CHAPTER 4

RESEARCH DESIGN AND METHODOLOGY

4.1 INTRODUCTION

This chapter analyses the research design and methodologies used in the testing of the theoretical and conceptual models discussed in Chapter 3. It also presented the methods used for data collection and analysis were explained and the steps for the scale modification which included in-depth interviews and the reasons for modifying the scale were also explained in detail. This chapter enumerated the sampling method and the sample size determining procedure and a detailed explanation about the different data analysis techniques were outlined followed by a discussion on the validity and reliability of the scale items.

4.2 THE RESEARCH METHODOLOGY AND METHODS

For any research the most important and a difficult decision to make was whether to use quantitative or qualitative research methods or a mixed method approach. The differences between the qualitative and quantitative approaches were based on the judgment of different authors because both approaches may have included different methods (Hanson and Grimmer 2007). The fundamental reason for Quantitative research was to test an objective theory which made the researcher to collect numerical data and analyze it statistically. Therefore, quantitative research generally involved instruments that were used for collecting data and quantitative methods were
usually concerned with making generalizations about a population under investigation (Zikmund et al 2010).

In general there were two types of quantitative research methods used for collecting data:

1. Survey: It involved the collection of data by using a questionnaire to discover the opinions of a population, based on a sample of the population. Moreover, there were many techniques used to collect data which included interviews, telephone calls and observation (Tharenou et al 2007).

2. Experiment: Experiments were carried out to test the relationship between cause and effect relationship. This involved a comparison of the control group to the treatment group (Tharenou et al 2007). According to Tharenou et al (2007), experiments can be either true experiments or quasi-experimental design.

4.2.1 Research Methodology Used in Service Quality Models

In a review of 19 service quality models between the years of 1984-2003, nine studies used the quantitative methods, four used the qualitative method and two used mixed methods and four did not report any analysis (Seth et al 2005). In another study exploring the 30 different service quality scales, the quantitative method was the main approach for data collection. Moreover, the most widely known quality service model, which was the SERVQUAL, used a quantitative method as the main approach for data collection and the qualitative method was used for generating items for the service quality scale (Chen et al 2007).
Qualitative methods were found it difficult to find small differences that quantitative research might be able to find. Additionally, qualitative methods usually used small samples which made the findings not generalizable to the population. Based on the literature it was found that, the data collection method used in several service quality studies were quantitative and the survey technique was used.

4.2.2 Use of Likert Scale in Service Quality Measurement

In a study that reviewed 19 service quality models, it was found that 10 models have used quantitative methods which all used Likert scales (Seth et al 2005). According to Caro and Garcia (2008) the number of items on the service quality scale depended on the context of the research. In this research for collecting the data regarding the customers’ opinion on retail service quality, a questionnaire with a 5 point Likert scale was used.

4.2.3 Rationale for Research Design and Research Issues

In this section, the reason for the research methods adopted for the data collection, and the justification of how far the research design addresses the research questions were explained. The research issues and questions explained in Chapter 1 were presented in this section for better understanding of the research design.

**Research Issue One: Can Western Concept of Retail Service Quality Scale (RSQS) applicable to emerging country like India?**

As mentioned in the literature, Dabholkar et al (1996) developed and tested the retail service quality model in the United States of America (USA) retail setup and hence, the degree to which the applicability of retail service quality
model in a developing retail market like India should be tested and the first research issue focuses on this area.

In the research design of study one, the existing retail service quality scale constructs were verified for their applicability in Indian retail setting. The aim of study one was to find out the validity and reliability of retail service quality scale model in India.

**Research Issue two:** *What dimensions, or factors, best measure the retail service quality construct in India? What scale items should be used in developing measures of retail service quality?*

The research method in study two was shaped by the findings from research issue one. The study two results provided the development of a scale instrument and the research questionnaire to measure retail service quality in India. The purpose of the research questionnaire was to test the proposed factor structure of retail service quality and also to find out the construct reliability and validity of the scale instrument.

The combination of research issue one and two provided the path for answering the research questions. The framework for the two research issues was guided by the selection of appropriate research methods with the appropriate qualitative and quantitative methods to provide a better answer to the research questions.

