CHAPTER TWO

Research Methodology
2.1- Introduction

This chapter covers the methodology used in this research; it also gives detailed information on the following sections: Importance and Objective of the subject, Research Hypotheses, Research Framework, Research Purpose, Research Approach, Research strategy, Research Type, Research Instrument, Research Design, Structural Equation Modeling, Scope of the Research, Terminology, Acronyms and Abbreviations, Sample selection, Analysis of data and Quality standard which include Validity and Reliability were discussed.

2.2- Important Concepts of the study

- **Small and Medium Enterprises:** According to the new definition by Micro, Small and Medium Enterprises Development Act, 2006, Small enterprises are classified broadly into; (i) Enterprises engaged in the manufacture / production of goods pertaining to any Industry & (ii) Enterprises engaged in providing / rendering of services. It defines MSMEs in the following way:

  - **Manufacturing Enterprises:** defined in terms of investment in plant and machinery (excluding land & buildings) and further classified into:-
    
    Small Enterprises - Investment above Rs. 25 lakh & upto Rs. 5 crore  
    Medium Enterprises - Investment above Rs. 5 crore & upto Rs. 10 crore  

- **Service enterprises**
    
    Small Enterprises - Investment between 10 Lakhs & 2 crore  
    Medium Enterprises - Investment between 2 crore & 5 crore  

- **Electronic Commerce** “is sharing business information, maintaining business relationship, and conducting business transactions by means of telecommunications networks”

- **Organizational performance** comprises the actual results or output of an organization as measured against its intended results or outputs.
• **Financial performance** Measuring the results of a firm's policies and operations in monetary terms. These results are reflected in the firm's return on investment, return on assets, value added, etc.

• **Operational performance**, Firm’s performance measured against standard or prescribed indicators of marketing effectiveness, product/service quality, and customer satisfaction.

• **Business-to-Business e-commerce** is simply defined as e-commerce between companies.

• **E-markets** are simply defined as Web sites where buyers and sellers interact with each other and conduct transactions.

• **Business-to-consumer** e-commerce or commerce between companies and consumers

• **Business-to-Government e-commerce** or B2G is generally defined as commerce between companies and the public sector.

• **Consumer-to-consumer e-commerce** or C2C is simply commerce between private individuals or consumers.

• **Mobile commerce** (M-commerce) is the buying and selling of goods and services through wireless technology-i.e., handheld devices such as cellular telephones and personal digital assistants.

• **E-Business** is: “The transformation of an organization’s processes to deliver additional customer value through the application of technologies, philosophies and computing paradigm of the new economy.”

• **Level of electronic commerce adoption** is: defined by the number of electronic commerce applications being used by the companies.

### 2.3- Importance of the Subject

As a legacy of Gandhian philosophy, since independence the small and medium sectors have played an important role in the economic development of the country. Especially since the commencement of planning for economic growth, adequate emphasis has been given on the development of Small and Medium Enterprises (SMEs) by policy makers, politicians and the intelligentsia alike. The multi pronged objectives of increased industrial output, generation of
employment, dispersal of industrial activities across regions and development of entrepreneurship has been successfully met through the propagation of SMEs\(^3\).

The importance of the Small and Medium Enterprises sectors is well-recognized from its significant contribution to the socio-economic objectives of growth in generation of employment, output, exports and fostering entrepreneurship. Currently, the sector accounts for around 95 per cent of the industrial units in the country contributing 40 per cent of the manufacturing sector output and approximately one-third of the nation’s exports\(^2\).

This research focuses on Small and Medium Enterprises for two key reasons. First, they are important to economic development in developing countries such as India. As per the United Nations Conference on Trade and Development, SME’s account for 60 to 70 percent of all employment in developing countries. It is critical for such businesses to be prepared for and take full advantage of any benefits offered by electronic commerce.

The second reason to focus on SME’s is that they are in a very good position to adapt to new technology; they may be able to adapt faster than larger companies that can slowed by bureaucracy and stricter staffing hierarchies. E-commerce may offer them comparatively more advantages to find new customers and suppliers especially in markets they have not easily been able to reach before – either internationally or regionally. Markets everywhere are globalizing partially due to the widespread use of the Internet. Electronic commerce can give SME’s a better chance to compete in their markets and, indeed, in some cases, is or will soon become a competitive necessity for survival. The potential benefits of electronic commerce to “level the playing field” for SME’s – allowing them to compete better – are critical for them to understand and sort out\(^3\).

### 2.4- Objective of the Study

In India, it is estimated that there are around 128.4 lakh Small and Medium Enterprises units in 2007-2008 and at the same period of time SME’s employ over 312.5 lakh persons. SME’s

---

1 : Gupta, Rahul, (2006), Scope of Cottage and Small Scale Industry in West Bengal in the Early2000, IBS Research Center, Kolkata, p1
2 : (http://www.sidbi.com)
account for 42 percent of Manufacturing sector turnover and 35 per cent country’s exports. (www.rbi.org.in) SMEs are increasingly seeing the benefits arising from e-commerce as expanded geographical coverage giving them a larger potential market into which they can sell their products and services. Some of the key industries that have high potential for early adoption of e-commerce are: financial (stock exchanges and banks), automobiles, retail, travel, IT and manufacturing.¹ The general purpose of the research is to study of organization and management of electronic commerce applications in Small and Medium Enterprise Sector. The main purpose of this research is to study the effect of e-commerce applications on the organizational performance of small and medium enterprises in Pune. There are many different types of e-commerce applications that may influence to organizational performances; in this research, we group them according to five different categories: electronic advertising, electronic payment system, electronic marketing, electronic customer support service and electronic order and delivery. According to literature review, most common types of organizational performance measures that used in recent empirical research are included: Financial or accounting performance, operational performance and market-based performance. In this research, we tried to indentify the relationship between application of electronic commerce and organizational performance of small and medium enterprises in Pune. Generally, the study will try to answer these questions:

1) Is there any relationship between application of E-commerce and organizational performance in small and medium enterprises?

