CHAPTER 3

RESEARCH METHODOLOGY

3.0 Introduction

This chapter discusses the design and methodology used in this research. Several aspects such as the research design, product and country selection, measurement of constructs, questionnaire design, sampling technique, data collection technique and data analysis technique are described in detail in this chapter. Furthermore, the hypotheses of the study are also presented in this chapter.

3.1 Research Design

According to Malhotra (2005), research design is a framework or blueprint for conducting the research project and it will specify the details of the procedures necessary for obtaining the information needed to structure and solve the research problems. Furthermore, Malhotra classified the research design into two broad categories, i.e., exploratory and conclusive research, where the objective of exploratory research is to provide insights into, and an understanding of, the problem confronting the researcher. Whereas, the conclusive research is designed to assist the decision maker in determining, evaluating and selecting the best course of action to take in a given situation.

According to Zikmund (2000), the research design is a master plan specifying the methods and procedures for collecting and analyzing the needed information and suggests four basic design techniques for the research, i.e., surveys, experiment, secondary data and observation. The research design is also intended to explain and discuss the relationship between the variables investigated, and, finally, it is designed to discuss and interpret the procedures of the study from developing the hypotheses to data analysis.

Survey research is widely used to determine specific characteristics of groups (Fraenkel and Wallen, 2003) and measure attitudes and opinions of groups towards certain issues (Ary, Jocobs, and Razaveih, 2002). Surveys are conducted through various methods including mail, telephone, and personal interviews based on the contents of the questionnaire, number of subjects, budget, time available and target response rates. Correlational research is performed to determine relationships among variables, whereas causal-comparative research is intended to detect the case for or the consequences of differences between groups of people (Fraenkel and Wallen, 2003). Correlational research is not different from casual-comparative research because both are based on the relationships among variables (Fraenkel and Wallen, 2003). They are effective to explore possible causes or results based on already existing differences among groups (Fraenkel and Wallen, 2003).

The present study was conducted using the survey approach with a multiple choice questionnaire requiring respondents to give fixed responses to the statements or questions asked. As a result, it will accomplish the objectives of the research and answer the issues raised and problem statements put forward through careful

analysis. The survey design will also help the researcher to achieve the objectives of the research and testing the hypothesized relationships.

The decision to use a survey research in the present study is also because it is the most common form of research design used in country of origin studies (e.g. Darling and Wood, 1990; Darling and Taylor, 1993; Jaffe and Martinez, 1995; and Balabanis and Diamantopoulos, 2004), consumer ethnocentrism studies (e.g. McLain and Sternquist, 1991; Razak, et al., 2002; and Klein at al., 2006), consumer animosity studies (e.g. Klein et al, 1998; Klein, 2002; and Hinck, 2005) and other consumer behaviour studies (Wang and Rao, 1995). Strengthening the decision to use a survey approach is that it is the most flexible means of obtaining data from respondents as well as providing an opportunity to examine causal relationships and performing the analyzing task of a complex and sophisticated analysis technique (Zikmund, 2000).

3.2 Country and Product Selection

One of the main focus in this study is to explore the effects of the animosity construct among Muslims towards the purchase of foreign products. Thus, how to select a foreign country as a producer is one of the major issues in this research. Consequently, it is very important to select the foreign country that might have an issue or problem with Muslims all over the world. Subjects' knowledge about countries plays an important role in participants' information processing and decision making. If there are no problems between consumers and foreign country producers, the result could be meaningless and the main objectives of the research might not be accomplished.

When the target population in the research is Muslims, it is important to make sure that the issues will directly give effect to the target group. In this case, it can be said and argued that the current relationship between Muslims and the US, as a whole, can influence all Muslims to express their dissatisfaction towards the US by using their purchasing power. As suggested by the animosity model of foreign products purchase, the animosity towards another country can have many sources such as military events or diplomatic disputes (Klein et al., 1998).

