Chapter 4

Research Design and Methodology

Introduction

In the previous chapters, discussion is focused on the origin of this dissertation, the dimensions and related variables of SERVTRUST, the conceptual model and the hypotheses of the study. This chapter will present the research design and methodology which will be employed to obtain and analyze the data. This chapter will also provide the sampling plan, survey questionnaire, and data collection procedure. As mentioned in earlier chapters, this study is primarily based on a scale development procedure suggested by Churchill (1979), Parasuraman et al. (1988), Hinkin (1995), DeVellis (2003), and Fornaciari et al. (2005). Following this, the procedure covers all necessary steps to specify the domain of construct, items generation, purifying measurement, reliability testing, and validity testing.
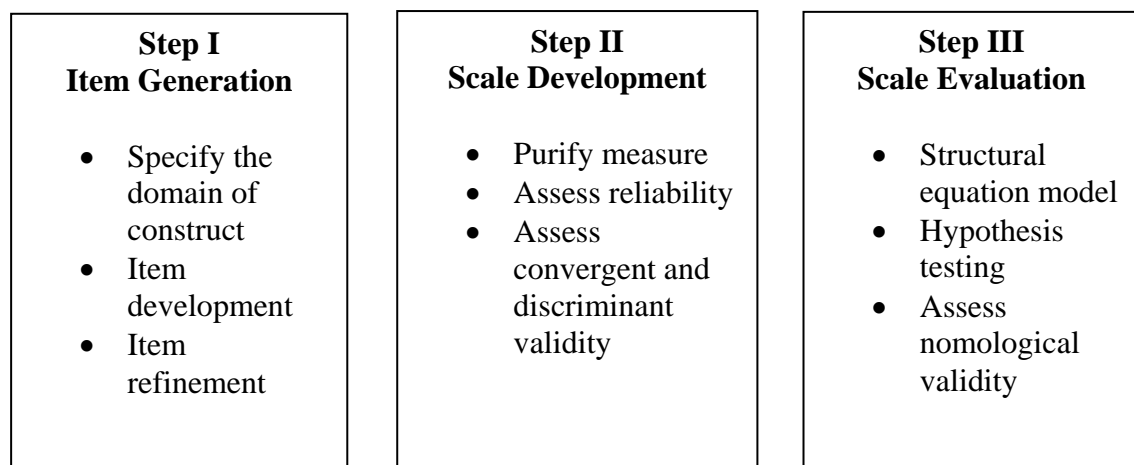
Research Design

The study is conclusive research which is designed as descriptive research (Malhotra, 2004) in order to identify and describe a consumer's trust in service providers within the area of business-to-consumer service in Thailand in particular. This research design is guided by the hypotheses shown in chapter 3. Descriptive research is widely used in marketing research because it provides an accurate snapshot of some environmental aspects of marketing (Aaker, D. A., Kumar, and Day, 2007). Survey research was employed because it is the most favorable methodology in education research (Abe, 2004). The data collection was a cross-sectional design, which used self-administered questionnaires to survey Thai consumers. This self-administered technique would reduce the cost of interviews, and also prove useful in the case of having no open-ended questions, allowing the respondent to interpret complex questions (Aaker, D. A. et al., 2007).

The marketing gurus who are the experts at scale development such as Churchill and Parasuraman, suggest procedures for developing better measures. The most cited author, Churchill (1979), suggests eight step for developing better measures. He starts from specifying the domain of a construct using a literature review. Then, the researcher can use that, experience survey, or focus group to generate a sample of items. Next, data collection is needed in order to purify a measure using factor analysis. Data collection should be done again in order to assess reliability and validity. Finally, the researcher should develop norms of measurement by summarizing the distribution of scores. In addition, Parasuraman et al. (1988) suggested 11 steps in developing a service quality scale. However, their suggestion covered Churchill's paradigm but more specific to the service quality context.

Because of its complexity, this study was categorized into three main distinct sections: item generation, scale development, and scale evaluation (Hinkin, 1995; Fornaciari et al., 2005). This procedure covers all suggested steps in scale development study by Churchill (1979), Parasuraman et al. (1988), Devellis (2003), Tian et al. (2001), and Malhotra (1981). However, the researcher modified a detail of each step for more understanding. Figure 4.1 shows the three sections together with the required tasks of each section.

Figure 4.1

Steps Employed in Developing the SERVTRUST Scale

Step I Item Generation	**Step II Scale Development**	**Step III Scale Evaluation**
• Specify the domain of construct • Item development • Item refinement	• Purify measure • Assess reliability • Assess convergent and discriminant validity	• Structural equation model • Hypothesis testing • Assess nomological validity

From Figure 4.1, Step I shows item generation. This stage consists of three important tasks (specify the domain of construct, item development, and item refinement) which must be finished before moving to the next stages. The domains or dimensions together with items were generated from a literature review, and were discussed with a person in marketing research. Then the questionnaire was launched to collect the data from Thai consumers.