**Research Issue One (Study One):** *Can Western Concept of Retail Service Quality Scale (RSQS) applicable to emerging country like India?*

The inherent research principles for this research was to find out the relationships between variables of retail practice and therefore, the need for using the exploratory and replication studies arose to assist in highlighting the
relationship between isolated variables (Malhotra et al 1996). The study one was an exploratory replication research to determine the degree to which RSQS developed for western markets fit into the emerging retail market like India. In the study one the degree to which a structural model of retail service quality had a similar factor structure to that of the existing service quality models (Dabholkar et al 1996; Finn and Lamb, 1991) were also examined.

The replication studies and its significance had been promoted in the marketing literature (Hubbard and Armstrong 1994), and there was a general consensus among the researchers that replication studies can be beneficial to the advancement of social science knowledge (Singh et al 2003). The purpose of replication study was not to merely examine the same issues in the same way (Tsang and Kwan 1999), but rather to examine the relevance of using the proven Western concept of RSQS to measure retail service quality construct in emerging retail market like India. The fundamental principle was to determine the relevance of an existing scale to another population, employing the same methods of measuring constructs and analyzing data and this type of research was referred as an empirical generalizations replication study (Tsang and Kwan 1999).

The purpose of research issue One was to use the cross cultural element into the study, but more focus was given for a replication study to empirically test the external validity of the Dabholkar et al (1996) retail service quality scale (RSQS) model and factor structure using the retail setup in India. It was supported by the literature mentioned in Chapter 2, that the previous replication studies of RSQS have questioned the validity of the factor structure as well as the stability of the scale items (Mehta et al 2000).

Survey questionnaire used in the pilot study included the scale items from the Dabholkar et al. RSQS model and some more additional items
specific to the Indian retail market were included based on the review of literature on retail service quality.

4.3 SAMPLING METHOD

A sample was taken from the target population being researched. If the sample was adequate it will have the same characteristics of the population (Zikmund 2003) and the findings were usually used to make conclusions about the population. Moreover, sampling techniques were methods that were used to select a sample from the population by reducing it to a more manageable size. According to De Leeuw et al (2008), sampling techniques were used when inferences are made about the target population.

The population for this research work was defined as like similar studies — comprising retail shoppers (Dabholkar et al 1996; Boshoff and Terblanche 1997; Mehta et al 2000; Kim and Jin 2002). The sample was collected from the Coimbatore and Erode Districts of TamilNadu state of India. Coimbatore and Erode were selected because of the presence of organized retail stores and was considered as a right place for understanding the customers’ opinion on RSQS.

A quota sampling procedure was used. The quota was fixed based on income, gender and age since these demographic characteristics were known to impact the perceptions of service quality (Gagliano and Hathcote 1994). The sample was divided equally among income, gender and age groups.
4.4 AREA OF STUDY

For the research purpose, Coimbatore and Erode of TamilNadu, India were selected.

Tamil Nadu is the Southern state of India and is among the most industrialized states in the country. The state is manufacturing base for a number of large and medium scale industries from diverse sectors such as automobile and auto-components, textile, cement, engineering, pharmaceuticals, etc. The state attracted cumulative FDI inflows of US$ 7.3 billion between April 2000 and April 2011 and is been amongst the highest FDI attracting states (Indian Brand Equity Foundation, Nov 2011 report). FDI inflows have helped achieve greater technological advancement. At current prices, the Gross State Domestic Product (GSDP) of Tamil Nadu was about US$ 97.8 billion in 2009-10. The GSDP increased at a CAGR (Compound Annual Growth Rate) of 14.9 per cent from 2004-05 to 2009-10.

The criteria used in selecting Coimbatore and Erode regions were given in the following table.