2) Is there any relationship between application of E-commerce and operational performance in small and medium enterprises

3) Is there any relationship between application of E-commerce and Financial performance in small and medium enterprises

4) Is there any relationship between application of E-commerce and market-based performance in small and medium enterprises

5) What is level of acceptance and adoption of e-commerce applications between the Small and Medium Enterprise.

6) What are the most impertinent benefits or potential benefits of e-commerce applications by the SMEs

7) What are the most impertinent problems or potential problems of e-commerce applications by the SMEs

2.5- Research Hypotheses

This research shall attempt to test one main hypothesis and three sub-hypotheses.

Main Hypotheses:

H: E-commerce Applications has a significant and positive impact on organizational performance in small and medium enterprises.

Sub-Hypotheses:

H1- E-commerce Applications has a significant and positive impact on operational performance in small and medium enterprises

H2- E-commerce Applications has a significant and positive impact on financial performance in small and medium enterprises

H3- E-commerce Applications has a significant and positive impact on market-based performance in small and medium enterprises

2.6- Research Model

A review of the literature indicates that there are many different types of e-commerce usage that may influential to organizational performance in small and medium sized enterprises; we can group them according to five different categories: 1) Electronic Advertising 2) Electronic Payment System 3) Electronic Marketing 4) Electronic Customer Support Service 5) Electronic Order and Delivery. On the other hand, organizational performance can be further subcategorized into operational performance, financial performance and market-based performance.
2.7- Research Purpose

The proposed research will use data gathered in a retrospective pattern during the recent decade in India. Research can be formulated in certain ways depending on the objectives as indicated in the research problem. According to Zikmund (2000)\(^1\) proposed many research activities help to categorize the purpose of the study in terms of the research into three namely exploratory, descriptive and explanatory. The classification in terms of the purpose of the study can often be distinguished between exploratory, descriptive and explanatory research purpose. The purpose of the study is in connection with the research problem and also depends on the objective and problem formulation, purpose of the study varies. Saunders et al. (2007)\(^2\) stated the question as to what research question to use is not a question of either/or other three classification tends to overlap Saunders et al (2007) stated that the purpose can be exploratory and become more descriptive later. An exploratory study is valuable means of finding out ‘what is happening; seek new insights; ask questions and to assess phenomena in a new light’\(^3\). Exploratory approach is helpful when the researcher wants to clarify the understanding of a problem. The great advantage is that it is flexible and adaptable to change.

---

\(^1\): Zikmund, W. (2000), business research methods, 6 editions, Dryden Press, Texas, USA
\(^3\): Robson, C. (2002), Real World Research: a resource for social scientists and practitioner- researcher (2 edn), Oxford
Descriptive research is ‘to portray an accurate profile of persons, events or situations’ (ibid). Furthermore, Saunders et al. (2007) also elaborated that descriptive research can be seen as an extension of exploratory research. With descriptive purpose the researcher already needs to have a clear picture of the phenomena he/she will further investigate. It also helps the researcher to draw conclusions from the data he or she is describing. Saunders et al. (2007) stated that explanatory researches are studies that establish causal relationships between variables. The emphasis here is on studying a situation or a problem in order to explain the relationships between variables. Anderson & Svensson (1999)\(^1\) postulated that research starts with exploratory phase to find what the study is about and continues to descriptive and explanatory depending on the objective of the study. Information on SMEs application of e-commerce in Pune tends to be limited; we are adopting exploratory and descriptive studies to attain the necessary information as regards to the research hypotheses. This approach would further help us to understand and describe the phenomenon. This thesis is both exploratory and descriptive since it seeks to pertain what is happening, seek insights and ask questions about e-commerce and levels of application by SMEs especially in Pune (India). The study also describes events, situations and draw conclusions on data collected. Also the study seeks to clarify understanding of problems, which the researchers are unsure of the precise nature of the problem.

### 2.8- Research Strategy

Research strategy is seen as plan that the researcher uses to carry out the study to answer the research questions\(^2\). The research strategy explains how the researcher collects and analyses data gathered. He further states that the type of research strategy to be used by researcher depends largely on the research purpose (exploratory, descriptive or explanatory). The type of research strategy to use depends on three important factors namely;

1- type of research question asked

2- the control the researcher has over behavioral events and

3- Degree of focus on contemporary by historical events.

---


The most important among the three categories is the type of research question asked. The type of research strategy used is linked to how the research question is posed. There are five primary strategies in the social sciences for conducting social science research. They are as follows:

1. Experiments
II. Survey
III. History
IV. Analysis of archival information
V. Case study

Each of these strategies has a different way of collecting and analyzing empirical evidence and each can be used for exploratory, descriptive and explanatory research\(^1\). Although usually, case studies are considered appropriate for the exploratory phase, surveys and histories fit the descriptive phase, and experiments are the only way of doing explanatory or causal inquiries. Yin (1994) argues that hierarchical view is incorrect and each strategy can be used for all three purposes: exploratory, descriptive, or explanatory.

Based on the formulation of research questions, survey and case studies strategies suit the purpose of the thesis which provides the possibility of gaining information from a large number of respondents which yields a more general view of the phenomenon under study.

Case study refers to the collection and presentation of detailed information about a particular participant or small group, frequently including the accounts of subjects themselves. Case studies typically examine the interplay of all variables in order to provide as complete an understanding of an event or situation as possible. This type of comprehensive understanding is arrived at through a process known as thick description, which involves an in-depth description of the entity being evaluated, the circumstances under which it is used, the characteristics of the people involved in it, and the nature of the community in which it is located. Thick description also involves interpreting the meaning of demographic and descriptive data such as cultural norms and mores, community values, ingrained attitudes, and motives.

Unlike quantitative methods of research, like the survey, which focus on the questions of who, what, where, how much, and how many, and archival analysis, which often situates the participant in some form of historical context, case studies are the preferred strategy when how or why questions are asked. Likewise, they are the preferred method when the researcher has

---

\(^1\) Yin, R.K. (2003), case study research: design and methods, thousand oaks, CA, sage publications
little control over the events, and when there is a contemporary focus within a real life context. In addition, unlike more specifically directed experiments, case studies require a problem that seeks a holistic understanding of the event or situation in question using inductive logic—reasoning from specific to more general terms\(^1\).

