CHAPTER 4

RESEARCH METHODOLOGY

This chapter presents the methodology used in the present study to empirically test the proposed model and the various hypotheses stated. This chapter covers; research design, development and description of constructs and their measurement, sample size and sampling technique and statistical techniques used for the purpose of analysis.

4.1. RESEARCH DESIGN

The research design lays out the blueprint or framework for conducting the research and providing the appropriate procedures for collecting and analyzing the required information for answering the research question (Churchill, 1979) and it relates to the decisions regarding the purpose of the study, i.e. exploratory or conclusive, descriptive or casual, cross-sectional or longitudinal (Saunders et al., 2003), study setting, type of investigation, extent of any research inference, unit of analysis, sampling size, data collection method and measurement and measures. A research design locates the researcher in the empirical world, and links the research question to data (Punch, 2005).

Research design can be classified on the basis of function and techniques. On the basis of function, research can be classified into exploratory, descriptive and causal studies. Based on technique, business research includes experiments, surveys and observational studies (Zikmund, 1997). Broadly, research design can be classified into exploratory and conclusive. Conclusive research is further categorized into descriptive and causal research.

Exploratory Research Design: Exploratory research design is used to explain the thoughts and opinions about the research problem or the respondent population or to provide insights on how to do more conclusive (causal) research (Nargundkar, 2003).
Descriptive research: Descriptive research is concerned with the research question starting with who, what, when, and where, usually using standardized questionnaires with the main aim of describing the characteristics of a population or a phenomenon (Zikmund, 1997). Descriptive research is characterized by a planned and structured research design, preceded by clear specification of the research problems, questions and prior formulation of specific hypothesis (Churchill, 1987). The purpose of descriptive research is the determination of relationship or association between two variables and the extent to which they covary (Churchill, 1987).

Descriptive research can be either cross sectional or longitudinal. A cross-sectional study is a positivistic design to gain information, can be done either just one time or over a period of time (Collis and Hussey, 2003) concurrently a longitudinal study is positivistic strategy which involves the study of a variable or group of subjects over a long period of time (Collis and Hussey, 2003). In longitudinal study, the data is collected for more than one time to study the effect of change. Causal research is concerned with ascertaining cause and effect relationships among variables. In addition to it, causal studies concerned with why question about how one variable have an influence or is responsible for changes in another variable, and their relationship can be depicted as either symmetrical, reciprocal or asymmetrical (Copper Schindler 2006; Emory & Cooper, 1991). A cross sectional survey will allow studying relationships between variables and making predictions regarding which antecedents lead to customer orientation among service employees. A cross sectional survey enable to collect data from many respondents in a relatively short amount of time.

In the present study, descriptive research in combination with causal research (hypothesis testing) is applied in order to answer the research questions. The scales to measure different constructs have been used for the purpose of survey and thorough statistical analysis has been followed to analyze the data. The hypotheses testing was first carried out by classical statistics analysis, e.g. factor analysis, reliability test, correlations test and followed by confirmatory factor analysis by structured equation.
modeling. Once the decision is made to use descriptive research then next decision needs to be considered is to whether use cross-sectional or longitudinal designs (Malhotra, 2007). Time and budget constraint is another reason to choose cross-sectional design to gather data (Saunders et al., 2003). Cross-sectional design has been used because relationship among the variables at a given point in time is studied it did not consider changes or development in the relationship between the study variables.

In the present study an attempt has been made to answer the main research question by testing a proposed model. The deductive research approach is followed which requires numerical data and is appropriate because existing theory is used to construct a research model and to state the hypothesis for the purpose of testing.

4.2 SAMPLING

Sampling refers to the methods that researchers use to select the sampling unit, i.e. institution, individuals, groups, objects, or phenomena that they actually want to study (Thyer, 2001). Population is the universe of units from which the sample is to be selected (Bryman and Bell, 2003). The target population should be accurately selected on the basis of elements, sampling unit, extent and time. The universe of the present study is all the public as well as private sector banks operating in India. The present study is confined to all banks including public as well as private banks operating in the state of Punjab. The following multiple decisions were required to be made before the final selection of the samples.

a. What should be the criteria for selection of banks?

b. How many banks to be selected for the study?

c. Which part of the state of Punjab to be covered for the selection of banks?

d. How many branches of each selected bank for the selected area of Punjab to be selected?

e. What should be the criteria of selection of branches for each selected bank for the selected area?
f. How many respondents including employees and customers to be selected per branch?

g. What should be the criteria of selection of respondents including employees and customers?

In order to collect required data, top 13 public and private sector banks from three cities - Amritsar, Jalandhar and Ludhiana - of Punjab were selected on the basis of the ranking reported by financial express (Dec, 2009). Ten percent of the total branches of each bank in each city were selected on the basis of their annual business. Thereafter, five employees from each bank in each city were selected. The respondents selected for the present study were front line employees which included single window operators, customer service representatives and tellers who spent their time dealing directly with customers for transactions or responding to customer queries. Total sample from 96 branches is 480 (96*5). A total of 500 questionnaires were distributed, 424 responses were retrieved to analyze the data. A multi stage random sampling technique was used to address the responses at the rate of 85%. The questionnaire designed for the study was distributed to the selected respondents. In case of customers, total sample size is 448. A total of 500 questionnaires were distributed, 448 responses were retrieved to analyze the data. A random sampling technique was used to address the responses at the rate of 90%. Table 4.1 shows the criterion used to select the banks, branches and employees from each bank in each city.
TABLE 4.1
DISTRIBUTION OF RESPONDENTS: BANKS, BRANCHES AND EMPLOYEES