Table 4.1 Relevance of Coimbatore and Erode for the study

<table>
<thead>
<tr>
<th>S. No</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Growth of Organized retail stores</td>
</tr>
<tr>
<td>2</td>
<td>Increase in urban population</td>
</tr>
<tr>
<td>3</td>
<td>Change in Lifestyle</td>
</tr>
<tr>
<td>4</td>
<td>Growth of Literacy rate</td>
</tr>
<tr>
<td>5</td>
<td>Social relevance of research</td>
</tr>
</tbody>
</table>
The first criterion for selecting the location was based on the growth of retail market. The region had witnessed many organized retail chains expanded their operations and the statistical information about the demographic profile of Coimbatore and Erode districts further justified the selection of two districts. The growth in population in urban areas in the last 10 years had forced many organized retail stores expanding their retail presence in these locations.

The growth in literacy rate and subsequent changes in the life style were considered as one of the reasons for selecting the locations. Coimbatore and Erode were considered as the educational hub and as the life style changes from the more conservative single income family to a more liberal double income family people were ready to have a more matured shopping than the regular Kirana store shopping.

4.5 SAMPLE SIZE DETERMINATION

The Sample size calculation was based on many rules. However, the most effective way to determine a sample size would be to make use of power analysis by using G power3 software.

Effect size was a measure of the strength of the effects being studied in the research. According to Cohen (1978) the value of effect size is 0.02 for small effects, 0.15 for medium effects and 0.35 for large effects.

The significance level $\alpha$ was also known as type I error. It was the probability of rejecting a correct hypothesis. The commonly used value for $\alpha$ is 0.05. The power of a test (1-$\beta$) was the probability of not accepting wrong hypothesis. The usually accepted minimum value of (1-$\beta$) was 0.8. Any lower values will indicate a test that was not powerful enough. The number of predictors is the number of indicators contained in the most complex
construct in the proposed path model. These four constants are fed into the
software to analyze the deviation of $R$ squared value from zero. The resultant
output contains the required sample size satisfying the input conditions along
with the actual testing power of the sample size.

After a priori sample size calculation, the sample is again retested for power
using a post hoc calculation after the purification of the constructs to yield the
net power of the sample being studied.

APRIORI SAMPLE SIZE CALCULATION

**F tests** : ANOVA: Fixed Effects, Omnibus, one-way

**Analysis** : A priori: Compute required sample size

**Input Parameters:**

- Effect size $f = 0.10$
- $\alpha$ err prob $= 0.05$
- Power ($1 - \beta$ err prob) $= 0.60$
- Number of Groups $= 5$

**Output Parameters:**

- Noncentrality parameter $\lambda = 8.000000$
- Critical F $= 2.383132$
- Numerator df $= 4$
- Denominator df $= 795$
- Total sample size $= 800$
- Actual power $= 0.601778$
POST HOC SAMPLE SIZE CALCULATION

**F tests**: ANOVA: Fixed Effects, Omnibus, one-way

**Analysis**: Post hoc: Compute achieved power

**Input Parameters**:

Effect size $f$ = 0.10  

$\alpha$ err prob = 0.05  

Total sample size = 800  

Number of Groups = 5

**Output Parameters**:

Noncentrality parameter $\lambda$ = 8.000000  

Critical F = 2.383132  

Numerator df = 4  

Denominator df = 795  

Power (1 - $\beta$ err prob) = 0.601778
Table 4.2 Sample Size from Retail Stores at Erode and Coimbatore

<table>
<thead>
<tr>
<th>S. No</th>
<th>Name of the Place</th>
<th>Name of the Store</th>
<th>Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Coimbatore</td>
<td>Kannan Departmental Store</td>
<td>139</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Spencers</td>
<td>119</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nilgiris</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td></td>
<td>More Store</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Overall</td>
<td>397</td>
</tr>
<tr>
<td>2</td>
<td>Erode</td>
<td>Kannan Departmental Store</td>
<td>197</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Spencers</td>
<td>74</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nilgiris</td>
<td>74</td>
</tr>
<tr>
<td></td>
<td></td>
<td>More Store</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Overall</td>
<td>395</td>
</tr>
</tbody>
</table>

**Total Sample Size Collected = 792**

4.6 DATA COLLECTION METHOD

Data were collected by using a structured questionnaire and the questionnaire consisted of both demographic variables and service components. The questionnaire was self-administered at the store location. The reason for this data collection method was based on the theory that respondents were more attractive to the task of completing a questionnaire and provides more meaningful responses when they were contextualized in the environment that they were evaluating (Dabholkar et al 1996).