According to Yin (1994), case studies can be single or multiple case studies. Both methods are frequently used; however there are certain distinct advantages of using multiple cases rather than single ones. Multiple cases are often considered to be more robust than single cases, and allow comparisons between cases\(^2\).

The “survey” is collection of large scale of data from fairy large population. It is generally used with deductive philosophy. This strategy is widely used in business and marketing research. Hence, questionnaire, structured observation and structured interviews often belong to survey\(^3\).

Moreover, the researchers resolved on survey, one of the reasons, being questions was based on the company’s perceptive. By studying from this perceptive, we decided that, by gaining better understanding of the companies’ processes, relationships, attitudes, and their dealing with their customers, we could get a holistic picture of the selected cases.

*Survey* is considered to be the most appropriate and suitable for this study hence decision to use it. This is due to the fact that it would enable the researchers to unearth insights and a deeper understanding about the application of e-commerce among Indian SMEs.

### 2.9- Research Purpose

This section describes type of research that can be chosen to suit best the purpose of the study. According to Yin (1994)\(^4\) researchers can choose their research strategy depending on those strategy best suits their research questions characteristic. Yin (1994) categorizes research to three categories: exploratory, descriptive, or explanatory. Description of these categories is described in the described in the following:

---

1: http://writing.colostate.edu/guides/research/casestudy/pop2a.cfm
2: Yin, R.K. (2003), case study research: design and methods, thousand oaks, CA, sage publications
2.9.1- Exploratory Research

This is when a study is undertaken with the objective to explore an area where little is known or to investigate the possibilities of undertaking a particular research study. When a study is carried out to investigate its feasibility, it is also called feasibility. It is usually used when a researcher wants to explore areas about which she/he has little or no knowledge. The goal and objective is to find models, ideas or hypotheses more willing than evaluating a hypothesis\(^1\).

Exploratory research is usually conducted to explore the problem situation. It is particularly helpful in braking broad, indistinct problem statements into smaller, more precise sub problems statement\(^1\). Exploratory research is identified as collecting information in an unstructured and informal way. Pilot projects can be used in determining the final protocols that will be used. Survey questions may be deleted or added based on the outcome of the pilot study. Selecting vases is a difficult process, but the literature provides guidance in this area\(^2\).

2.9.2- Explanatory Research

Explanatory research tries to clarify casual relationship (why and how) there is a relationships between two variable of a situation or phenomenon. This type of researches attempts to explain, for example, why stressful living results in heart attack? It usually accomplishes this goal through laboratory and field experiments\(^3\).

According to Yin (2003b)\(^4\) explanatory research goes beyond descriptive research. The researcher exceeds than just describing the characteristics, it analyses and explains why or how something is occurring. Hence, analytical research aim is to comprehend phenomena by detecting and assessing causal relations among them. Same as descriptive study the research must be knowledge able about the research subject. The research purpose in this study has been assessed\(^5\).

---

2.9.3- Descriptive Research

Descriptive research describes systematically a situation, problem, phenomenon, service or program, or provides information about, like the living condition of a community, or describes attitudes toward the issue and obtains information on the characteristics of a particular issue. Objective of descriptive is to represent an accurate profile of individual person’s events or situations like market characteristics and the goal are to describe events that have occurred or happening in present time. It may answer such questions; who, what, when, where, and how. This study supposes that the researcher has a prior knowledge about the problem situation. There should be a clear statement of the problem, specific hypothesis, and specification of the information needed and plenty of literature.

This research is frequently used when a problem is well structured and there is clear statement of the problem and it doesn’t investigate cause/effect relations and what caused the relation. The gathered data are frequently quantitative and statistical analysis is applied to evaluate and reach conclusion about the information. Given that descriptive research determines and expresses the subject we admit that descriptive research goes beyond exploratory research.

The topic under study in this research presides under the broad research field of social science with a specific emphasis on an employee’s intent to turnover based on various individual-structural- and environmental factors. Much has been published on these factors and their interdependence and it follows that a descriptive research methodology should be followed. The time allowed for this research study did not allow for a longitudinal study on the subject matter and a cross-sectional sample will be taken as input to this research.

Considering that adequate amount of literature is available and this thesis internationally relies on pervious theories and research model and variables are drawn from theory. The picture of factor affecting e-commerce adoption between India SMEs is clear and description of extent of association between variables is going to be conducted. Hence, regarding to aforementioned explanation this thesis is going to be a descriptive research.

---

2.9.4- Exploratory Factor Analysis

Many scientific studies are featured by the fact that “numerous variables are used to characterize objects”\(^1\). Examples are studies in which questionnaires are used that consist of a lot of questions (variables), and studies in which mental ability is tested via several subtests, like verbal skills tests, logical reasoning ability tests, etcetera\(^2\). Because of these big numbers of variables that are into play, the study can become rather complicated. Besides, it could well be that some of the variables measure different aspects of a same underlying variable.

For situations such as these, (exploratory) factor analysis has been invented. Factor analysis attempts to bring inter-correlated variables together under more general, underlying variables. More specifically, the goal of factor analysis is to reduce “the dimensionality of the original space and to give an interpretation to the new space, spanned by a reduced number of new dimensions which are supposed to underlie the old ones”\(^1\), or to explain the variance in the observed variables in terms of underlying latent factors\(^3\). Thus, factor analysis offers not only the possibility of gaining a clear view of the data, but also the possibility of using the output in subsequent analyses\(^4\). In this research will be given of the use of factor analysis. This will be done by carrying out a factor analysis on data from a study in the field of applied linguistics, using SPSS for Windows. For this to be understandable, however, it is necessary to discuss the theory behind factor analysis.