For example, the dissatisfaction can be seen by referring to the statement of one of the most prominent scholars in the Muslim world, Dr. Yousef Al-Qaradawi:

> *"Each riyal, dirham …etc. used to buy their [U.S., Israeli] goods eventually become a bullet fired at the hearts of a brother or a child in Palestine," he said. "For this reason, it is an obligation not to help them. To buy their goods is to support tyranny, oppression and aggression. Buying goods from them will strengthen them; our duty is to make them as weak as we can." "American goods, exactly like 'Israeli' goods, are forbidden. It is also forbidden to advertise these goods". Furthermore, Al-Qaradawi added "the time has come for the Islamic Ummah (people) to say "No" to America, "NO" to its companies, and "NO" to its goods, which swamp our markets"* (Waheed, 2002).

Another prominent Muslim scholar in Syria said that it is not permissible to purchase American products manufactured in the Arab and Islamic world as long as part of its profits goes to the mother American company. Dr. Mohammad Saeed Al-

Bouti, head of the Beliefs and Religions Department in Islamic Law (Shariaa) School, Damascus University said that:

> *"The American products which must be boycotted are those whose revenues go to the U.S. such as American cigarettes and restaurants. There are too many of these companies in our countries". He added "these products, which are manufactured in Arab and Islamic countries by franchise agreements given by American companies, must not be purchased, because part of its profit goes automatically to these companies. Boycotting American and Israeli products are compulsory and it is the Jihad [struggle] that every Muslim can carry out in face of the Israeli aggression and those who support it"* (Waheed, 2002).

From the perspective of the US official, Assistant Secretary of State for East Asian and Pacific Affairs, James A. Kelly, responding when he was asked a question on whether anti-American feelings among Muslims, which resulted from the US invasion of Iraq, would change with the US helping tsunami victims when the giant wave hit several countries including Indonesia and Malaysia, he said:

> *"The United States is not looking to change the anti-American sentiments of people in Indonesia and other Muslim countries by helping tsunami victims"* (The Star, 2005: 10).

In Malaysia, one campaign organized by the Muslim Consumer Association of Malaysia (PPIM), urging Muslim consumers not to consume and to boycott US

goods, in response to US interference in the internal affairs of Muslim countries (Malay Mail, 2002). The president of the Association of Muslim Restaurant Operators of Malaysia (PRESMA), Jamarulkhan Kadir said that at the beginning of the boycott campaign only the Coca-Cola brand was involved. He added, in future they might consider the boycotting of other US food products. They have not decided what brands but the boycott will be extended to a few products gradually. He reiterated that the boycott was a symbolic protest against brands that symbolise US hegemony (Malay Mail, 2002).

In a later event, PPIM again launched the campaign of boycotting US made products in August, 2006. This time, they called on Malaysians to stop buying products from three high-profile US companies, Starbucks, Coca Cola and Colgate-Palmolive. The boycott aims to protest US support of Israel's military actions against Palestine and Hezbollah in southern Lebanon (Islam Online, 2006). Azmin Ibrahim, the group's organizing committee member, said that:

> *"To buy and trade with US goods is to actively support the illegal Israeli occupation in Palestine and the war in Lebanon, Israel and the US has consciously destroyed Palestine and Lebanon and undermined peace in the region" (Islam Online, 2006).*

Ben-Meir (2005) argues that a number of reasons can provide the "rationale" for this negative kind of behaviour. To understand the depth of hatred towards the US, we must first consider what precipitates such sentiments and precisely how hate infused with religious zeal is used to spread anti-Americanism, transforming people

psychologically to the point where they are ready to commit unspeakable crimes. Certainly, the continuing rise of anti-American sentiment impedes US political manoeuvrability and undermines its influence, with potentially disastrous implications for its strategic national interests. According to several recent polls taken in many Arab and Muslim countries, 85 to 90 percent of the people have extremely negative views of the US. Hating the US is fashionable in this part of the world, and few dare to say anything positive.