Data collection was more meaningful when the responses were taken from customers in the store after the shopping was completed (Dabholkar et al 1996; Boshoff and Terblanche 1997). Asking shoppers to complete a questionnaire in the shop provided them a chance to pay attention to the dimensions while answering the questionnaire as this also eliminated problems with customers trying to recall the shopping experience.
4.6.1 Data Collection Instrument

For using RSQS in the Indian retail context, two items out of 28 items in RSQS were found inapplicable. The research process used similar procedure used by Dabholkar et al (1996) for examining the face validity of the RSQS items. For the better use of RSQS, apart from the literature, two other sources were used to examine the face validity of the items. One independent expert in the field of service quality and one senior retail store manager of a leading retail chain were consulted before removing 2 items of RSQS. The two items removed by the experts as being inapplicable for South Indian retail are “telephonic interaction with customers” and “store own credit card”. All the other 26 items were found applicable to South Indian retail context by the experts.

The questionnaire was developed in English. Since the targeted retail customers of Coimbatore and Erode Districts speak, read or write in Tamil, it was necessary to translate the questionnaire into Tamil by a translator who was from the marketing background. To reduce the translation bias, the Tamil version of the questionnaire was again translated into English by another expert. The final instrument consisted of 26 items and additional items to assess the convergent and discriminant validity of the retail service quality questionnaire. These items are based on the study by Boshoff and Terblanche (1997). All items were measured using five point Likert scale’ from ‘1-Strongly Disagree’ to ‘5-Strongly Agree’. 
4.7 DATA ANALYSIS

Data analysis occurs after the data has been collected (Field, 2009) to make sense of the study and reach certain findings. The different techniques used for data analysis by the researcher in this study were explained in detail.

4.7.1 Correlation Analysis

Correlation analysis is a measure of the size and direction of the association between variables. It showed the strength of the association between two continuous variables. Correlation analysis was carried out to verify the discriminant validity. A correlation analysis was conducted to all the dimensions of RSQS which indicate that different constructs were comparatively distinctive from each other and the unidimensionality of the scale items were also assessed based on the correlation values between the items which indicated the relatedness of the questions asked in the dimensions. The correlation coefficient values range from -1 to 1; the strength of the association depends on the absolute value of the correlation coefficient and the direction of the correlation depends on the sign of the correlation (Burns and Bush 2010). According to Burns and Bush (2010) the table below was a rule of thumb about the strength of the association.

<table>
<thead>
<tr>
<th>Range</th>
<th>Strength of association</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.81-1.00</td>
<td>Strong</td>
</tr>
<tr>
<td>0.61-0.80</td>
<td>Moderate</td>
</tr>
<tr>
<td>0.41-0.60</td>
<td>Weak</td>
</tr>
<tr>
<td>0.21-0.40</td>
<td>Very weak</td>
</tr>
<tr>
<td>0-0.20</td>
<td>None</td>
</tr>
</tbody>
</table>
4.7.2 Confirmatory Factor Analysis (CFA)

Factor analysis is a common statistical method used to find a small set of unobserved variables (also called latent variables, or factors) which can account for the covariance among a larger set of observed variables (also called manifest variables).

Confirmatory factor analysis (CFA) is a statistical technique used to verify the factor structure of a set of observed variables. CFA is a special case of the structural equation model (SEM), and allows the researcher to test the hypothesis that a relationship between observed variables and their underlying latent constructs exists. CFA allows the researcher to test the hypothesis that a relationship between the observed variables and their underlying latent construct(s) exists. The researcher uses knowledge of the theory, empirical research, or both, postulates the relationship pattern a priori and then tests the hypothesis statistically.

Each of the five Retail Service Quality Scale (RSQS) dimensions was reviewed for their unidimensionality by using the Confirmatory Factor Analysis (CFA). Confirmatory factor analysis was conducted to assess the factor structure of the RSQS scale and the four component models were subjected to confirmatory factor analysis (CFA) examining the associations between overall service quality, the dimensions, and sub-dimensions.