<table>
<thead>
<tr>
<th>Bank Top 10 Banks</th>
<th>Branches of Banks</th>
<th>Total No of Branches in selected three cities</th>
<th>10 % selection of Branches of Banks Each cities</th>
<th>Total No. of selected branches</th>
<th>No. of Employees (5 Employees per Branch)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Amritsar</td>
<td>Jalandhar</td>
<td>Ludhiana</td>
<td>Amritsar</td>
<td>Jalandhar</td>
</tr>
<tr>
<td>SBI</td>
<td>33</td>
<td>42</td>
<td>54</td>
<td>129</td>
<td>3</td>
</tr>
<tr>
<td>BOB</td>
<td>5</td>
<td>7</td>
<td>12</td>
<td>24</td>
<td>1</td>
</tr>
<tr>
<td>UBI</td>
<td>12</td>
<td>20</td>
<td>17</td>
<td>49</td>
<td>2</td>
</tr>
<tr>
<td>BOI</td>
<td>17</td>
<td>12</td>
<td>32</td>
<td>59</td>
<td>3</td>
</tr>
<tr>
<td>PNB</td>
<td>41</td>
<td>46</td>
<td>57</td>
<td>144</td>
<td>4</td>
</tr>
<tr>
<td>IDBI</td>
<td>2</td>
<td>5</td>
<td>9</td>
<td>16</td>
<td>1</td>
</tr>
<tr>
<td>PSB</td>
<td>37</td>
<td>26</td>
<td>29</td>
<td>92</td>
<td>4</td>
</tr>
<tr>
<td>OBC</td>
<td>20</td>
<td>31</td>
<td>38</td>
<td>89</td>
<td>2</td>
</tr>
<tr>
<td>CB</td>
<td>10</td>
<td>16</td>
<td>13</td>
<td>39</td>
<td>1</td>
</tr>
<tr>
<td>HDFC</td>
<td>24</td>
<td>27</td>
<td>35</td>
<td>86</td>
<td>2</td>
</tr>
<tr>
<td>AXIS</td>
<td>12</td>
<td>11</td>
<td>22</td>
<td>45</td>
<td>1</td>
</tr>
<tr>
<td>YES</td>
<td>2</td>
<td>7</td>
<td>5</td>
<td>14</td>
<td>1</td>
</tr>
<tr>
<td>ICICI</td>
<td>10</td>
<td>12</td>
<td>16</td>
<td>38</td>
<td>1</td>
</tr>
<tr>
<td>Total No. of branches</td>
<td>225</td>
<td>262</td>
<td>339</td>
<td>824</td>
<td>26</td>
</tr>
</tbody>
</table>
4.2.1 Sampling Technique

The most important decision is to select the sampling method to be used to select the sample for the study. The selection of a sampling method depends on the nature of the research question, time and money, desired accuracy level, and the data gathering method. There are two main sampling methods, probability and non-probability sampling (Zikmund, 2000). For this study, the probability method was chosen due to the universal acceptance of results and high generalizability of results due to the availability of the sampling frame. The most common probability sampling techniques includes simple random, systematic, stratified and cluster. From the list of probability sampling methods, multi stage random sampling approach was used to collect the data.

4.2.2 Determine Sample Size

The sampling size is critical in achieving statistical significance. Roscoe (1975) reported four rules of thumb in determining the proper sample size: (1) the size of sample that is appropriate for most research is between 30 to 500 samples, (2) a minimum sample size of 30 is necessary for each sub-group when samples are broken into subsamples, (3) in multivariate research, the sample size should be several times as large as the number of variables in the study, (4) sample size as small as 10 to 20 is possible for the simple experimental research with tight experimental control. The techniques that can be used to determine the adequacy of sample size depends on the statistical methods used. Numerous scholars have reported that much larger sample is required when using structural equation modeling techniques to obtain stable parameter estimates and standard errors (Schumacker & Lomax, 2004). Different scholars have suggested different guidelines for determining proper sample size in applying SEM. Boomsma (1983) recommended 400 would be sufficient. Hair et al., (2010) recommended five considerations to determine the sample size when using Structural Equation Modeling (SEM) techniques. First, if the distribution of the data deviates from the assumption of multivariate normality, then 15 respondents for each parameter is an acceptable number to minimize the problem of deviation from normality. Second, the sample size should range from 150 to 400 respondents, if the estimation technique is to be used, if the sample size exceeds 400, then the MLE method becomes more sensitive and results of the
goodness-of-fit measures become poorer. Third, the more constructs model has, the more parameters should be used in the analysis and as a result the more sample size is needed to conduct the analysis. Fourth, the more missing data research has, the greater sample size a study needs. Fifth, researchers required to consider communalities (average error of variance of indicators, and represent the average amount of variation among the measured/indicator variables explained by the measurement model) before deciding the proper sample size. Communalities should be above 0.5 (equals 0.7 standardized loading estimates); otherwise the study requires more sample size. For instance, Hair et al. (2006) assert that if any communality is between 0.45-0.55, or the model has constructs with fewer than three items, then the sample size should be above 200. On the other hand, if the communalities are lower than 0.45 then the minimum sample size should be 300 or more.