In his article, Ben-Meir (2005) proposes seven sources of hatred towards the US among Muslims, i.e., i) US as an imperial power – US is at best suspected by certain segments of the Muslim and the Arab world as embodying imperialism, colonialism, and capitalism; ii) Arrogance – The US is seen as an arrogant and morally decadent state, Muslims accused the US of demeaning, humiliating and inconsiderate policies that negatively affect their lives; iii) Corrupting culture – Muslim religious radicals fear the pervasiveness of US culture and what they perceive as its disastrous influence on Muslim youth; iv) Self-indulgence – The US is also seen as self-indulgent, uncaring, it neither understands nor shows any interest in understanding other people's needs, culture and aspirations; v) Lack of even-handedness – failure to be even-handed in its foreign policy, the US is accused of favouring Israel over the Arab States, specifically the Palestinians. Moreover, the US is always seen as being prejudiced towards Muslims and always labelling Muslims as terrorists; vi) US as a "rogue" state – most Muslims feel that America is one itself. They view both Gulf wars, which resulted in the deaths of thousands of Iraqis, as criminal acts; and lastly, vii) Envy – The US is envied because of its wealth, freedom, outreach, and

immense human and material resources as well as their economic and military power.

Put all these together, and it's not hard to see how the US becomes the target for its detractors. From time immemorial, leaders of all political and philosophical persuasions have attempted to blame others for their own shortcomings and endemic problems. The bigger the problem they faced, the larger and more significant the "responsible" party had to be. From a Muslim perspective, the United States represents all that is bad and evil in their society not simply because it is a superpower with unprecedented influence, but because its power is so visible and domineering (Ben-Meir, 2005).

This shows to us that the feeling of victimization and dissatisfaction among Muslims in the world towards the US existed and possibly it could negatively affect their attitude towards US made products by taking proactive action such as boycotting the products. Thus, after identifying, considering and assessing several countries, the US has been chosen as the foreign country that is going to be used in this study.

Meanwhile, for product selection, in the country of origin and foreign product evaluation research, the effects seems to exist when the products selected are in general (Nagashima, 1977; Wall and Heslop, 1986; Darling and Wood, 1990; Howard, 1989; Kaynak et al., 2000; Suh and Kwon, 2002; Balabanis et al., 2002; Shin, 2001; Ettenson and Klein, 2005; Laroche et al., 2005; Kinra, 2006), for certain product categories (Cordell, 1992; Hong and Wyer, 1990; Roth and Romeo, 1992; Kim and Pysarchik, 2000; Teas and Agarwal, 2000; Nijssen and Douglas, 2004),

specific brands (Chao, 1993; Han and Terpstra, 1988; Tse and Gorn, 1993; Knight and Calantone, 2000; Watson and Wright, 2000; Cervino, Sanchez and Cubillo, 2005), consumer products (Bannister and Saunders, 1978; Papadopoulos, Heslop and Bamossy, 1989; Ahmed and d'Astous, 2004) and also for industrial products (White, 1979; Chasin and Jaffe, 1987; Cattin et al., 1982; Ahmed and d'Astous, 1995; Edwards et al., 2006). Studies have claimed the generalizability of their findings in the area of foreign products' evaluation and the influence on consumer behaviour when evaluating quality, determining the willingness to buy, and the final purchase decision for products.

For the current study, general product evaluations were used in order to evaluate the products made in the US. This is consistent with previous studies (e.g. Kaynak et al., 2000 and Leonidou, Hadjimarcou, Kaleka, and Stamenova, 1999; Ettenson and Klein, 2005; Laroche et al., 2005; Kinra, 2006) where researchers did not focus on the specific type of products in evaluating foreign made products. This can avoid the bias towards certain products coming from certain countries. For example, the favourability of electrical appliances from Japan or the favourability of computer software developed by a US company.

3.3 Sources of Research Data

The sources of research data are a combination of both primary and secondary data gathered as follows:

3.3.1 Secondary Data

Secondary data is somewhat historical, already assembled and does not require direct access to the respondents. The data was previously collected and assembled for some projects other than the one at hand. Secondary information or data can often be found inside the company, in the library, on the Internet or it can be purchased from firms that specialize in providing information. Among the sources used to gather the information needed were on-line journals (for example, Journal of International Business Studies, Journal of Marketing Research and International Journal of Research in Marketing), and other related periodicals from libraries and resource centres as well as from the online Internet news sources (for example, www.islamonline.net). In addition, local newspapers such as The Star, Malay Mail, Berita Harian and Utusan Malaysia were also used as secondary data.

