CHAPTER 4 RESEARCH METHODOLOGY

4.1 Introduction

This chapter present the research methodology employed in the study. The chapter comprises of three major sections. First section includes research designs, design of the survey instrument and result of the sampling and respondents' profile. The second section describes results from testing on data screening. Finally, results from reliability, validity on survey data are presented.

4.2 Research designs

Research design is the blueprint developed to fulfil the research objectives by answering research questions. Quantitative approach was employed as research strategy in this study. A correlational study research method is undertaken in order to determine the relationship between the independent variables and the dependant variable. The instrument used in this study was a self-administered questionnaire that distributed to respondents in the form of survey and the data collected was primary data that used for analysis. The survey instrument was adapted from well established journals and previous literatures.

4.3 Design of the survey Instrument

A structured questionnaire was used as the research instrument in the survey. Questionnaire method was used as this method enable the researcher knows

accurately what is required and how to measure the variables of interest. In addition, questionnaire collection method offers quick response from respondents, easy to manage and well structures that enable respondent to answer the questions within defined alternatives. The questionnaire contained three major sections as listed below (refers to **Appendix I**):

Section A: Antecedents that influence the initial trust formation in Internet shopping
Section B: Ranking order of importance antecedents
Section C: Characteristics of respondents

Section A of the questionnaire was developed to measure and compare the importance of the antecedent that influence respondents' attitude in initial trust formation in Internet shopping that lead to behavioral action to form intention to purchase online. The questionnaire comprises of 40 items and 8 variables. The 40 measureable items were adaption of the previous research questionnaires done by Cheung and Lee (2001), Chen and Barnes (2007) and McKnight e*t al.,* (2002a). The 40 items of all variables were measured based on seven point Likert scale, in an increasing order, ranging from 1 "strongly disagree" to 7 "strongly agree". The seven scales were displayed in a numerical format to make the survey easy and convenient for the respondents.

In section B, a ranking table was provided to capture respondents' perceived importance of factor influencing their attitude in initial trust formation in Internet shopping. Respondents were asked to place the ranking order in the blank column, ranging from one to seven in a decreasing order of importance, where one

represented the most important factor, whereas seven represented as the least important factor. An example shows how to answer the ranking table correctly was prepared to provide respondent guidance in answering this section. Section C consisted of a series of questions addressing respondents' demographic information. Respondents were asked to provide information on their gender, age group, ethnic group, religion, highest academic qualification, technical background (from qualification), occupation and gross monthly income. These questions were composed to assist in understanding on the characteristics of the respondents.

4.4 Sample selection and sampling method

A survey was conducted in March and April, 2010. Convenience sampling method was employed in this study. Survey instrument was directly administered to the respondents. The target respondents of this study are general public from Klang Valley areas. Klang Valley is the most developed region in this country that Internet infrastructure is easily available, therefore the people staying in this region has relatively high exposure to e-commerce. As a result of that, for the purpose of this study, people staying in this region are a proper sample that enables to adequately represent the population of Malaysia. A convenience sample of respondent (N=700) was targeted. Each respondent was briefed about the survey to ensure questionnaire was completed correctly. Sample collection was carried out in three major townships in Klang Valley areas that with the highest number of Internet users in the country, namely Kuala Lumpur, Petaling Jaya and Klang. Fast food outlets, Cafes and shopping centres in Kuala Lumpur, Petaling Jaya, and Klang were selected to

conduct the survey. The samples were collected by employing convenience sampling in the selected locations.

4.5 Results of sampling

Table 4.1 summarised the results of the sampling. The table shows the results by presenting number and percentage of questionnaire distributed and returned in each selected location.