4.3 DEVELOPMENT OF THE QUESTIONNAIRE

Data for present study was collected using a questionnaire. Questionnaire was developed based on the procedures suggested by Churchill and LaCubucci (2005). The early stage of development of the questionnaire involved planning in which were considered the research objectives, conceptual model and the hypotheses to ensure that data gathered would address all issues. According to Czaja and Blair (1996), “at this stage we must decide the goals of the research and determine how best to accomplish them within the available time and resources”. The next important step is to determine what to include in questions and how to ask questions so desired information can be gathered. Malhotra (2010) points out that self-administered questionnaire must include simple questions and detail instructions. Self-administered questionnaire is described as a data collection technique in which the respondent reads the survey questions and records his or her own responses without the presence of a trained interviewer” (Hair et al., 2006). Self-administered questionnaire method of collecting data is used because it fits the current study philosophy (positivism), and approach (deductive). The self-administered questionnaire form used within this thesis is called a drop-off survey. This method involves the researcher traveling to the respondents’ location and a representative of the
researcher (i.e., front-desk staff in this research) hand delivering survey questionnaires to respondents. Following this, the completed surveys are picked up by the representative after the respondents have finished (Hair et al., 2006; Zikmund, 2003). Furthermore, other means of survey data collection such as mail, web-based survey, and telephone were inappropriate to collect the data.

To test the hypotheses, the questionnaire was developed, as shown in appendix 1. There were two sections in the questionnaire. The first section of the questionnaire consisted of major variables established for the study; viz., customer orientation of service employees, job satisfaction. Constructs were measured by using multiple-item measures. Second section was mainly to obtain the demographic information of the respondents. The demographics include age, gender, education, working experience in banking industry, length of service. The demographic question was put last because of the general understanding that the respondents would feel relaxed in answering the simple questions and would feel that the survey was nearly over.

All questions in the questionnaire were designed as closed questions with multi choice answers. The closed-end question is a question in which respondents must choose the answer from a number of predetermined alternatives (Collis and Hussey, 2003). Closed question are very convenient and reliable for collecting factual data and are usually easy to analyze (Collis and Hussey, 2003) and more appropriate to the current study’s philosophy (positivism) and approach (deductive) (Saunders et al., 2003; Zikmund, 2003). As a result, closed questions with multi-choice answers were employed in the current study questionnaire.

4.4 MEASUREMENT OF VARIABLES

A list of scales used to measure the different constructs of the proposed model of customer orientation, its antecedents and consequences is provided in table 4.4. Though, the scales used in the present study are tested and validated scales. However, for the scales to be appropriate in the specific context of the present study, certain modifications of the scales where needed are done. Constructs were measured by using multiple-item measures. A five-point Likert scale anchored from strongly disagree to strongly agree to
measure respondents’ level of agreement or disagreement with the items in question. A symmetric scale, a neutral midpoint, and which considers the distances between the categories to be equal is used. That is, the distance between “Strongly Agree” to “Agree” is the same as the distance between “Strongly Disagree” and “Disagree”.

**TABLE – 4.4 CONSTRUCTS**

<table>
<thead>
<tr>
<th>No.</th>
<th>Constructs</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Employees CO</td>
<td>Saxe and Weitz (1982)</td>
</tr>
<tr>
<td>2</td>
<td>Job Satisfaction</td>
<td>Hartline and ferrell (1996)</td>
</tr>
<tr>
<td>3</td>
<td>Affective Commitment</td>
<td>Hunt, Chonko &amp; Wood (1985)</td>
</tr>
<tr>
<td>4</td>
<td>Organizational Identification</td>
<td>Wieseke et al., 2007</td>
</tr>
<tr>
<td>5</td>
<td>Self-efficacy</td>
<td>Hartline and Ferrell, 1996</td>
</tr>
<tr>
<td>7</td>
<td>Customer Oriented Culture</td>
<td>Schneider, Parkington &amp; Buxton (1980)</td>
</tr>
<tr>
<td>8</td>
<td>Personality Traits (Extraversion, agreeableness and conscientiousness)</td>
<td>Shogo Iwanaga, 2007</td>
</tr>
<tr>
<td>9</td>
<td>Employee Selling Orientation</td>
<td>Saxe and Weitz (1982)</td>
</tr>
<tr>
<td>10</td>
<td>Superior Value Creation</td>
<td>Guenzi, Luca &amp; Troilo (2011)</td>
</tr>
<tr>
<td>11</td>
<td>Long term Strategic Orientation</td>
<td>Guenzi, Luca &amp; Troilo (2011)</td>
</tr>
<tr>
<td>12</td>
<td>Employee Customer orientation</td>
<td>Michaels and Day (1985)</td>
</tr>
<tr>
<td>13</td>
<td>Employee Selling Orientation</td>
<td>Michaels and Day (1985)</td>
</tr>
<tr>
<td>14</td>
<td>Customer Satisfaction</td>
<td>Donavan et al., (2001)</td>
</tr>
<tr>
<td>15</td>
<td>Customer Commitment</td>
<td>Huang, 2008</td>
</tr>
<tr>
<td>16</td>
<td>Customer Retention</td>
<td>Palmatier, 2006</td>
</tr>
</tbody>
</table>

The methodology adopted in order to test the underlying twenty hypotheses is discussed in the following chapter. This includes an overview of the methodology used,
measurement development, data collection tool, sampling design, data collection procedures, analytical techniques, and finally issues related to reliability and validity.