3.3.2 Primary Data

Primary data refers to the data collected directly from the original sources for a specific purpose. In other words, primary data is data gathered and assembled specifically for the project in hand. The primary data used for this research was gathered through the distribution of questionnaires to selected consumers. The scales to measure the constructs in this study were generally taken from previous studies. However, most scales were modified to suit the Malaysian environment. The scales were basically modified from earlier research conducted by Rusnah (2005), Klein et al (1998), Shimp and Sharma (1987), Kosterman and Feshbach (1989), Darling and Arnold (1988), Darling and Wood (1990), Wilde and Joseph (1997), and Pullman et

al., (1997). The details of the questionnaire will be discussed in a later section of this chapter.

3.4 Type of Research

Research can be classified into two basic categories: quantitative and qualitative research. According to Smith, Smoll, and Curtis (1983), quantitative research employs the traditional, the positivist, the experimental, or the empiricist method to inquire into an identified problem, based on testing a theory, measured with numbers, and analyzed using statistical techniques. It emphasizes the objectivity and reproducibility. Meanwhile, Fraenkel and Wallen (2003) argue that the goal of quantitative methods is to determine whether the predictive generalizations of a theory hold true. Thus, quantitative research is more concerned with issues of how much, how well, or to whom an issue applies. Furthermore, Kerlinger and Lee (2000) explained that quantitative research is deductive in nature in which researchers make inferences based on direct observations with the primary goal to describe cause and effect.

By contrast, a study based upon a qualitative process of inquiry has the goal of understanding a social or human problem from multiple perspectives (Denzin and Lincoln, 2000). Thus the qualitative researchers deploy a wide range of interconnected interpretive practices, hoping always to get a better understanding of the subject matter at hand. Additionally, they also explained that qualitative research is conducted in a natural setting and involves a process of building a complex and

holistic picture of the phenomenon of interest as well as being inductive in nature in which researchers focus on delving into the issues of interest in depth and detail.

According to Fraenkel and Wallen (2003), quantitative research can be classified as either descriptive or experimental research. The purpose of descriptive research is to become more familiar with phenomena, to gain new insights, to formulate a more specific research problem or hypothesis. In contrast, experimental research is to test cause and affect relationships among variables. In descriptive research, researchers do not have direct control over independent variables because their manifestations have already occurred or because they are inherently not manipulable (Kerlinger and Lee, 2000). Quantitative research can be used to establish causal relationship among variables, so that the applicability in examining the process allows the researchers to test the hypothesized relationship between independent variables, mediating variable and the dependent variables.

Based on the literature reviews and the quantitative trend in research on consumer behaviour and marketing study, the current study adopted a quantitative research approach.

3.5 Measurement of the Constructs

This section discusses the measurement of the constructs used in the study. Measurement items in this study were generally generated from previous research. However, minor modifications were done in order to suit the current study environment. This was done following the feedback from the pre-testing. In this

case, faculty members who are experts in this area of study were asked to compare and evaluate the items in the scale. Three expert judges from the consumer behaviour field were, therefore, appointed to refine the generated items in an attempt to ensure that the scales developed have content and face validity. Nevertheless, the modifications do not alter the content of the constructs. Even though most of the measurement items were modified from existing scales, some of the items were newly developed based on the current study perspectives.

The discussion of the constructs is based on the sequence that was presented in the conceptual framework. The following sub-sections discuss the measurement of the constructs used in this study, that is Muslim religiosity, consumer animosity, consumer ethnocentrism, patriotism, US product judgment, purchase willingness and purchase action.