Table 4.1: Results of sampling

	Sampling location	Number of questionnaires distributed	Return of questionnaires (in number)	Return of questionnaires (in percentage)	Total respondents subjected to analysis
1	**Kuala Lumpur**				
	(A)Fast foods outlets	100	38	38%	36
	(B)Cafes	100	52	52%	47
	(C)Shopping centres	50	28	56%	24
2	**Petaling Jaya**				
	(A)Fast foods outlets	100	52	52%	52
	(B)Cafes	100	47	47%	46
	(C)Shopping centres	50	24	48%	19
3	**Klang**				
	(A)Fast foods outlets	80	23	29%	21
	(B)Cafes	80	38	48%	37
	(C)Shopping centres	40	24	60%	23
	Total	**700**	**326**	**47%**	**305**

In total, 700 questionnaires had been distributed. 326 questionnaires or 47% of the total questionnaires were returned. However, a total 13 of respondents were

excluded from analysis due to incomplete questionnaires. Eight respondents were excluded due to incorrect answer in their questionnaires. Ultimately, the total samples qualified for analysis is 305 samples or 43.60 per cent from the total sample collected.

4.6 Respondents' profile

Table 4.2 shows the respondents' characteristics in this study. Descriptive analysis was conducted and the result is presented in frequency and percentage. In this survey, female respondents and male respondents represent 54 per cents and 46 per cents of the samples collected respectively. Respondents are from the age group of 25 to 30 years old (32.50 per cents) and age group from 31 to 40 years old (31.10 per cents) of the total samples collected. This is followed by respondents from the age group 41 years old and above that represents 22.30 per cents, age group 19 to 24 years old (13.80 per cents) and age group 18 years old and below (0.30 per cents).

Respondents with degrees holders or professionally qualified are contributed 64.30 per cent in our survey. Respondents with diploma accounted for 17.40 per cents, whereas STPM and SPM or below accounted 8.80 per cents and 9.50 per cents respectively. The respondents with diploma qualifications or higher were asked to indicate the background of their qualification whether is technical relevant. The objective of this question is to enable analysis can be performed to investigate the relevance of technical background in this study. Respondents with education

without technical background contributed 69.80 per cent, the remaining 30.20 per cents were reported their qualifications are technical relevant.

Majority of respondents are come from officer/executive and managerial positions. This group of respondents accounted 54.80 per cents of the total respondents where officer/executive (38.40 per cents) and managerial (16.40 per cents). Professional and business owner are accounted for 13.40 per cents and 8.20 per cents respectively. Student respondents contributed 13.80 per cents, whereas retired respondents accounted for 9.80 per cents. Respondents with monthly income ranged from RM 2,001 to RM 4,000 contributed 35.40%. Another 22.00 per cents reported their monthly earning is between RM 4,001 to RM 6,000. The result indicated that majority of the respondents are middle income earner. Respondents reported their monthly earning between RM 8,001 to RM 10,000 (4.20 per cent) and RM 10,001 and above (4.90 per cent). Monthly income from RM 2,001 or less is accounted for 18.70 per cents.

Table 4.2: The demographical profiles of respondents (*N*=305)

		Frequency, *n*	Percentage, %	Cumulative, %
Gender	Male	139	46.00	46.00
	Female	166	54.00	100.00
Age Group	18 years and below	1	0.30	0.30
	19-24 years old	42	13.80	14.10
	25-30 years old	99	32.50	46.60
	31-40 years old	95	31.10	77.70
	41 years old and above	68	22.30	100.00
Ethnic Group	Malay	74	24.30	24.30
	Chinese	164	53.80	78.10
	Indian	52	17.00	95.10
	Others	15	4.90	100.00
Religion	Islam	81	27.00	27.00
	Buddhism	118	39.00	66.00
	Hinduism	35	11.00	77.00
	Christianity	55	18.00	95.00
	Others	16	5.00	100.00
Highest qualification	SPM or below	29	9.50	9.50
	STPM	27	8.80	18.30
	Diploma	53	17.40	35.70
	University degree/ professional qualifications	196	64.30	100.00
Technical Background	With technical background	92	30.20	30.20
	Without technical background	213	69.80	100.00
Occupation	Student	42	13.80	13.80
	Officer/Executive	117	38.40	52.20
	Managerial	50	16.40	68.60
	Professional	41	13.40	82.00
	Business owner	25	8.20	90.20
	Retired	30	9.80	100.00
Gross monthly income	RM 2,000 or less	57	18.70	18.70
	RM 2,001 to RM 4,000	108	35.40	54.10
	RM 4,001 to RM 6,000	67	22.00	76.10
	RM 6,001 to RM 8,000	45	14.80	90.90
	RM 8,001 to RM 10,000	13	4.20	95.10
	RM 10,001 and above	15	4.90	100.00

4.7 Descriptive Statistics of the Survey Data

4.7.1 Data preparation

Data preparation involves checking the data for accuracy, entering the data, and developing and documenting a database structure that integrates the various measures using standards statistical programs (*i.e.,* SPSS). The raw data were checked for data entry errors and all errors were corrected before analysis.