4.5 DATA ANALYSIS TECHNIQUES

In order to analyze quantitative data gathered from questionnaire successive stages of analysis: preliminary analysis (screening data prior to analysis), descriptive analysis, and multivariate analysis were performed. SPSS and AMOS software packages were used to analyze the data. Descriptive statistics were initially conducted to provide an overview of the sample. Then, the reliability tests and EFA (exploratory factor analysis) were applied to the data in order to conduct a preliminary test of the validity and reliability of the instrument. The purpose of this test is to assess the scales used to measure the constructs, i.e. to refine the measures (Churchill, 1979; Peter, 1979); the refinement is based on reliability and dimensionality. Refined measurement scales were then subjected to a validation phase through confirmation factor analysis (CFA) as a method to finalise the scales (Marsh et al., 1988). The following section discusses the procedures for scale refinement and validation. Once each scale is assessed to be uni-dimensional, reliable, and valid, structural model was employed to investigate the hypothesized relationships between constructs. The results of these data analysis methods are presented in the next chapter.

4.5.1 Descriptive Statistics Analysis

Descriptive statistics is useful for summarizing large sets of data into simple and meaningful figures. In order to understand the distributions of each construct, descriptive statistics analysis is utilized to describe the characteristics of each variable, means, variations and standard deviations. Each of these measures is briefly explained below:

Mean

Mean the sum of the measurements divided by their number, it is the most frequently used method to find the accurate average of a set of data. However, as the only measure of central tendency that uses all the data values in a sample or population, the mean has one great weakness as it influenced by extreme values (outliers).
The Standard Deviation

The Standard Deviation is a value that measures how far away each value in a set of data is from their mean, where a large standard deviation indicates that the data points are spread far from the mean and a small standard deviation indicates that they are clustered closely around the mean. If all the data values are equal, then the standard deviation is zero. Standard deviation is the most frequently used method to measure variability (dispersion) of a set of data as it gives a good picture of how the data is spread around, but it is still influenced by extreme scores (outliers) (Bryman and Cramer, 2005).

Descriptive statistics also provide some information concerning the distribution of scores (skewness and kurtosis). Skewness gives an indication of the symmetry of the distribution; a skewed variable is a variable whose mean is not in the middle of the distribution, while Kurtosis, on the other hand, provides information about the peakedness of a distribution; a distribution can be too peaked (will show short, thick tails) or too flat (will show long, thin tails) (Tabachnick and Fidell, 2007). Positive skewness values show positive skew where the scores clustered to the left-hand side of the graph, while negative skewness values on the other hand show a clustering of scores at the right-hand side of a graph (Pallant, 2007).

4.5.2 Purification and Reliability of Measurement Scale

The first step is to refine and purify the measurement scales. In order to purify measurement scales, the researcher conducted two analyses: reliability and EFA. The reliability of each scale was assessed via Cronbach alpha and item to total correlation.

4.5.2.1 Factor Analysis

Factor analysis is a variable reduction technique which identifies the number of latent constructs and the underlying factor structure of a set of variables during the initial stage of scale development (Netemeyer et al., 2003). To reduce the number of items and extract factors, the researcher employed the principal component analysis technique (Netemeyer et al., 2003; Hair et al., 2006, p.112). A Varimax rotation was then applied to initially-
extracted factors. In addition, Fabrigar et al., (1999) stated that the pattern of loadings was almost always found to be the same with Varimax as with oblique rotation. To assess the factorability of items, Kaiser-Meyer-Olkin (KMO) measure of sample adequacy and Bartlett’s test of Sphericity and communalities were used to test the suitability of the data for factor analysis. In addition, the researcher evaluated the factorial solutions (e.g., item loadings and percentage of variance extracted) obtained from SPSS.

4.5.2.2 Cronbach Alpha

The Cronbach alpha measures the correlations between the items belonging to a factor (Iacobucci & Churchill, 2010). Coefficient alpha is a measure of squared correlation between observed scores and true scores. Reliability is measured in terms of the ratio of true score variance to observed score variance. It can evaluate the internal consistency of each factor (Robinson, Shaver and Wrightsman, 1991). Nunnally (1978) states that a coefficient alpha greater than 0.70 represents a good indication of internal consistency. DeVellis (1991) recommends the following guidelines for coefficient alpha values: “below 0.60, unacceptable; between 0.60 and 0.65, undesirable; between 0.65 and 0.70, minimally acceptable; between 0.70 and 0.80, respectable; between 0.80 and 0.90, very good”.

4.5.2.3 Item to Total Correlation

Another separate-item measure of the scale’s consistency is the correlation of each item to the sum of the remaining items in one factor, commonly referred to as the item-to-total correlation. The approach assumes that the total score is valid and thus the extent to which the item correlates with the total score is indicative of convergent validity for the item. If low item-to-total correlations are evident, these results indicate that the items do not come from the domain of the same construct and will lead to higher levels of error and low levels of reliability (Churchill, 1979).