2.9.5- Factor analysis versus principal component analysis

After having obtained the correlation matrix, it is time to decide which type of analysis to use: factor analysis or principal component analysis. The main difference between these types of analysis lies in the way the communalities are used. In principal component analysis it is assumed that the communalities are initially. In other words, principal component analysis assumes that the total variance of the variables can be accounted for by means of its components.

---

(or factors), and hence that there is no error variance\(^1\). On the other hand, factor analysis does assume error variance. This is reflected in the fact that in factor analysis the communalities have to estimated, which makes factor analysis more complicated than principal component analysis, but also more conservative.

### 2.9.6- Technical aspects of principal component analysis

In order to understand the technical aspects of principal component analysis it is necessary be familiar with the basis notions of matrix algebra. As Rietveld & Van Hout (1993)\(^2\) state, “the number of positive eigenvalues determines the number of dimensions needed to represent a set of scores without any loss of information”. Hence, the number of positive eigenvalues determines the number of factors/components to be extracted. The construction of the factor itself is then calculated via a transformation matrix that is determined by the eigenvectors of the eigenvalues. After constructing the factors it is possible to determine the factor loadings simply by calculating the correlations between the original variables and the newly obtained factors or components. Rietveld & Van Hout (1993)\(^1\) furthermore name two criteria for dimensionality reduction in principal component analysis: 1) The new variables (principal components) should be chosen in such a way that the first component accounts for the maximum part of the variance, the second component the maximum part of the remaining variance, and so on. 2) The scores on the new variables (components) are not correlated.

### 2.9.7- Technical aspects of factor analysis

In factor analysis the different assumption with regard to the communalities is reflected in a different correlation matrix as compared to the one used in principal component analysis. Since in principal component analysis all communalities are initially, the diagonal of the correlation matrix only contains unities. In factor analysis, the initial communalities are not assumed to be 1; they are estimated (most frequently) by taking the squared multiple correlations of the variables


with other variables\textsuperscript{1}. These estimated communalities are then represented on the diagonal of the correlation matrix, from which the eigenvalues will be determined and the factors will be extracted. After extraction of the factors new communalities can be calculated, which will be represented in a reproduced correlation matrix.

The difference between factor analysis and principal component analysis is very important in interpreting the factor loadings: by squaring the factor loading of a variable the amount of variance accounted for by that variable is obtained. However, in factor analysis it is already initially assumed that the variables do not account for 100\% of the variance. Thus, as Rietveld & Van Hout (1993)\textsuperscript{1} state, “although the loading patterns of the factors extracted by the two methods do not differ substantially, their respective amounts of explained variance do!”.

2.9.8- Which type of analysis to choose?

According to Field\textsuperscript{2} (2000) strong feelings exist concerning the choice between factor analysis and principal component analysis. Theoretically, factor analysis is more correct, but also more complicated. Practically, however, “the solutions generated from principal component analysis differ little from those derived from factor analysis techniques”. In Rietveld & Van Hout (1993)\textsuperscript{1} this is further specified: “the difference between factor analysis and principal component analysis decreased when the number of variables and the magnitudes of the factor loadings increased”. The choice between factor analyses thus depends on the number of variables and the magnitude of the factor loadings. After having made this choice, the question arises how many factors there are to be retained.

2.9.9- Factor rotation

After factor extraction it might be difficult to interpret and name the factors / components on the basis of their factor loadings. Remember that the criterion of principal component analysis that the first factor accounts for the maximum part of the variance; this will often ensure that “most
variables have high loadings on the most important factor, and small loadings on all other factors”\(^1\). Thus, interpretation of the factors can be very difficult.

A solution for this difficulty is factor rotation. Factor rotation alters the pattern of the factor loadings, and hence can improve interpretation. Rotation can best be explained by imagining factors as axes in a graph, on which the original variables load. In orthogonal rotation there is no correlation between the extracted factors, while in oblique rotation there is. It is not always easy to decide which type of rotation to take; as Field (2000)\(^1\) states, “The choice of rotation depends on whether there is a good theoretical reason to suppose that the factors should be related or independent, and also how the variables cluster on the factors before rotation”. A fairly straightforward way to decide which rotation to take is to carry out the analysis using both types of rotation; “if the oblique rotation demonstrates a negligible correlation between the extracted factors then it is reasonable to use the orthogonally rotated solution”.

**2.9.10- Results of factor loadings and factor scores**

In the last sector it was already reported that the factor loadings are represented in the rotated component matrix. As may be known by now, these factor loadings are important for the interpretation of the factors, especially the high ones. One can wonder, however, how high a loading has to be in order to determine the interpretation of the factor in a significant way. This is dependent of the sample size, the bigger the sample the smaller the loadings can be to be significant. Stevens (1992)\(^2\) made a critical values table to determine this significance. On the other hand, Field (2000)\(^1\) states that “the significance of a loading gives little indication of the substantive importance of a variable to a factor”. For this to determine, the loadings have to be squared. Stevens (1992)\(^2\) then “recommends interpreting only factor loadings with an absolute value greater than 0.4 (which explain around 16% of variance)”. This is only possible in principal component analysis, though. In factor analysis the amount of explained variance is calculated in a different way. Thus, Stevens (1992)\(^3\) recommendation should be approached with

care! The second result of factor analysis is the factor scores. These factor scores can be useful in several ways. Field (2000) and Rietveld & Van Hout (1993) name the following:

1. If one wants to find out “whether groups or clusters of subjects can be distinguished that behave similarly in scoring on a test battery, [and] the latent, underlying variables are considered to be more fundamental than the original variables, the clustering of factor scores in the factor space can provide useful clues to that end.  
2. The factor scores can serve as a solution to multicollinearity problems in multiple regressions. After all, the factor scores are uncorrelated (in the case of orthogonal rotation); 
3. The factor scores can also be useful in big experiments, containing several measures using the same subjects. If it is already known in advance that “a number of dependent variables used in the experiment in fact constitute similar measures of the same underlying variable, it may be a good idea to use the scores on the different factors, instead of using the scores on the original variables”.