Step II is data analysis and test of scale reliability and validity. The data was analyzed by using Exploratory Factor Analysis, EFA. Consequently, SERVTRUST, which is a scale to measure trust of Thai consumers, was presented. Internal consistency, which is concerned with the homogeneity of the items within a scale (DeVellis, 2003), was examined. Then, convergent and discriminant validity was assessed based on Confirmatory Factor Analysis, CFA. As a rule of thumb suggested by Hair et al. (1998), the scale will have a convergent validity if the standardized loading is greater than 0.5, variance extracted is higher than 0.5, and construct reliability is higher than 0.7. For discriminant validity, they suggested that a variance extracted estimate for two factors has to be higher than the square of their correlation.

Finally, Step III is the scale evaluation. Structural equation modeling was employed to test the relationship among SERVTRUST and related variables. Then, the hypothesized relationships were tested. Finally, nomological validity was examined to assess the degree to which the measurement behaved as expected (Churchill, 1979). If the structural equation model had a good fit and the hypotheses were as expected, nomological validity will be concluded.

Sampling Plan

Sampling Frame

This study targets Thai consumers and a sampling frame must include males and females who are at least 20 years old and have used a service related to health care and banking service providers. A quota sampling, which is a non-probability judgmental sampling technique (Joppe, 2001; Aaker, D. A. et al., 2007),

was used to distribute samples by gender, and age as shown in Table 4.1. This sampling technique eliminates gross biases (Aaker, D. A. et al., 2007) and ensures that the sample is representative of the population in the same proportion (White, 1998; Churchill and Iacobucci, 2005). Therefore, the result of the study can be partially inferred to the population (Aaker, D. A. et al., 2007); refer Table 4.1.

Table 4.1

A Sample Distribution

Age	Gender		Total
	Male (Percentage)	Female (Percentage)	(Percentage)
20-29	12.25	12.25	24.50
30-39	12.75	13.00	25.75
40-49	10.50	11.25	21.75
50-59	6.50	7.00	13.50
>60	6.50	8.00	14.50
Total	48.50	51.50	100.00

Source: Adapted from National Statistic Organization (2005)

Many researchers have tried to categorize target service industries by using different dimensions (Lovelock, 1983; Stell and Donoho, 1996; Michael, C., Diane, and Darryl, 2000; Patterson and Smith, 2001a, 2001b; Lovelock and Wirtz, 2004). However, this study categorized the service sectors by the degree of interpersonal contact because this is proposed as an important dimension which impacts on benefits received (Patterson and Smith, 2001a, 2001b). In order to represent a broad cross-section of services (Parasuraman et al., 1988), this study is conducted within two service industries, which are a health care (high contact), and a banking service (low contact) provider. These two service industries are involved with people in everyday life, not only in Thailand, but also in other countries. Thus, it is

easier for future research to test the equivalence of measurement model in different countries. Moreover, these two services were employed in the study of Patterson and Smith (2001a) about the relationship benefits in service industries in Thailand. They indicated that this typology is "the only one that is empirically based in a domestic services context" (p. 431). However, medium contact service was not included here because it can be easily transformed into low contact service, as internet and communication technology have changed the way of customer engagement with service providers (Lovelock and Wirtz, 2004).

Sample Size

Because of the limitation of factor analysis which is used to analyze the dimensions of trust in Step II, the minimum sample size for data collection was suggested to be a one-to-five ratio (Grimm and Yarnold, 1998; Hair et al., 1998). In contrast, Hair et al. (1998) suggested that a sample size of 100 would be suitable. Although only a small sample size is required when degree of freedom is large, a moderate sample size is needed for an adequate power analysis (MacCallum, Browne, and Sugawara, 1996), and to avoid an error which can occur during the process (Hair et al., 1998). However, in order to decrease the chances of over-fitting the data and to obtain a fully proper calculation, a sample size of one-to-three ratio, or approximately 400, was favorable for this study. This is supported by Marsh et al. (1998), who suggested that the sample size should be large in case of a moderate number (3-4) of indicators per factor.

Survey Questionnaire

Although questionnaire design is an important part of every dissertation (Simmonds-Moore, 2006), it is only briefly discussed here, as this study focuses on scale development. As such, item generation, which is mostly provided here, will be presented again, in detail in chapter 5.