4.6 STRUCTURAL EQUATION MODELING

Structural Equation Modeling (SEM) has grown to be one of the main techniques of data analysis that attract many scholars across different disciplines and progressively
more in the social sciences. Structural Equation Modeling (SEM) is “a collection of statistical techniques that allow a set of relationships between one or more independent variables, either continuous or discrete, and one or more dependent variables, either continuous or discrete, to be examined” (Tabachnick and Fidell, 2001). Structural equation modeling (SEM) is a powerful statistical technique that combines the measurement model (confirmatory factor analysis) and the structural model (regression or path analysis) into a simultaneous statistical test (Garver, 1999).

Therefore, researchers have found SEM to be an appropriate technique to examine their hypothesized models (Palmatier et al., 2006; Wang et al., 2006). SEM has the ability to assess the uni-dimensionality, reliability and validity of each individual construct (Anderson and Gerbing, 1988; Bollen, 1989; Hair et al., 1998; Kline, 1998). Further, it provides an overall test of model fit and individual parameter estimate tests simultaneously, thus, providing the best model fits to the data adequately. In this thesis, SEM using confirmatory factor analysis, therefore, has been conducted. Structural equation modeling software AMOS was used.

The use of SEM has numerous advantages over conventional means of analysis, such as conducting a series of multiple regression analysis for each dependent variable in the model. SEM can be used as a confirmatory factor analysis tool to test the dimensionality and validity of each construct within the model (Kline, 1998). SEM is also capable of examining a system of hypothesized equations with multiple dependent variables simultaneously. SEM permits the assessment of the model’s performance as a whole by providing multivariate goodness-of-fit indices and permits the researcher to control for measurement error for each construct in the model (Hair et al., 1998). It also allows for the comparative analysis of the proposed model to other equivalent and restricted models as well as thorough evaluation of potential model modifications (Kline, 1998). According to Anderson and Gerbing (1988), SEM is a confirmatory method providing “a comprehensive means for assessing and modifying theoretical models”.
4.6.1 Two-Stage Structural Equation Modeling

In this thesis, the two-stage approach adhered by Anderson and Gerbing (1982) was adopted to conduct the analysis. The first (measurement model) stage of the analysis was conducted by specifying the relations between the observed variables and latent variable (Hair et al., 2006). The aim of this stage was to verify the uni-dimensionality, reliability and validity of the constructs.

4.6.1.2 Confirmatory Factor Analysis (First Stage of SEM)

Confirmatory factor analysis was used to estimate the measurement properties of latent construct prior to testing the structural relationships among them (Hair et al., 2010). Following the two stage approach adhered by Anderson and Gerbing (1988), confirmatory factor analyses was performed on each of the construct of interest as well as the on all the constructs of interest simultaneously to check the dimensionality of the constructs and validity of the measures. The use of CFA and construct validity assessments enables the researcher to assess the quality of their measures within a measurement model. Confirmatory factor analysis (CFA) is a better method for use to validate and improve the psychometrics properties. Therefore, CFA was used to confirm the existence of a specific factor structure. Items that loaded weakly on the hypothesized factors were removed from the scale, thus resulting in a uni-dimensional scale (Dunn et al., 1994). In using CFA, a factor loading of 0.50 and above on a specified factor has been considered acceptable (Hair et al., 1995), and thus this level is used as the cut off value within this thesis. The measurement model was tested by using CFA was conducted in two steps i.e., (i) assessment of the uni-dimensionality (ii) assessment of reliability and validity.

4.6.1.2.1 First Step- Assessment of the Uni-dimensionality

Uni-dimensionality refers to the existence of single construct/trait underlying a set of measures (Hattie, 1985; Sureshchander, Rajendran & Anantharaman, 2002). Uni-dimensionality occurs when a set of items forming an instrument all measure just one thing in common (Paulin, Ferguson & Bergeron, 2006). Traditionally, EFA was employed to explore dimensionality of constructs. However, as Anderson and Gerbing
(1988) suggest that exploratory factor analysis cannot assess uni-dimensionality directly, but aims to assess the factor structure of a scale. Confirmatory factor analysis (CFA) is a more rigorous and precise test of uni-dimensionality as compared to traditional techniques such as exploratory factor analysis (EFA) (Garver, 1999). Exploratory analysis typically does not provide an explicit test of uni-dimensionality. In addition, exploratory factor models do not provide explicit test statistics for ascertaining convergent and discriminant validity (Gerbing and Anderson, 1988). Hunter and Gerbing (1982) stated that “EFA is a poor ending point for the construction of a uni-dimensional scale”. However, exploratory factor analysis can only suggest, not confirm, dimensions (Churchill, 1979) whilst scale uni-dimensionality can only be established via confirmatory factor analysis (Gerbing and Anderson, 1988). Once uni-dimensionality has been achieved, the next step is to assess the reliability and validity because uni-dimensionality is necessary for construct validity (McGartland, Berg-Weger & Tebb, 2001) but alone is not sufficient to ensure the usefulness of a scale (Gerbing and Anderson, 1988). For this purpose, confirmatory factor analysis using maximum likelihood estimate was performed (Anderson and Gerbing, 1988; Kline, 2005). Following this, relationships between the underlying theoretical latent constructs were specified in the structural model (second stage).