4.7.3 Regression analysis

Regression analysis was used to find the relationship between one dependent variable and one or more independent variables and was popular in many research areas (Hair et al 1998). Regression was another way to determine the association between variables; this was similar to bivariate correlation as it assumes that there was a link between the dependent and
Regression analysis was called simple regression analysis when there was only one independent variable and was called multiple regression when there was more than one independent variable (Robson et al. 2008). Multiple regression analysis was more complex than correlation and was used to find the ability of a set of independent variables in predicting the dependent variable.

According to Zikmund et al. (2010) simple regression analysis and correlations were mathematically the same in many respects; however the correlation was an interdependence technique and regression was a dependence technique. The interdependence technique does not differentiate between dependent and independent variables; whereas the dependence technique distinguishes between dependent and independent variables.

The R square ranges from 0-1 and this showed how much of the dependent variable was explained by the independent variables. The higher the R square the stronger the association between the dependent variable and the independent variable.

To perform multiple regression analysis, the following assumptions must be adhered to:

- Sample size: the sample size has to be large enough for the results of the regression analysis to be meaningful. The sample size must be \(N \geq 50 + 8 \times m\) for testing multiple correlations, where \(m\) was the number of independent variables.
- Multicollinearity: Multicollinearity should not exist because Multicollinearity problems weaken the regression model. Multicollinearity exists when there is a strong relationship between the independent variables. The independent variables must not correlate highly with each other and \(r\) value is
preferred to be below 0.7 and an r value of above 0.9 indicates strong correlation which weakens the regression model. To perform the regression analysis, there should also be some correlation between the independent variables and the dependent variable and it should not fall below 0.3. To test Multicollinearity, the tolerance level must be greater than .10 and the VIF must be less than 10.

A study that measured retail service quality by using the RSQS in discount stores in Korea and the US used regression analysis to find the relationship between service quality, behavioral intentions and satisfaction (Kim and Jin 2001). In a study by Mehta et al (2000) regression analysis was used to find the significance of the RSQS dimensions in explaining the variance in the overall perception of the retail environment. Akbaba (2006) used regression analysis to test the relationship between the overall service quality against the service quality dimensions.

A stepwise regression analysis was performed to test the relationship between the dependent variable, the retail service quality, on the independent variables, Physical Aspects, Reliability, Personal Interaction, Problem Solving, and Policy.

4.7.4 Multivariate Analysis of Variance (MANOVA)

A MANOVA or multivariate analysis of variance is a way to test the hypothesis that one or more independent variables, or factors, have an effect on a set of two or more dependent variables. MANOVA allows the researcher to test hypotheses regarding the effect of one or more independent variables on two or more dependent variables. The effect of Age group, Educational Qualification and occupation and combined effect of demographic variables on different parameters of retail service quality across two different places.
4.8 THE VALIDITY AND RELIABILITY OF THE SCALE

Reliability and validity of the service quality scale was important for the measurement of service quality and for obtaining meaningful results. Furthermore, reliability was more important when the questionnaire adopted a Likert scale because there were many variables testing the concept. A questionnaire was considered reliable if it gave similar results when repeated (Hair 2006).

4.8.1 Reliability

Reliability was an indication of how consistent the findings were based on the method of data collection and analysis (Saunders et al 2007). The most common method for testing the internal consistency of a scale for reliability was the Cronbach alpha coefficient (Hair et al 1998).

The Cronbach alpha coefficient range were from 0 to 1 with a minimum of 0.6 while other studies suggest that anything above 0.7 indicate high levels of internal reliability (Hair et al 1998). They had stated that reliability was a test of the amount of consistency between the different measurements of an individual’s response, with the purpose of making sure that these responses were consistent and similar over a period of time and across situations. Furthermore, if the same test was conducted under the same circumstances, then the researcher should find a similar result which was called the test-retest method (Zikmund and Babin 2010a).