2.10- Structural Equation Modeling

The causal relationship between constructs and quantification of the effects between them is of interest to almost all researchers, regardless of discipline. Sewell Wright invented path analysis (1918) as a methodology to analyze systems of structural equations and to describe how a number of interesting constructs relate to each other. Three of the most important aspects of path analysis are the path diagram, the equations relating correlations or covariance’s to parameters, and the decomposition of effects. Methodological desires by researchers using path analysis have for example been: a) to be able to measure the latent variables of interest through multiple manifest variables in order to get better measurement, b) to be able to accommodate for measurement error, and c) to be able to statistically compare alternative models. Structural Equation Modeling (SEM) is a quantitative statistical method that was developed to satisfy these methodological desires. SEM combines the

benefits of path analysis, factor analysis and multiple regression analysis\(^1\) \& \(^2\). SEM is based on correlation statistics, i.e. the linear relationships between variables, and the common variance between the variables forms the basis for the analyses. SEM analyses and presents the degree of relationship between variables in terms of explained variance. A hypothesized model is tested statistically in a simultaneous analysis of the entire system of variables, to determine the extent to which the covariance or correlation matrix stipulated by the model, is consistent with the matrix based on the empirical data. If the statistical goodness of fit between the two compared matrices is adequate the model is a plausible representation of the relations between variables that the model developer has specified\(^3\) \& \(^4\). While most other multivariate procedures essentially are descriptive by nature (e.g. exploratory factor analysis), SEM takes a confirmatory (i.e. hypothesis-testing) approach to data analysis, although exploratory aspects can be addressed. Whereas traditional multivariate procedures are incapable of either assessing or correcting for measurement error, SEM provides explicit estimates of these parameters.

Kline (2005)\(^5\) describes three main differences between structural equation modeling and other approaches. First, SEM requires formal specification of a model to be estimated and tested. It forces the model developer to think carefully about their data and to formulate hypotheses regarding each variable. Second, SEM has the capacity to estimate and test relationships between latent variables. Third, SEM is a more comprehensive and flexible approach to research design and data analysis than any other single statistical model in standard use by social and behavioral scientists. Hoyle also describes SEM as similar to correlation analysis and multiple regression analysis in four specific ways. First, SEM is based on linear statistical models. Second, requirements such as independence of observations and multivariate normality will have to be met. Third, SEM promises no test of causality. It merely tests relations among different

---

variables. Finally, like any other quantitative analysis, post-hoc adjustments to a SEM model require cross-validations.1

The development of a structural equation model is supported by special software packages. The first and most widely spread software is LISREL which is an acronym for Linjära Strukturella Relationer (Linear Structural RELations). LISREL was originally developed by the two Swedish professors Karl Gustaf Jöreskog and Dag Sörbom. One of the earliest references to LISREL methodology is Joreskog (1973), and since then LISREL has been developed in several generations. Currently version 8.80 is the latest version. Several other software packages have been developed, where AMOS (SPSS, 2008) and EQS (MVSOFT, 2008) probably are the most widely spread, apart from LISREL. Structural equation modeling has been used for many years and is a popular methodology for non-experimental research, where methods for testing theories are not well developed, and ethical or practical considerations make traditional experimental designs unfeasible. Human factors researchers at the Swedish Defense Research Agency, Maud Angelborg-Thanderz and Erland Svensson, have used LISREL since 1984. A structural equation model may have one or more components. One component that is present in all structural equation models is the measurement model that defines the latent constructs through different manifest variables. Another important component is the structural model. The structural model tests relationships between the different latent variables.2 & 3

2.10.1- Structural equation model development process

Structural equation modeling (SEM) is almost a research field in itself and therefore only a brief introduction to the model development process is provided here. Introductionary texts concerning the SEM development process accessible for non-experts, are, for example, provided in Diamantopoulos & Siguaw (2000), Byrne (1998, 2001), and on the Internet4. Joreskog (1993)5

---

(1993)\(^1\) distinguishes between three scenarios of SEM use that he termed Strictly Confirmatory, Alternative Models, and Model Generating. In the Strictly Confirmatory scenario the researcher formulate a single model based on theory, collects the appropriate data, and then test the fit of the model to the collected data\(^2\). The researcher does no modifications to the model and either accepts or rejects the model. However, as other unexamined or nested models may fit the data as well or better, an accepted model is only a model that has not been rejected. In the Alternative Models scenario the researcher proposes several alternative competing theory-driven models. Based on the analysis of the collected data, the most appropriate model is chosen. Although this approach is desirable in principle, a problem is that in many specific research topic areas, the researcher does not find two or better developed alternative models to test. In the Model Generating scenario, the researcher proceeds in a more exploratory fashion, often after first having had to reject an initial model after assessment of its poor fit. Joreskog notes that although respecification may be either driven by theory or data, the goal is to find a model that is meaningful and statistically well fitting. The problem with the model development approach is that models developed in this way are post-hoc models, which may not be stable and may not fit new datasets\(^3\). By the use of a cross validation strategy, where the initial model is developed using one data sample and then tested against an independent sample, some of this concern can be addressed. For the model(s) presented in this thesis, the approach, as in many cases, most closely matches the Model Generating scenario. Regardless of which of these three approaches that have been chosen, SEM does not in itself provide clues concerning causality in a model, i.e. in what directions the effects go (and specifically in the modeling software, in which directions the arrows point). The causality has to be justified by theory and the good judgment by the researcher. In a description of the SEM development process, Joreskog & Sorbom (1993)\(^1\) describe the validation of the measurement model and the fitting of the structural model as the two main steps\(^3\). The validation of the measurement model and the fitting of the structural model as the two main steps\(^3\). The validation of the measurement model and the fitting of the structural model as the two main steps.
through path analysis with latent variables. The model that is being developed is specified on the basis of available theory. Constructs are chosen and operationalized by multiple manifest variables and tested through confirmatory factor analysis to establish that indicators seem to measure the corresponding constructs. The researcher proceeds to development of the structural model only when the measurement model has been validated. Two or more alternative models (one of which may be the null model) are then compared in terms of model fit, which measures the extent to which the covariances predicted by the model correspond to the observed covariances in the data. Modification indexes, suggested by the analysis software, may be used by the researcher to alter one or more model specifications to improve fit, but only if supported by theory. A solid theoretical foundation is thus needed before a structural equation model is developed, as theory warns us of potential problems such as, for example, excluded variables. Theoretical support is also necessary in order to distinguish between statistically equivalent models. Good definitions are also helpful when identifying appropriate manifest variables/measures. In another description of the SEM development process Diamantopoulos & Siguaw (2000) describes eight relatively distinct but related steps that a researcher goes through when developing a structural equation model:\footnote{Muthén, L.K., & Muthén, B.O. (2008). Mplus user’s guide (5th ed.). Los Angeles: Authors.}:

1. Model conceptualization
2. Path diagram construction
3. Model specification
4. Model identification
5. Parameter estimation
6. Assessment of model fit
7. Model modification
8. Model cross validation

Brief descriptions of the basic outline and considerations of each of Diamantopoulos’ & Siguaw’s steps will be provided below.
2.10.1.1- Model conceptualization

In this initial step the researcher defines his or her conceptual model, which translates theoretical assumptions into a conceptual framework. This conceptual model needs to be identified based on existing literature and theory. In this step, the researcher decides which latent variables or constructs that will need to be included, and how they are to be operationalized through manifest variables. During this stage, it is crucial to make every effort to include any important factors that can affect the variables that are included in the model. An omission of important factors represents a specification error and the result can be that the proposed model in the end does not represent the “whole” truth. Successful development of a structural equation model is to a large extent based on a sound model conceptualization. It is rare that a modeling process that does not start from well-established concepts or constructs and tested measures, result in a useful model.

2.10.1.2- Path diagram construction

In this second step of the modeling process the model developer can describe his or her model graphically as a path diagram. This is not a mandatory step, but it is helpful in order to make the model more explicit for the model developer.

2.10.1.3- Model specification

The third step is model specification, where the researcher specifies which effects that is null, which are fixed to a constant and which ones that vary through the specification of a command file for the analysis software. The researcher now needs to be very explicit on which variables that will be included and how they shall relate. The specification of a command file can be either through a text or a graphical format. Effects are represented by an arrow in a path diagram, while null effects result in the absence of an arrow. Note that the existence or absence of an arrow represents a rather strong theoretical assumption. A model where no effect is constrained to zero will always fit the data, and the closer one is to this most complex model, the better the fit of the model to the data. Thus, for a model where many effects are included in the specification, the fit

---

indices reported are better, but the model is also more complex and harder to grasp for the researcher.

### 2.10.1.4- Model identification

The fourth step in the process is model identification, which is performed by the analysis program, e.g. LISREL or AMOS. In this step the empirical data is investigated to see whether there is enough information in the data to do the parameter estimation that is performed in the next step, i.e. that a unique value can be identified for each parameter in the model. If there is a lack of information, i.e. the number of parameters estimated is less than the number of variances and covariance’s, the model, becomes under-identified and the analysis is cancelled. The model can also become just-identified or over-identified. If the number of parameters estimated are greater than the number of variances and covariance’s then the model is over-identified. To exemplify what is done during the model identification the following simple example can be used: Is there enough information to uniquely identify the values of A and B in the equation $A \times B = 100$? The answer is no, as there are several different possible solutions and this would equal to when a model is unidentified. However, if A is fixed to 10 you know that B has to be 10 and the equation can be identified.

### 2.10.1.5- Parameter estimation

If the model can be identified, the parameter estimation step can be executed. During the parameter estimation the analysis software create a covariance matrix based on the specified model. If there is no relation between two variables specified during the model specification the covariance is set to zero. The covariance matrix that is proposed by the model is then compared to the matrix produced by the data. The selection of method of estimation is also an important factor.

---

component of the model specification. Several methods of estimation can be used and ordinarily one will get similar estimates by any of the methods\(^1\). Maximum Likelihood estimation is by far the most common method and Garson (2008) recommends that it is used, unless the researcher has good reason or counterarguments. Unlike some of the other estimation methods, Maximum Likelihood does not assume uncorrelated error terms. Key assumptions are large samples, manifest variables with multivariate normal distribution, valid specification of the model, and manifest variables on an interval or ratio scale, although ordinal variables are widely used in practice. If ordinal data are used, they should have at least five categories and not be strongly skewed\(^1\).

### 2.10.1.6- Assessment of model fit

A large class of omnibus tests exists for assessing how well the model matches the observed data. \(\chi^2\) is a classic goodness-of-fit measure to determine overall model fit. The null hypothesis is that the implied or predicted covariance matrix \(\Sigma\) is equivalent to the observed sample covariance matrix \(S\), \(\Sigma=S\). A large \(\chi^2\) and rejection of the null hypothesis means that model estimates do not sufficiently reproduce sample covariance; the model does not fit the data well. By contrast, a small \(\chi^2\) and failure to reject the null hypothesis is a sign of a good model fit. However, the \(\chi^2\) test is widely recognized to be problematic\(^2\)&\(^3\). It is sensitive to sample size, and it becomes more and more difficult to retain the null hypothesis as the number of cases increases. The \(\chi^2\) test may also be invalid when distributional assumptions are violated, leading to the rejection of good models or the retention of bad ones. \(\chi^2\) is based on a very stringent hypothesis of \(\Sigma=S\)\(^4\). Due to these drawbacks of \(\chi^2\) test many alternative fit statistics have been developed, though each has its own advantages and disadvantages\(^5\). Another commonly reported statistic is the Root Mean Square Error of Approximation (RMSEA), a measure of fit introduced by Steiger and


Lind (1980)\(^1\). RMSEA “incorporates a penalty function for poor model parsimony” and thus becomes sensitive to the number of parameters estimated and relatively insensitive to sample size\(^2\). The Amos User’s Guide suggests that “a value of the RMSEA of about 0.05 or less would indicate a close fit of the model in relation to the degrees of freedom;” although “this figure is based on subjective judgment” and “cannot be regarded as infallible”\(^5\).