**Two step approach to SEM (Structural Equation Modeling)**

- **Stage-I**
  - Assessment of Measurement Model (CFA)
    - Step I: Uni-dimensionality
    - Step II: Assessing Reliability and Validity
      - (CR >0.50, AVE >0.50, Convergent and Discriminant Validity)

- **Stage-II**
  - Assessment of Structural Model
    - Testing Hypothesized Relationships between constructs
4.6.1.2.2 Step-II Reliability and Validity

Reliability is the degree to which measures are free from error and therefore yield consistent results (Zikmund 2000; McDaniel and Gates 2006). That means reliability is the extent to which research findings would be the same if the research were to be repeated. Therefore, the main objective of reliability is to minimize the errors and biases in a research. The reliability of each scale was assessed via Cronbach alpha and item to total correlation as discussed in section (4.5.2.2 and 4.5.2.3). Other authors such as Steenkamp and Van Trijp (1991) suggested that CFA provides a better estimate of reliability than coefficient alpha. Furthermore, assessing reliability by using CFA is essential to ensure that all the individual items are consistent in their measurements (Hair et al., 1998). Internal consistency has been assessed using confirmatory factor analysis (CFA). Reliability as being evaluated in CFA is the composite reliability and average variance extracted.

4.6.1.2.3 Composite Reliability

Composite reliability is a better indicator than Cronbach alpha because it is free from the assumption of equal item reliabilities (Gerbing and Anderson, 1988; Hair et al., 1998). CR measures the internal consistency of a set of measures rather than the reliability of a single variable to capture the degree to which a set of measures indicates the common latent construct (Holmes-Smith et al., 2006). Here, a main advantage is that CR is based on estimates of model parameters and has wide applicability.

\[
\text{Composite Reliability} = \frac{(\Sigma \text{standardized loading})^2}{((\Sigma \text{standardized loading})^2 + \Sigma e_j)}
\]

4.6.1.2.4 Average Variance Extracted

AVE estimate is a more conservative indicator of the shared variance in a set of measures than construct reliability. Hence, the variance-extracted estimate reflects the overall amount of variance in the items accounted for by the latent construct. In this thesis, CR and AVE have been calculated separately for each multiple item construct because AMOS does not compute these two measures directly (Hair et al., 1998). Bagozzi and Yi
(1988) recommended that CR should be equal to or greater than 0.60, and AVE should be equal to or greater than 0.50. As this threshold is widely accepted, it has been used in this thesis. Composite reliability and AVE of a scale is calculated by the following (Hair et al, 1998).

\[
\text{AVE} = \frac{\text{Summation of square factor loading}}{\text{Summation of square factor loading} - \text{Summation of error variances}}; \frac{\text{Square of summation factor loading}}{\text{Square of summation of factor loadings} + \text{Summation of error variances}}, \text{whereas measurement error is} (1-R \text{ square}).
\]

The standardized loadings are obtained directly from the program output; and \( \varepsilon_j \) is the measurement error for each indicator. The measurement error is 1.0 minus the square of the indicator’s standardized loading.

4.7 VALIDITY

Validity is the ability of the measuring instrument to measure what is to be measured” (Zikmund, 2000). According to McDaniel and Gates (2006) validity means the degree to which what the researcher was trying to measure was actually measured. Reliability alone is not sufficient to consider that an instrument is adequate (Churchill, 1979; Anderson and Gerbing, 1988; Hair et al., 1998). Therefore, validity is required to validate the constructs of this thesis. Hair et al. (2006) and Fornell and Larcker (1981) posit that assessing construct validity is a product of two validities; convergent and discriminant validities. Construct Validity is concerned with what the instrument is actually measuring (Churchill, 1995). Although measuring reliability develops ‘internally consistent’ sets of measurement items, it is not sufficient for construct validity (Nunnally, 1978). Construct validity was therefore examined in this thesis by analyzing both convergent validity and discriminant validity.

Convergent validity examines whether the measures of the same construct are correlated highly, and discriminant validity determines that the measures of a construct have not correlated too highly with other constructs.
4.7.1. Convergent Validity

Convergent validity related to the internal consistent validity between each construct items, i.e., high or low correlations (Fornell and Larckers, 1981). Convergent validity is defined as the degree to which multiple attempts to measure the same concept are in agreement (Bagozzi, 1994). Convergent validity was assessed based on the indicators estimated coefficients of each measurement scale used in this research (composite reliability), average variance extracted and Cronbach alpha. Convergent validity will be achieved if regression coefficients (factor loadings) of the measurement items are significant and substantial, i.e. >0.5 (Hair, 1998), as well as if the model receives a satisfactory level of fit. Composite reliability is a measure of the overall reliability of a set of heterogeneous but similar indicators, while, individual variable reliability can be tested using Cronbach alpha, the composite reliability is concerned with testing the reliability of a construct/ latent variable. The average variance extracted reflects the overall amount of variance in the manifest variables accounted for by the latent construct (Hair et al., 2006). Both the composite (construct) reliability and the average variance extracted have been calculated in this study by using the following two formulas (Hair et al., 2006).

4.7.2 Discriminant Validity

Discriminant Validity was conducted to make sure that each construct and its indicators, in the proposed model, differ from any other construct and its indicators. Discriminant validity is ‘the degree to which measures of different concepts are distinct’ (Bagozzi 1994, p. 20). Discriminant validity has been assessed in three ways. First, following the guidelines offered by Anderson and Gerbing (1988), discriminant validity is assessed by constraining the correlation between two constructs equal to 1 and then employing a chi square difference test on chi square values obtained from the constrained and unconstrained model. The discriminant validity is achieved when the chi square statistics for the unconstrained model is significantly lower than the constrained model (Bagozzi and Phillips, 1982). Second, discriminant validity is established if the AVE of each construct is greater than the square of the standardized correlations between constructs (Fornell and Larcker 1981). Third, According to Gaski (1984) discriminant validity is
achieved when the alpha coefficients for constructs are greater than their correlation coefficients.