Therefore, if the scales have a high correlation, the test was reliable (Zikmund 2003). For the purpose of this research, the Cronbach alpha coefficient, the most common method for testing reliability, and 0.7 was used as the minimal accepted level. Many studies have used reliability to test their modified service quality scale that ranged from 0.6 to 0.96 (Dabholkar et al
Validity was concerned with the test capable of finding what it was designed for, which was not as simple as it seems (Hair 2006). Normally, researchers refer to three types of validity for the testing of a scale (Zikmund and Babin 2010b).

1. Face validity or content validity: It refers to an agreement between experts that the scale measures what it was intended to and seems to be a good reflection of the scale (Zikmund and Babin 2010b). To verify face validity, the researcher reviewed the literature to find similar scales used by experts in the field and where possible, consults with the experts (Parasuraman et al 2005).

2. Criterion validity: It refers to checking the scale to ascertain whether it can give accurate predictions, usually of the dependent variable. Moreover, it could refer to obtaining similar results using different measurements. Criterion validity could be classified as either predictive validity or concurrent validity. The classification depended on the time in which the scale and criterion validity were correlated. Concurrent validity was verified when the new measure and the validity were carried out at the same time and the measure turns out to be valid. Predictive validity was checked when the measure predicts a future result (Sureshchandar et al 2002, Jabnoun and Khalifa 2005 and Parasuraman, Zeithaml and Malhotra 2005).
3. Construct validity: It refers to the analysis of the data supporting a part of the study's hypothesis or the scale answering some of the research questions. Construct validity included convergent validity and discriminant validity. To have convergent validity, the scores for the category 'excellent' must be higher than the category 'very good' for each of the dimensions (Parasuraman et al 1988; Aldlaigan and Buttle 2002). The scores for the category 'good' must also be higher than the scores for the category 'poor' and so on.

Discriminant validity showed that the measure was unique in some way. Studies that have measured construct validity include Parasuraman et al (2005); and Caro and Garcia (2007). Using a scale outside its cultural context runs the risk of the instrument not measuring what was intended to measure. Therefore, it was crucial to test the instrument for construct validity before using it in a different country (Cui Lewis and Park 2003).

In this research, the researcher referred to the three types of validity tests while evaluating the validity of the new scale.

4.9 SCALE MODIFICATION

Evidence from past research suggested that using a generic scale to measure service quality across industries was not suitable without modification (Akbaba 2006; Caro and Garcia 2008). Therefore the more specific the measure was, the more valuable the potential information could be (Karatepe et al 2005). It was common for scales to be modified while measuring service quality, and this had been done by many researchers who felt the need for industry specific measures (Karatepe et al 2005; Chowdhary and Prakash 2007) and culture specific measures (Cui Lewis and Park 2003; Karatepe et al 2005; Prayag 2007). To customize a scale or adapt questions
that have already been used in other studies was more efficient than using questions that have not been tested.

The SERVQUAL scale extended across many industries and cultures and for this reason it became a weaker instrument that was used as a base for developing new service quality scales in different cultures (Gaur and Agrawal 2006). The SERVQUAL and the RSQS are two scales that were most frequently used to measure retail service quality. However, they were not suitable for other cultures without modifications (Gaur and Agrawal 2006).

A study conducted in Singapore set out to discover what the most appropriate scale was, RSQS or SERPERF, which only measured the perception level of service, in supermarkets and electronic retailers. They found that RSQS was superior in an environment where there are more goods and less services e.g. supermarkets, while SERVPERF was more appropriate for environments where services were more dominant than goods e.g. electronic retailers (Mehta et al 2000).

This chapter provided an overview of the research methodology used in this study. A survey was used as part of the quantitative approach. The original RSQS scale was modified to be more context specific and culturally sensitive. The researcher followed the steps recommended by Churchill (1979) while modifying the scale, by reviewing the literature to find the most suitable scale, then generating items to be used for the new scale was done based on the interviews. Data were collected by distributing a total of 850 questionnaires and 792 were found to be usable based on the sample size calculation. The data was subjected to reliability and validity tests to make sure it was appropriate for the main study. The data which was collected was then administered to partial disaggregation techniques for the Confirmatory Factor analysis. The findings from this chapter are carried forward to the next chapter for the validation of conceptual models developed in Chapter 3.