4.7.2 Characteristics of the variables scores

The starting point to understanding the nature of any variable is to characterize them in relation to normality, homoscedasticity, linearity, and multicollinearity and singularity. Understanding these characteristics of the data creates awareness of any assumption violations and the implications they may have for the estimation process or the interpretation of the results. The model developed in Chapter 2 contained 8 constructs. Each construct was described in detail in Chapter 5. The data collected in relation to each construct were examined for normality, homoscedasticity, linearity and multicollinearity and singularity.

4.7.3 Normality of Data

Normality refers to the shape of the data distribution for an individual metric variable (scale), and it correspondence to normal distribution (Pallant, 2005; Tabachnick ad Fidell, 2001). Normality is tested by examining the skewness and kurtosis of the

variables. According to Hair *et al*., (1998), skewness for the data distribution is within the range of -1 to +1, the data distribution is considered normal, similarly the kurtosis for a data distribution within the range of -1 to +1, the data distribution is considered normal.

An examination of the skewness and kurtosis value indicate all the variables listed are within the range of -1 to +1, therefore, the results showed that the data collected are normal, or near normal is this study. Similarly, all variables showed results within -1 to +1 in the kurtosis test that showed normality is assumed in the data collected.

4.7.4 Linearity

This study test linearity by running a simple regression analysis and examines all the independent constructs to the dependents constructs. Partial regression plots (refer Appendix IV) demonstrate linear association between the relationships of all the single independents constructs to the dependent constructs. The partial regression plots suggest linear relationships among the constructs.

4.7.5 Homoscedasticity

Homoscedasticity refers to assumption that dependent variable exhibit equal levels of variance across the range of independent variables(s). Examination of the scattplot showed points are randomly and evenly dispersed throughout the plot indicative of a situation in which the assumption of homoscedasticity has been met.

4.7.6 Multicollinearity and Singularity

With multicollinearity, the constructs are very highly correlated. With singularity, the constructs are redundant, *i.e.*, one of the constructs is a combination of two or more constructs. Multicollinearity and singularity was assessed using, tolerance index (TI), variance inflation factor (VIF) and correlation coefficients.

First, examine the TI and VIF by referring to **Table 5.1**, the tolerance values ranged from 0.382 to 0.768, well within the acceptable ranges of greater than 0.10, and VIF values ranged from 1.302 to 2.619, and are well within the acceptable ranges of less than 10. Second, the correlations between the dependent construct and independent constructs have most correlations around the preferable coefficient of 0.30 (Tabachnick and Fidell, 2001; Hair *et al*., 1998). **From Table 4.3**, the results from examination of correlations between measures of constructs indicates the correlations between each independent construct are below 0.68 which is below the recommended value of 0.70 suggested by Tabachnick and Fidell (1996). Therefore, suggesting the non-existence of multicollinearity assumption.

4.7.7 Examination of outliers

The presence of outliers was examined using histogram, normal Q-Q plot, the detrended normal Q-Q plot, and box plot. The histograms, normal Q-Q plot, the detrended normal Q-Q plot, and box plot indicate the presence of no data considered to be outliers.

4.8 Reliability Test and Correlation

4.8.1 Reliability Test

Cronbach's alpha coefficient was utilised to measure the internal consistency of the scales employed in this study. The Cronbach's alpha value for each variable is presented in **Table 4.3**. As Nunnally (1978) and DeVellis (2003) recommend a minimum level of 0.70, then scale of the construct can be considered high reliable. **Table 4.3** shows all the constructs revealed Cronbach's alpha values greater than 0.80, which is exceed the recommended value of 0.70. The Cronbach's alpha value for construct, Initial Trust in Internet shopping was calculated based on the 4 items retained.