4.8 STRUCTURAL MODEL

The structural model has been defined as “the portion of the model that specifies how the latent variables are related to each other” Evaluation of the model fit is the most essential event in SEM testing (Hair et al., 2006). The structural model aims to specify which latent constructs directly or indirectly influence the values of other latent constructs in the model (Byrne, 2001). There are two ways to think about the model fit. The first is to examine the adequacy (fit) of each individual parameters of the model, while the second is concerned with examining the goodness-of-fit (GOF) of the entire model (Schumacker and Lomax, 2010; Byrne, 2010). The discussion of examining the fit of each parameter is discussed below, while the criteria to evaluate the entire model fit are discussed later in this section.

4.8.1 Test Adequacy of Each Parameter Estimate

Schumacker and Lomax (2010) identified three key features of the adequacy of each parameter. One feature is whether a free parameter is significantly different from zero (Byrne, 2010). Once parameter estimates are attained, standard errors for each estimate are also obtained. A ratio of the estimated parameter to the standard error estimated can be calculated as a critical value (C.R.), which is assumed normally distributed; that is the critical ratio (C.R.) can be calculated through dividing parameter estimate by its standard error (Byrne, 2010; Schumacker and Lomax, 2010). As such, it functions as a z-statistic in testing that the estimate is statistically different from zero. Based on a probability level of .05, the test statistic must exceeds the value of ± 1.96 before the null hypothesis (that the estimate equals 0.0, in other words, no relationship exists) can be rejected (Byrne, 2010). The parameter estimate, standard error, and critical value are automatically provided in the AMOS output for a model.

A second feature is whether the sign (positive/negative) and the direction of the estimate are consistent with what is anticipated from the theoretical model (Schumacker and Lomax, 2010).
A third feature is that parameter estimates should be logical, that is, they should be within an anticipated range of values (e.g. no negative values obtained and correlations should not exceed the value of 1.00) (Byrne, 2010). Thus, all free parameters should be in the expected positive/negative direction, be statistically different from zero, and make practical sense (Schumacker and Lomax, 2010). The AMOS program provides also squared multiple correlations ($R^2$) for each single observed variable separately. These values shows how well each single observed variable serves as measure of the latent variables and range from 0 to 1 (Byrne, 2010). Squared multiple correlations are also specified for each endogenous variable separately. These values also range from 0 to 1 and serve as an indication of the strength of the structural relationships (Schumacker and Lomax, 2010).

4.8.2 Second, Test the Model as a Whole

The goodness-of-fit for the entire model (GOF) describes how well the hypothesized model reproduces the covariance matrix between the indictors’ items. In other words, the model is first specified (based on a theory) and then the sample data is utilized to test the model to determine the goodness-of-fit between the hypothesized model and the sample data (Byrne, 2006).

The model fit compares the theory to reality as characterized by the sample data. In other words the estimated covariance matrix ($\Sigma_k$) is mathematically compared to the actual observed covariance matrix ($S$) to supply an estimate of model fit, where the closer the values of these two matrices are to each other, the better the model fit (Hair et al., 2006). GOF measures for the whole model can be classified into three groups: absolute measures, incremental measures and parsimony measures (Arbuckle, 2008). In the following section, some basic elements in calculating GOF are reviewed, followed by an explanation of each category of GOF.

4.8.3 SEM Assumptions

Like any statistical method, a number of assumptions need to be met before conducting SEM. For example, SEM requires the sample size to be adequate, as covariance and
correlations are less stable when estimated from small sample sizes (Tabachnick and Fidell, 2001). While some authors believe that SEM could be used for sample sizes as small as 50 (Anderson and Gerbing, 1984), it has been generally accepted that 100 is the minimum sample size to ensure the appropriate use of maximum likelihood estimation (MLE) (Hair et al., 1998). However, Boomsma (1983) suggests that the estimation of SEM by using maximum likelihood methods can be used only when the sample size is at least 200. Bentler (1995) suggested that instead of thinking about number of participants per measured variable, it is worthwhile to thinking about how many subjects there are per estimated parameter. While there is no agreement among the scholars about sample size, Hair et al., (1998) considered a number of 200 to be ideal. The sample size of this thesis is 424 (bank employees) and 448 (bank customers), which is considered appropriate for using SEM.

4.8.4 Evaluating the Fit of the Model

In SEM, there are a series of goodness-of-fit indices, which identify whether the model fits the data or not. There are many indices provided by SEM, although there is no agreement among scholars as to which fit indices should be reported. For example, Anderson and Gerbing (1988) suggest that researcher might assess how well the specified model accounts for data with one or more overall goodness-of-fit indices. Kline (1998) recommends at least four such as GFI, NFI, or CFI, NNFI and SRMR. In order to reflect diverse criteria and provide the best overall picture of the model fit, Jaccard and Wan (1996), Bollen and Long (1993), Hair et al., (1998), and Holmes-Smith (2006) recommend the use of at least three fit indices by including one in each of the categories of model fit: absolute; incremental; and parsimonious (these are discussed below). The first category of absolute values includes chi-square ($\chi^2$), GFI, and RMSEA; the second category (incremental) includes AGFI, NFI, CFI, TLI; and the third category (parsimonious) includes $\chi^2$/df. These are described in more detail below.