Comparative fit index (CFI) evaluates “the fit of a user-specified solution in relation to a more restricted, nested baseline model,” in which the “covariance’s among all input indicators are fixed to zero” or no relationship among variables is posited\(^3\). CFI ranges from 0 for a poor fit to 1 for a good fit. Tucker-Lewis index (TLI) is another index for comparative fit that “includes a penalty function for adding freely estimated parameters”. TLI can be interpreted in a similar fashion as CFI, but it can have a value outside of the range of 0 to 1. There is no single evaluation rule on which everyone agrees. Hu and Bentler (1999)\(^4\) provide rules of thumb for deciding which statistics to report and choosing cut-off values for declaring significance. When RMSEA values are close to .06 or below and CFI and TLI are close to .95 or greater, for example, the model may have a reasonably good fit. Therefore, it is recommended to report not only χ\(^2\) but RMSEA and CFI/TLI\(^5\).

### 2.10.1.7- Model modification

When a model have been evaluated with respect to its fit, the modeler can decide whether the model is acceptable or that it needs to be modified in order to fit the empirical data better. LISREL presents suggestions for model improvement, so called modification indices. These modifications are entirely data driven and careful deliberation and theoretical support must substantiate any changes to the model based of the modification indices\(^6\).

---

2.10.1.8- Model cross-validation

The last step of the modeling process is to do cross-validation of the proposed model against a new dataset, or a part of the dataset that have been kept aside for cross validation purposes. This step is extra important if major changes have been done to the model as a result of the model modification phase\(^1\) &\(^2\).

2.11- Research Type

Qualitative measures are measures without numerical data recording. Qualitative measures are including short written responses of surveys; interviews; video and audio data recording; and many other approaches in non-numerical format. So the qualitative research is a type of research that uses qualitative measures. Both quantitative and qualitative measures occur in almost any research topic that one can think of. In fact, in almost every applied social research project both of these approaches uses\(^3\). So there are three types of researches: Qualitative, quantitative and mixed mode.

Qualitative is predominantly used for any data collection such as interview or data analysis procedure that generates or use non-numerical data. According to Leavy (1994)\(^4\) in qualitative approach the aim is to explain rather than to predict phenomena and understand things rather than measure.

Quantitative research is most often used in studies with clearly stated hypotheses that can be tested. It focuses on well-defined, narrow studies. A quantitative method discusses the problem from a broader perspective, and often by providing a survey questionnaire with specific answer alternatives.

Tashakkori & Teddlie (2003)\(^5\) maintain individual quantitative and qualitative techniques and

---

1: Bae, B.R (2005), LISREL structural equation model: understanding and application. Chunglam Publishing Company, Korea
procedures do not exist in isolation. In choosing research methods researchers will either use single data collection technique and corresponding analysis procedures (mono method) or use more than one data collection technique and analysis procedures to answer their research questions (mixed methods).

Anyway, this study tends to use quantitative method. The researcher would use a questionnaire to get the needed information.

2.12- Research Design

This research has benefited from designing stated below. First with a review of literature and then with a definition on the problems which this research is going to target the context starts. Then by better defining the problems and with help of literatures different research questions extracted that in the next phase narrow down to several hypotheses. Different literatures presented to study this context including different effects that formulated in hypotheses. Then with help of gathered data different definitions on the statistical analyzes of this thesis will be given which clear the way that must be taken in chapter 4 which is analyzing these data. Finally, suggestions about the research problems and similar researches will be in the future.
2.13- Scope of the Research

The research is limited to Small and Medium Enterprises in India especially those around Pune.
2.14- Statistical Population

Sampling design begins with defining the target population. This is the collection of elements which contain the information needed by the researcher, about the conclusion and presumption to be made. A statistical population is a set of entities concerning which statistical inferences are to be drawn, often based on a random sample taken from the population. According to Saunders et al. (2000), the target population should be defined in terms of elements, sampling units, extent and time. The object from or about which the information is respondent. A sampling unit is an element or unit containing elements, which is available for selection at some stages of the sampling process. Extent refers to the geographical boundaries of the research and the time refers to the period under consideration. According to the explanations mentioned above, the target population of the current study is defined as follow

2.15- Sample Selection

The population of the research will be all the Small and Medium Enterprises in Pune which are using e-commerce applications and own internet website which is estimated to be 395 units. Out of 395 units, a sample of 120 SME units is to be taken to obtain the primary data. The scale used in this research is called the Likert scales, where the five alternatives are strongly disagreeing (1), disagree (2), neutral (3), agree (4), and strongly agree (5) respectively.

Therefore, Standard deviation:

\[ \sigma = \frac{\max(x_i) - \min(x_i)}{6} = \frac{5-1}{6} = 0.667 \]

And size of population is:

\[ n = \frac{N \times \sigma^2 \times (x_{1/2})^2}{\varepsilon^2 (N-1) + (x_{1/2})^2 \times \sigma^2} = \frac{395 \times (1.96)^2 \times (0.667)^2}{(0.1)^2 (395-1) + ((1.96)^2 \times (0.667)^2)} = 120 \]

Where:

To conduct research, primary and secondary data will be obtained. Primary data will be obtained from the sample SMEs using a survey questionnaire. Secondary data will be collected using various Government related publications, websites and magazines.

