measures.

PLS has the advantage that it “involves no assumptions about the population or scale of measurement” (Fornell & Bookstein, 1982) and consequently works without distributional assumptions and with nominal, ordinal, and interval scaled variables.

The sample size is not a very stringent criteria for PLS. Chin and Newsted (1999) indicated that PLS can be performed with a sample size as low as 50. H.Wold had even “analyzed 27 variables using two latent constructs with a data set consisting of ten cases” (Chin & Newsted, 2003). A research based on 6 subjects has been published by Tenenhaus, Pagès, Ambroisine & Guinot (2005). Lee (1994) has also used a sample size of 18. As a popular rule of thumb for robust PLS-SEM estimations, Barclay et al. (1995) suggest using a minimum sample size of ten times the maximum number of paths aiming at any construct in the outer model (i.e., the number of formative indicators per construct) and inner model (i.e., the number of path relationships directed at a particular construct). According to this
thumb rule, the sample size of this study is 40 (maximum number of path relationship directed at a particular construct is four).

All these examples support the usage of PLS SEM for this research.

The process of application of the PLS-PM has two parts ‘Formal Model’ and ‘Operative Model’ (Sanchez, 2013). The formal model illustrates the theoretical part which requires assumptions and specifications. The operative model explains the execution of the theoretical concept which is the computation side.

**Formal Model**

In the course of description of this part, there are three components which are important to it i.e. Inner Model (structural model), Outer Model (measurement model) and the weight relations.

**Inner Model**

It establishes the basic structure of the model proposed through demonstration of the relationship between all the exogenous and endogenous constructs present. This is also called the ‘structural model’. The structure could be depicted in the form of mathematical connotations as follows

\[ \text{LV}_j = \beta_0 + \sum_{i \rightarrow j} \beta_{ij} \text{LV}_i + \text{error}_j \]

The subscript \(i\) of \(\text{LV}_i\) refers to all the latent variables that are supposed to predict \(\text{LV}_j\). The \(\beta_{ji}\) are the path coefficients and they represent the “strength and direction" of the relations between the response \(\text{LV}_j\) and the predictors \(\text{LV}_i\). \(\beta_0\) is just the intercept term, and the \(\text{error}_j\) term accounts for the residuals.
PLS-SEM does not permit a recursive relationship between the latent variables of the structural model. In other words, no causal loops should be present in the Inner model.

**Outer Model**

The outer model or measurement model presents the relationships between the constructs and their measuring indicators or manifest variables. It includes the unidirectional predictive relationships between each latent construct and its associated observed indicators. Multiple relations are not permitted therefore; indicator variables are associated with only one latent construct. This also helps in avoiding the multi-co-linearity issues.

These manifest variables can be of two types i.e. reflective and formative (as discussed earlier). PLS-SEM can handle both formative and reflective measurement models, but as mentioned earlier, this research has only reflective indicators. The path diagram below shows both of these indicators.

![Path Diagram](image)

As this introduces the measuring indicators in the model so it is also called the measurement model. Similar to the inner model it can also be depicted through mathematical connotations

\[ X_{jk} = \lambda_{0jk} + \lambda_{jk} LV_j + \text{error}_{jk} \]  

Reflective Manifests  

Formative Manifests
LV_j = \lambda_{0j} + \lambda_{jk}X_{jk} + \text{error}_j \quad \text{Formative}

The \lambda_{jk} are called loadings; \lambda_{0jk} is just the intercept term, and the error terms account for the residuals.

**The Weight Relations**

As described earlier, the latent variables are estimated as the linear combination of a block of manifest variables through their weighted sum. This weight relationship of the manifest variables with their latent construct is called the weight relation. This can be illustrated through mathematical connotations as below

\[
LV_j = \sum_k w_{jk} X_{jk}
\]

The \( w_{jk} \) represents the weight coefficient of the manifest variable \( X_{jk} \).

**Operative Model**

This model is sequential to the formal model. After getting the values for the manifest variables (data), the estimation of the path coefficient and loadings is done under this part. This process of estimation is divided in three stages

- Stage 1: Get the weights to compute latent variable scores
- Stage 2: Estimating the path coefficients (inner model)
- Stage 3: Obtaining the loadings (outer model)

The first stage consists of obtaining the weights that will be used to get the scores of the latent variables. The second stage has to do with estimating the path coefficients of the inner model. And the third stage involves the computation of the loadings (outer model).

There are several software available for the analysis through PLS SEM. Few of the widely used once are AMOS, LISREL, SAS, SMARTPLS, R.
For this study R has been used for analysis. The reasons for using this over others are described in the section below.

‘R’ and the Reasons of its Usage

R is free open source software for data analysis, statistical computing, and graphics. The R project started in 1995 by a group of statisticians at University of Auckland and has continued to grow ever since (Seefeld & Linder, 2007).

- The R project is largely an academic endeavor and as it is an open source so people from all over the world contribute and share their own work in the form of packages. Most of the contributors are statisticians. This makes it constantly evolving.

- Most of the alternate software available is costly. Thus, being free software, R has no competitor for reaching the marginalized academic places of the world.

The above mentioned reasons motivated the researcher to use this software for the analysis. The detailed process of execution of the PLS-PM has been described in the chapter of quantitative data analysis.