This study used a self-administered questionnaire. Because of the length of this questionnaire, a self-administered design allows the respondent to answer in

their own time (Hair, Robert, and David, 2006b). The questionnaire first featured the Thammasat University logo to reassure the respondent that this was only for educational purposes. Then, the respondent was screened by asking whether he/she has used their main health care or banking service provider within the last three to six months. The screening questions ensured that the respondent still recognized his/her last service encounter.

The rest of the questionnaire focused on the development of a scale to measure consumer's trust. It contained the 13 dimensions of trust (benevolence, confidentiality, contractual, credibility, experience, expertise, friendship, information sharing, integrity, power, reliability, signal, and timeliness) together with 66 items, with each dimension measured by a rating scale. Respondents were presented statements that apply to each service industry and they were asked to respond to these with the choices: a lot, somewhat, not at all, or don't know. The "don't know" category was provided in case the respondent had insufficient experience to make a meaningful attitude judgment (Aaker, D. A. et al., 2007). However, the placement of items were rotated to avoid an order effect (Patterson and Smith, 2001a; Aaker, D. A. et al., 2007). Then, this questionnaire was pre-tested with the 30 respondents. This revealed that the provided rating scale was not reliable due to the respondent confusing judgment. From this analysis, it was suggested that the scale must be changed to the following three categories: applied a lot; applied a little; and don't know. Next, a new version of the questionnaire was developed and distributed to 500 respondents. Unfortunately, the result showed that there was low or no correlation among the items because only 45 of 500 respondents did not answer "don't know". As a result, this version of questionnaire seemed to be unusable and another revision was needed. The previous problems of questionnaire design will be presented clearly again in chapter 6.

A third version of the questionnaire was developed. From the previous two trials, each item was re-considered and some new items were added to increase the face validity of the construct. The revised version was divided into five main sections: screening, SERVTRUST, antecedents, consequences, and demographic information. As usual, the screening section included the introduction to the questionnaire and some other screening questions in order to obtain a respondent who

has used a health care and banking service from his/her main service provider. A valid respondent had used a health care service within the previous six months and a banking service within the previous three months. This would ensure that the respondent still recognized his/her last health and banking service experience.

The SERVTRUST, antecedents, and consequences sections were then developed. These three sections contained the operational definitions and measurement items of 19 constructs which were originally based on the conceptual definitions and items established in the western literature. These 19 constructs included 13 dimensions of trust (benevolence, confidentiality, contractual, credibility, experience, expertise, friendship, information sharing, integrity, power, reliability, signal, and timeliness), four antecedent latent variables (communication, problem solving, quality of the service, and satisfaction), and two consequent latent variables, (commitment and loyalty). After a review by marketing academic, Professor George P. Moschis, from the Georgia State University, the items pool contained 108 items which were 75 of SERVTRUST, 21 of antecedents and 12 of consequences. A seven-point Likert scale was employed because it is easier for respondents to answer than the previous rating scale (Abe, 2004). It is also the best for self-administered research design (Hair et al., 2006b). The respondent was asked to show their degree of agreement with the statements by using such categories markers as: strongly disagree; disagree; somewhat disagree; neither agree nor disagree; somewhat agree; agree; and strongly agree. Finally, demographic information questions were developed. In this section, closed-ended questions were used in combination with open-ended questions.

However, before administering the questionnaire, it was carefully translated into Thai which is the native language of the respondents. In order to avoid translation errors, the questionnaire was translated back to English, and compared with the original (English version). This procedure was suggested by Behling and Kenneth (2000). Three professional translators, who are bilingual, were hired to complete this task. The first one is a lecturer at Kasesart University, who has a part-time job as an English tutor at a private school. He also has a Masters degree in Business Administration (marketing), so he obviously understands the technical terms. He was responsible for the translation from English into Thai. The second translator is a lecturer at Prince of Songkla University. He has a very high skill in

English as an AFS exchange student, international program student and he is a Fulbright scholar. He was responsible for the translation from the previously translated Thai back into English. Moreover, the researcher discussed the questionnaire with a third translator, who is an English lecturer at Prince of Songkla University and a Ph.D. candidate in English Education at Chulalongkorn University. This was done to detect any differences in the translation of the complete questionnaire into Thai.

Data Collection

After the questionnaire was revised and addressed unfamiliar wording, ambiguous meanings, double-barreled questions, leading questions, confusing instructions (Aaker, D. A. et al., 2007), two research assistants were employed to collect data from the target respondents. These two research assistants were professional data collectors. They had been employed as field data collectors at a well-known research and private company such as INFOSEARCH Limited, Tayler Nelson Sofres Plc., Boon Rawd Brewery Company Limited, and AC Nielson (Thailand) Company Limited. However, they were re-trained by the author. Moreover, the author has a close control in order to avoid any mistakes which could occur during the data collection process. The research assistants were informed of the distribution quota needed, which was the same proportion of Thailand's population structure as reported by the National Statistic Organization of Thailand (2005).