4.8.2 Correlation

Pearson's correlation was used to explore the relationship between all the variables in this study. Correlations coefficients are able to provide numerical summary of the direction and strength of the linear relationship between all the variables. Pearson's correlation coefficients (*r*) can take on values from -1 to +1, and the sign out the front indicates the directions, positive correlation or negative correlation (Pallant , 2007). Cohen (1998) suggests the following guidelines to determine the strength of the relationship.

r = .10 to .29 or *r* = -.10 to -.29 small

r = .30 to .49 or *r* = -.30 to -.49 medium

r= .50 to 1.0 or *r* = -.50 to -1.0 large

Table 4.3 shows that all constructs are positive correlated to each others. Initial trust in Internet shopping (IT IS) was significantly correlated with Intention to Trust (IPO), r = .63, p (two-tailed) <.01. Perceived technical competence on Internet vendors (PTC) was significantly correlated with Initial trust in Internet Shopping (IT IS), r=.68, p (two-tailed) <.01. There was significant relationship between Perceived technical competence on Internet vendors (PTC) and Perceived Benevolence of Internet vendors (PB), r=.59, p (two-tailed) <.01. Similarly, third party recognition (TR) was also significantly correlated with perceived technical competence on Internet vendors (PTC) r=.57, p (two-tailed) <.01. Cultural environment (CE) was significantly correlated with Propensity to trust (PTT), r=.54, p (two-tailed) <.01.

Table 4.3: Cronbach's alpha value of constructs and correlations

Reliability Test			Correlations between measures of constructs (*N*=305)										
Items	Cronbach's alpha value		**PI**	**PC**	**PB**	**PTC**	**TP**	**LF**	**PTT**	**CE**	**EXP**	**ITIS**	**IPO**
3	.81	Perceived Integrity of Internet vendors (**PI)**	1.000										
3	.88	Perceived Competence of Internet vendors **(PC)**	.42	1.000									
3	.90	Perceived Benevolence of Internet vendors **(PB)**	.56	.45	1.000								
6	.88	Perceived Technical Competence on Internet vendors **(PTC)**	.51	.40	.59	1.000							
3	.89	Third Party Recognition **(TP)**	.38	.30	.44	.57	1.000						
3	.93	Legal Framework **(LF)**	.31	.23	.43	.47	.52	1.000					
4	.91	Propensity to Trust **(PTT)**	.28	.11	.28	.17	.194	.37	1.000				
3	.80	Cultural Environment **(CE)**	.24	.20	.29	.20	.24	.25	.54	1.000			
3	.93	Experience **(EXP)**	.28	.31	.21	.19	.19	.05	.15	.32	1.000		
4	.85	Initial Trust in Internet Shopping **(IT IS)**	.50	.34	.55	.68	.48	.42	.34	.37	.34	1.000	
3	.94	Intention to Purchase Online **(IPO)**	.45	.36	.50	.52	.33	.26	.17	.23	.44	.63	1.000

** Correlation is significant at the 0.01 level (*p*<.01) (2-tailed).

4.9 Principal Component Analysis

The aim of Principal Component Analysis (PCA) is to determine that the questions developed in the measurement instrument, *i.e.,* questionnaire are tapping the right concept and not something else (Sekaran, 2003). Validity test is employed by researchers to determine the wellness of an instrument used in measuring a particular concept that supposed to measure.

According to Pallant (2007), sample size and the strength of the relationship among the items or variables are the two main issues to be considered for a set of data is suitable for factor analysis. Tabachnick and Fidell (2007) suggested that a sample size of at least 300 is comfortable for factor analysis, however, smaller sample size is still acceptable if condition presents in high loading marker variable (above 0.80). In this study, 305 samples were qualified for the analysis, hence the sample size is considered sufficient for factor analysis. A correlation coefficient (loading level) greater than 0.30 is considered acceptance for factor analysis (Tabachnick and Fidell, 2007). An inspection of the correlation matrix revealed the presence of many coefficients of 0.3 and above.