4.8.4.1 Absolute fit indices

The chi-square ($\chi^2$) is considered the most fundamental measure of overall fit (Bollen, 1989). This is a test of whether the matrix of implied variance and covariance ($\Sigma$) is
significantly different to the matrix of empirical sample variance and covariance (S). It is calculated to determine the discrepancy between Σ and S. If the probability (P) is greater than .05, this indicates that the discrepancy between Σ and S is very small, meaning that the actual and predicted input matrices are not statistically different. Although this type of statistical index is the most important one to evaluate fit of the model, it has been criticized for being too sensitive to sample size (Fornell and Larcker, 1981; Marsh et al., 1988), especially in cases where sample size is over 200 (Bagozzi and Yi, 1988; Hair et al., 1998). Thus, marketing researchers do not solely use the value of chi-square to reject or accept their models (Bagozzi, 1981), but use it in conjunction with other indices to evaluate overall fit.

4.8.4.2 Goodness-of-Fit Index

The second measure of absolute fit index used within this thesis is the Goodness-of-Fit Index (GFI). The GFI measure indicates the relative amount of variance and covariance together explained by the model. The GFI value is calculated by comparing the discrepancy value for the model under test to the discrepancy value for a saturated version of the model which is counted as representing a 100% fit (or 1.0). However, this measure is not adjusted for degrees of freedom (Hair et al., 1995; Holmes-Smith, 1996), ranging from 0 (indicating a poor fit) to 1 (indicating a perfect fit), where a recommended level of acceptance is .90 (Hair et al., 1995).

4.8.4.3 Root Mean Square Error of Approximation

The third measure of absolute fit index used is Root Mean Square Error of Approximation (RMSEA). This measure assists in correcting the tendency of chi-square to reject specified models. It takes into account errors of approximation in the population, and relaxes the stringent requirement that the model holds exactly in the population. While Holmes-Smith et al., (2006) recommend that RMSEA should be less than 0.05; MacCallum and Browne (1993) suggest a value of up to 1.0 as reasonable. However, it has been found that a value ranging from .05 to .08 is commonly acceptable (Hair et al., 1998).
4.8.4.4 Incremental Fit Indices

The second category of indices includes incremental fit measures. These measures provide a comparison between the proposed model and the null model. Adjusted Goodness-of-Fit Index (AGFI), for instance, is one of the incremental indices, which has been found important, and is adopted in this thesis. This is because it takes into account adjustment for degrees of freedom, which GFI from the absolute fit indices category cannot do (Marsh et al., 1988; Holmes-Smith, 2006). The quantity 1-GFI is multiplied by the ratio of the model’s df divided by df for the base line model, the AGFI is 1 minus this result. Similar to GFI, this measure range from 0 (indicating a poor fit) to 1 (indicating a perfect fit), where a recommended level of acceptance is .90 (Hair et al., 1995). In addition to AGFI, Normed Fit Index (NFI) is one of the most popular incremental measures (Bentler, 1980, Byrne, 2001). NFI reflects the proportion to which the researchers’ model fit compared to the null model. For example, NFI = .50 means the researcher’s model improve fit by 50%. However, this index does not control for degrees of freedom (Bollen, 1989).

In order to overcome this shortcoming, Bentler (1990) has used it with the Comparative Fit Index (CFI). CFI compares the covariance matrix predicted by the model to the observed covariance matrix. Therefore, both of NFI and CFI are reported in this thesis. They range from 0 (poor fit) to 1 (perfect fit) having a commonly recommended level of 0.90 or greater (Hair et al., 1998). Another important incremental measure also used in this thesis is the Tucker-Lewis Index (TLI) (Tucker and Lewis, 1973). TLI is known as a non-normed fit index (Marsh et al., 1988). TLI combines a measure of parsimonious into a comparative index between the proposed or hypothesized and null models, resulting in values ranging from 0 (not fit at all) to 1 (perfect fit). Similar to NFI and CFI, the commonly recommended level is 0.90 or greater (Hair et al., 1995). It has been adopted in this thesis due to its ability to provide a nonbiased indication of model fit at all sample sizes (Finch and West, 1997).

4.8.4.5 Parsimonious Fit Indices

The third category of parsimonious fit indices tests the parsimony of the proposed model by evaluating the fit of the model to the number of estimated coefficient required to
achieve the level of fit (Hair et al., 2006). In this category, the normed chi-square ($\chi^2/df$) is the most popular parsimonious fit index used to evaluate the appropriateness of the model (Hair et al., 1998). In this measure, a range of acceptable values for the $\chi^2/df$ ratio have been suggested, ranging from less than 2.0 (Bollen, 1989; Hair et al., 2006; Tabachnick and Fidell, 2007), through less than 3.0 (Carmines and McIver, 1981), to more liberal limits of less than 5.0 (Wheaton et al., 1977). Since $\chi^2$ is the main component of this measure, $\chi^2/df$ is also sensitive to the sample size. Therefore, this thesis has used this measure as an indicator of overall fit (in conjunction with other measures), not as a basis for rejecting or accepting the model.