Furthermore, additional criteria were used by the research assistants, the author, and screening questions in the questionnaire to determine eligibility. The respondents had to be at least 20 years old with Thai nationality, and to have had experience with the service providers of health care and banking services within six and three months respectively before the study. If they passed these screening criteria, they were asked to check whether the statements they were presented to them applied to each service industry. The contact sheet was developed and distributed to the research assistant for recording their job performance. From the contact sheet, 1,534 prospective respondents were approached. Specifically, 241 prospects refused to participate in the survey, while 893 prospects did not qualify. Finally, 400 eligible

respondents were used for the study. The target respondents were also asked for their contact address for calling-back. The data collection area covered the main area of Bangkok such as Thonburi Park, Siam Square, Pracha-uthid, Victory Monument, Ramkamhang, and Lumpini Park.

Data Analysis Approach

In Step I, the item pool was generated from the literature review. Then, a marketing professor, and six Ph.D. candidates in marketing at Thammasat University were asked to review all of the items in the item pool (Spake et al., 2003) because Churchill (1979) suggested that a person in marketing research should be involved in this stage of a scale development. The unclear items were discussed and adjusted while the weak items were deleted. Pre-testing the questionnaire was required in this step for checking the reliability of the questionnaire and also the understanding of the respondent. Then, the item pool was re-considered and some new items were added.

Step II then commenced after data collection had been completed from the samples. For purification of scales, the data was coded separately by the health care and banking service industries. The coded data were screened for missing values, inconsistent response, skewness, and kurtosis. It was tested using the assumptions of multivariate analysis as suggested by Hair et al. (2006a). The SPSS program was employed to complete this task and all descriptive statistics were presented simultaneously. Then, factor analysis, correlation and reliability were analyzed (Churchill, 1979). Exploratory factor analysis was analyzed also using SPSS statistical software. A basic statistic for determining the reliability suggested by Churchill (1979) was coefficient alpha, but test-retest reliability should not be used because of the respondents' memories. Then, confirmatory factor analysis was employed using LISREL statistical software (Joreskog and Sorbom, 1996) for convergent and discriminant validity testing (Gerbing and Anderson, 1988; Peter et al., 2002). However, the need is to calculate construct reliability as recommended. Construct reliability was calculated based on a formula suggested by Hair et al. (2006a). For convergent validity, they suggested that standardized loading estimates should be 0.5 or higher, and ideally 0.7 or higher, and that variance extracted should

be 0.5 or higher, and coefficient alpha for construct reliability should be 0.7 or greater (Nunnally, 1978). To provide evidence of discriminant validity for the total scale, they also recommended the rule of thumb that the variance extracted estimates for two factors should be more than the square of the correlation between the two factors.

Finally, in step III, Structural Equation Modeling (SEM), which is a preferable technique for testing the structural model and measurement model, was used based on the LISREL software package (Hair et al., 1998). In SEM analysis, three models were proposed: the first one is for health care service industry, the second one is for banking service industry, and the last one is for average scale. In order to support the assumptions of separate analyses, each model contained the same antecedent and consequent constructs/items. All error terms were not allowed to correlate with each other in order to confirm that there was not some semantic overlap. Then, the hypotheses were tested. T-value was employed to examine the significance level of the proposed relationships. Furthermore, the nomological validity was tested by relating measurements to a theoretical model that leads to further deductions, interpretations, and tests (Malhotra, 1981).

Conclusion

This chapter presented the research methodology including the research design, sampling, research tool, data collection, and the analysis methodology. For research design, it is a descriptive research using a survey. The study was divided into three steps: item generation; scale development; and scale evaluation. The target sample is Thai consumers who are at least 20 years old with relatively recent experience with health care and banking service providers. The questionnaire was developed based on the literature reviewed and it was examined by a marketing professor and doctorate students. A seven-point likert scale was used with the perceptual items. The questionnaire was translated from English into Thai and distributed to the 400 target respondents. Quota sampling was employed based on the proportion of Thailand's population structure. The data analysis approaches depended on the objective of each step. For step I, SPSS statistical software was employed to analyze the descriptive statistics and exploratory factor analysis. LISREL was used to

analyze the confirmatory factor analysis in step II and structural equation modeling was used in step III. Construct reliability, convergent validity, discriminant validity, nomological validity, and hypotheses were also examined and presented.

In the next chapter, Step I, Item generation will be discussed. The generating of the item pool and refinement of items will be presented in detail. The pre-testing results will also be explained.