Two statistical measures, namely Kaiser-Meyer-Olkin (KMO) and Bartlett's test of Sphericity are important in assessing the factorability of the data. KMO measures sampling adequacy, whereas Bartlett's test of Sphericity. Bartlett's test of Sphericity should be significant (p<0.05) in order for the factor analysis to be considered appropriate. The KMO index ranges from 0 to 1, with 0.60 is the minimum value for a good factor analysis (Pallant, 2007). In this study, individual construct was tested

independently. KMO value and Bartlett's test of Sphericity were first inspected to ensure data sets are suitable for factor analysis, then follow by determining internal validity, and number of components extracted by using Kaiser's criterion that have an eigenvalue of 1 or above.

Table 4.4: KMO and Bartlett's Test of Sphericity for variables

		Bartlett's Test of Sphericity		
	Kaiser-Meyer-Olkin Measure of Sampling Adequacy	**Approx. Chi-Square**	***df***	**Sig.**
Perceived Trusting Belief on Internet Vendors	0.83	1723.342	36.000	0.000
Perceived Technical Competency on Internet Vendors	0.84	970.754	15.0000	0.000
Perceived Organisational Compliance	0.81	1396.528	15.000	0.000
Propensity to trust	0.83	836.352	6.000	0.000
Cultural Environment	0.66	320.019	3.000	0.000
Experience	0.75	712.549	3.000	0.000
Initial Trust in Internet Shopping	0.72	646.305	15.000	0.000
Intention to Purchase online	0.75	824.114	3.000	0.000

Table 4.4 shows that the Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO) value for all data were above .60 or above, therefore all data sets are suitable for factors analysis. For example, the highest KMO value is .84 from perceived technical competence from Internet vendors, whereas the lowest KMO value reported is .66 from cultural environment. Bartlett's Test of Sphericity values from all constructs are significant as $p<0.05$, therefore factory analysis is appropriate in this

study. However, two observations made from the inspection of the eigenvalues from tables of Total Variance Explained and Component Matrix of two constructs that required further investigation, namely Perceived Technical Competency on Internet Vendors and Initial Trust on Internet shopping.

Perceived Technical Competency on Internet vendors have two dimensions according to the proposed conceptual framework, there are perceived security control and perceived privacy control. However, total variance explained showed only one component with eigenvalue above 1.0 and that component extracted explained a total of 63 per cents of the variance. The result shows Perceived Technical Competency on Internet vendors is a unidimensional construct rather than two dimensions as what proposed in the conceptual framework. As a result, the two dimensions of Perceived Technical Competency on Internet vendors were merged as one dimension for data analysis. Furthermore, Component Matrix for Initial trust in Internet shopping reported that two components are extracted. Based on the result from PCA, a moderated conceptual model is proposed in **Figure 4.1** shows perceived technical competency on Internet vendors is a unidimensional construct, and accordingly hypothesis $\mathbf{H_2}$ is maintained and tested whereas hypotheses $\mathbf{H_{2a}}$ and $\mathbf{H_{2b}}$ will not be tested.

On the other hand, Initial Trust in Internet shopping should be a unidimensional construct that represent the potential attitude formed of Internet shoppers after influence by those antecedents proposed in this study. However, PCA reported that two measurement items for initial trust in Internet shopping, (1) I can not rely on e-vendors to keep the promise that they make, and (2) Anyone trusting

Internet shopping is asking for trouble, did not provide any explanatory power to justify the multi dimensional of the attitude action in this study, therefore, the two measurement items will not be included in testing.

PCA was then re-performed for all constructs and **Table 4.5** presents the results of the analysis done on goodness of data, factor loading, KMO value, eigenvalue and variance explained of all constructs were presented. Perceived Technical Competency on Internet Vendors was tested as an unidimensional construct and reported eigenvalue of 3.782 and variance explained of 63.00 per cent. On the other hand, initial trust in Internet shopping showed only one component with eigenvalue reported 2.752 and that component extracted explained a total of 68.81 per cents of the variance. All constructs showed eigenvalues larger than 1 and KMO value above 0.60. The result shows the constructs validity is now acceptable.