### Table 2.1: A comparison of Primary and Secondary Data

<table>
<thead>
<tr>
<th>Collection purpose</th>
<th>Primary Data</th>
<th>Secondary data</th>
</tr>
</thead>
<tbody>
<tr>
<td>For the problem at hand</td>
<td>For the problems</td>
<td></td>
</tr>
<tr>
<td>Collection process</td>
<td>Very involved</td>
<td>Rapid and easy</td>
</tr>
<tr>
<td>Collection cost</td>
<td>High</td>
<td>Relatively low</td>
</tr>
<tr>
<td>Collection time</td>
<td>long</td>
<td>Short</td>
</tr>
</tbody>
</table>

#### 2.16- Sampling Technique

In this study, we used **Stratified Random Sampling** for collecting necessary data. This sampling technique (also sometimes called proportional or quota random sampling) is a **probability sampling** method that involves dividing the population of study into homogeneous subgroups and then taking a simple random sample in each subgroup. There are several major reasons why we prefer stratified sampling over simple random sampling. First, it assures that we will be able to represent not only the overall population, but also key subgroups of the population, especially small minority groups. Second, stratified random sampling will generally have more statistical

---

precision than simple random sampling. This will only be true if the strata or groups are homogeneous. If they are, we expect that the variability within-groups are lower than the variability for the population as a whole. Stratified sampling capitalizes on that fact. Advantages over other sampling methods:

- Focuses on important subpopulations and ignores irrelevant ones.
- Allows use of different sampling techniques for different subpopulations.
- Improves the accuracy/efficiency of estimation.
- Permits greater balancing of statistical power of tests of differences between strata by sampling equal numbers from strata varying widely in size.

2.17- Research Instrument

Questionnaires are one of the most popular methods of conducting scholarly research. They provide a convenient way of gathering information from a target population. The questionnaire for the thesis is divided into four sections. Section I is again divided into subsection A and subsection B. Subsection A is designed to collect information related to the various usage of electronic commerce namely electronic marketing, electronic advertising, electronic customer support service, electronic order and delivery and electronic payment system. The questions are made up of simple type, which every item required the respondents to choose one answer out of five alternatives. Subsection B is to collect information related to organizational performance. There are 39 items, which can be categorized into operational performance, financial performance and market-based performance. Every item required the respondents to choose one answer out of five alternatives. The scale used is called the Likert scales, where the five alternatives are strongly disagreeing, disagree, neutral, agree, and strongly agree respectively. Section II is designed to collect information related to the benefits received from using e-commerce applications. There are 13 items where the respondents choose one answer from each item, which is of Likert scale type. There is also one open-ended question. Section III: is designed to collect information related to the problems of e-commerce application in small and
medium enterprises. Section IV is designed to collect information related to the company profile. There are 8 questions where the respondents are required to tick one answer or fill in the blank.

2.18- Reliability and Validity Analysis

Reliability refers to the credibility of the test, and it mainly tests measurements result and measurement tools’ stability and consistency.

Validity refers to correctness; it means the measurement tool can correctly measure the characteristics and functions.

Cronbach’ α coefficient is the most extensive measurement method of reliability. A coefficient is distributed between 0-1, on the minimum acceptable reliability coefficient, scholars are not unanimously. Here we use the theoretical more common practice, namely under general circumstances, if Cronbach’ α coefficient ranges between 0.80-0.90 and it is considered very good, ranges between 0.700-0.80 it is relatively good, and between 0.65-0.70 is acceptable, 0.60-0.65 is unacceptable. For not yet proven variables, as long as Cronbach’ α is above 0.60, then it can be accepted. For determining reliability of the study Cronbach’s Alpha method has been used. Followings have been resulted from this analysis:
Table 2.2: The reliability of the pretest research

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Number of Questions</th>
<th>Cronbach’s coefficient alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electronic Marketing (EM)</td>
<td>3</td>
<td>0.809</td>
</tr>
<tr>
<td>Electronic Advertising (EA)</td>
<td>4</td>
<td>0.702</td>
</tr>
<tr>
<td>Electronic Costumer Service (ECS)</td>
<td>5</td>
<td>0.768</td>
</tr>
<tr>
<td>Electronic Order &amp; Delivery (EOD)</td>
<td>4</td>
<td>0.750</td>
</tr>
<tr>
<td>Electronic Payment System (EPS)</td>
<td>4</td>
<td>0.870</td>
</tr>
<tr>
<td>All questions related to Applications of E Commerce (AeC)</td>
<td></td>
<td>0.858</td>
</tr>
<tr>
<td>Operational Performance (OPP)</td>
<td>15</td>
<td>0.883</td>
</tr>
<tr>
<td>Financial Performance (FP)</td>
<td>15</td>
<td>0.754</td>
</tr>
<tr>
<td>Market Based Performance (MBP)</td>
<td>9</td>
<td>0.744</td>
</tr>
<tr>
<td>All questions related to Organizational Performance</td>
<td></td>
<td>0.845</td>
</tr>
</tbody>
</table>

For determining validity of the questionnaire content credit has been used. Content credit of this questionnaire has been justified by guide professors and also initial distribution of questionnaire among number of experts, scholars and considering their corrective comments, it has the necessary credibility.

2.19- Data Gathering

Data gathering may range from a simple observation at one location to a grand survey of multinational corporations in different parts of the world. The method selected will vastly determine how the data are collected. Questionnaires, standardized tests, observational forms, laboratory notes and instrument calibration logs are among the devices used to record raw data.\(^1\)

As mentioned above, data gathering will be performed by referring to reliable official references in the country and abroad. Thus, data includes both primary and secondary sources. The primary

source, which is the case study conducted that includes quantitative data as well as the official Government data will be entered in SPSS 11.5 statistic software for performing, needed analysis.

2.20- Data Analysis

For analyzing the data derived from questionnaire regression analysis will be used and the software which will be used for analyzing the data is LISREL 8.5 and SPSS 16.

All data from the respondents will be coded and entered into the data sheet of the SPSS program for statistical analysis. The following analysis and tests will be carried out:

i) In the first part of the analysis shall consist of various tabulation data, where frequency and percentage will be displayed.

ii) In the second part of the analysis, cross tabulation will be carried out to establish relationship between the level of usage of e-commerce and organizational performance. At the same time, chi-square test and t-test will be carried out to determine whether the relationship is significant or not.

iii) In the third part of the analysis, Friedman test will be carried out to prioritize the factors that are benefitted from the e-commerce applications in terms of influence percentage.

iv) In the fourth part of the analysis, Friedman test will be carried out to prioritize the factors that are cause problems due to the use of the e-commerce applications in terms of influence percentage.