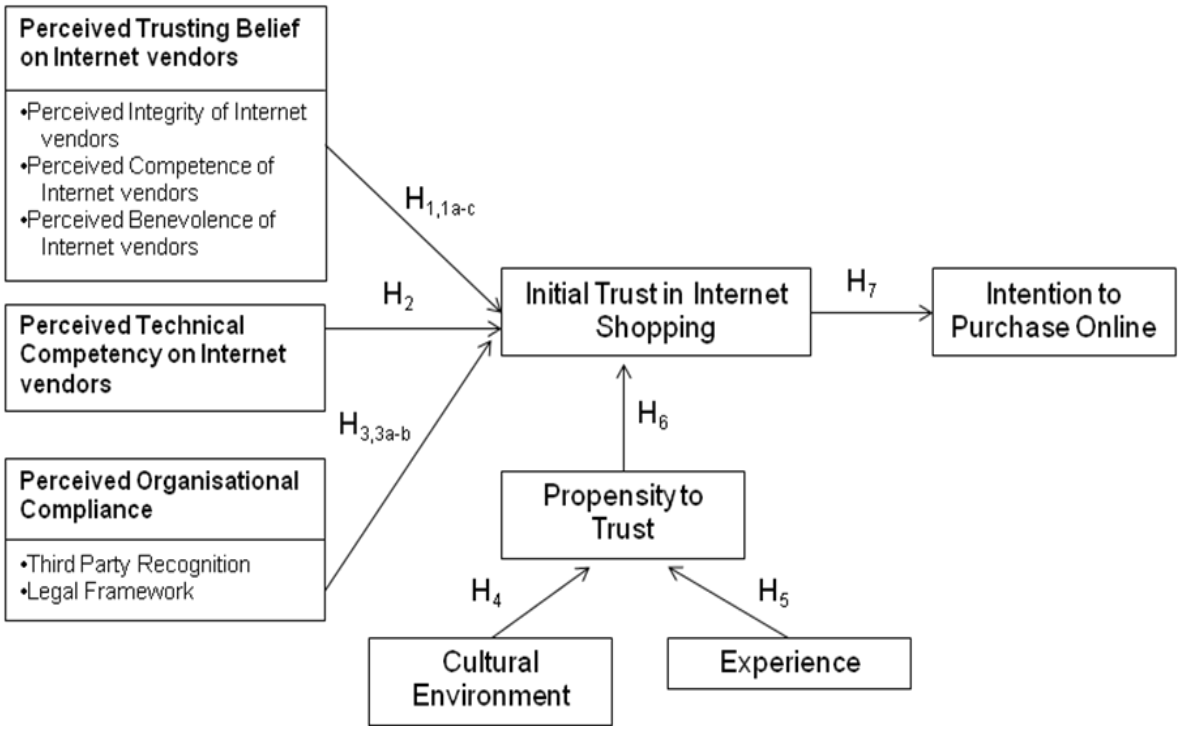

Figure 4.1: Moderated conceptual model

Table 4.5: Goodness of Data

	Constructs	Items	Factor Loading	KMO	Eigenvalue	Variance explained
1	Perceived Trusting Belief on Internet vendors					
1.1	-Perceived Integrity of Internet vendors	3	0.758-0.907	0.657	2.209	73.62%
1.2	-Perceived Competence of Internet vendors	3	0.859-0.938	0.688	2.407	80.25%
1.3	-Perceived Benevolence of Internet vendors	3	0.907-0.927	0.751	2.513	83.71%
2	Perceived Technical Competency on Internet vendors	6	0.669-0.844	0.842	3.782	63.04%
3	Perceived Organisational Compliance					
3.1	-Third Party Recognition	3	0.888-0.936	0.722	2.466	82.20%
3.2	- Legal Framework	3	0.924-0.956	0.741	2.624	87.48%
4	Propensity to trust	4	0.881-0.904	0.843	3.177	79.42%
5	Cultural environment	4	0.752-0.890	0.660	2.134	71.13%
6	Experience	3	0.917-0.946	0.753	2.609	86.97%
7	Initial Trust in Internet Shopping	4	0.800-0.861	0.783	2.752	68.81%
8	Intention to purchase online	3	0.916-0.957	0.746	2.661	88.69%

4.10 Chapter Summary

This chapter first discussed the research design, sample selection and sampling, result of sampling, followed by descriptive statistics of data testing and ended with tests of reliability, validity on survey data. A moderated conceptual was proposed based on the result from PCA. The following chapter presents data analysis and results generated from SPSS and Structural Equation Modeling (SEM) statistical techniques.