3.5.1 Measuring the Muslim Religiosity Construct

Religiosity is a difficult construct to measure (Scutte and Hosch, 1996) since there are several definitions of religiosity and it has been a key issue in the field of sociology since the 1960s. Among the dimensions that have been used and included in measuring religiosity are religious devotion, affiliation, ideology and others (King and Hunt, 1972).

From the perspective of consumer behaviour research, Caird (1987) proposes three different measures of religiosity: cognitive (focus on religious attitudes or beliefs), behavioural (evaluate church attendance or private prayer), and experiential (query

as to mystical experiences). Hirschman (1982) uses religious affiliation as opposed to religiosity in studying consumer consumption process. Wilkes et al. (1986) measured religiosity by church attendance, importance of religious values, confidence in religious values, and self-perceived religiousness. A few other studies focused on four factors, i.e., belief in the religious doctrine, religious practice or activity, the moral consequences, and an experience dimension or self-rating of one's religiosity (De Jong, Faulkner, and Warland, 1976; King and Hunt, 1972).

Sood and Nasu (1995), used a nine-item 5-point Likert scale to measure religiosity among respondents in the US and Japan anchored by 1 = strongly disagree and 5 = strongly agree. Some of the questions were different for the US and Japanese respondents based on their religious belief. In a study conducted by Vitell et al., (2005), religiosity was measured by the intrinsic/extrinsic religiousness scale adopted from Allport and Ross (1967) using a 5-point Likert type scale. Slight changes were made in the wording so that the scale would not appear to be measuring any specific religion, but rather a general religious commitment. For example, the wording “attending church” was changed to “attending religious services”. The intrinsic scale has eight items and is exemplified by items such as, “I try hard to live my life according to my religious beliefs”. The extrinsic dimension contains six items and is exemplified by items such as, “I go to religious services because it helps me to make friends”. In total, they used fourteen items in measuring religiosity (Vitell et al, 2005).

A lot of researchers have explained that the behaviour and beliefs displayed by different religious traditions were highly diverse and specific. In that context, the

various aspects of religion may not naturally fall into a continuum within specific dimensions (Hood et al., 1996; and Loewenthal, 2000). However, several researchers have suggested that a unidimensional conceptualization of religiosity should be considered when the relationship between religiosity and general cultural variables is examined or when the relations are tested in different religious groups (Schwartz and Huismans, 1995; and Cukur et al., 2004). Therefore, Cukur et al. (2004) measured religiosity using a single item, i.e., "My religious beliefs are very important to me" and a 7-point Likert scale ranging from 1 = strongly disagree to 7 = strongly agree. The item focuses on the nature of the respondents' belief about their strength of religious belief.

From the perspective of Muslim religiosity, Asyiq (2004) examine the role of self religiosity and father's religious attitude in the moral behaviour of youth from religious and nonreligious schools in Malaysia. In measuring the self religiosity, the scale consisted of nine positive and nine negative items in the form of statements. Each statement had to be rated on a Likert type rating scale comprising categories of 1 = strongly disagree, 2 = disagree, 3 = do not know, 4 = agree, and 5 = strongly agree. A high score on the scale indicated the religious orientation of the subject. An example of the questions in their study are; "Religion gives me comfort and security in life" and "When I face a problem, religion is the last thing on my mind".

Furthermore, in Asyiq (2004), for father's religious attitude scale, a fifteen-item scale was given consisting of nine positive and six negative items. Each item had to be rated on a 5-point Likert type scale with the categories of 1 = strongly disagree, 2 = disagree, 3 = do not know, 4 = agree, and 5 = strongly agree. A high score

indicated the religious attitudes of the father. An example of a positive item of the scale was "I send my children to mosque to attend religious activities", and a negative item was, "Religion does not play a role in the upbringing of my children."

In Khashan and Kreidie (2001), their study focuses on the social and economic correlate of Islamic religiosity. They have constructed a continuum of degree of religiosity, the most extreme category being occupied by individuals accepting violence as a legitimate means of achieving Islamic goals. The opposite end of the continuum is anchored by individuals with secular or simply noncommittal religious tendencies. The continuum includes five parts divided into 12 components, i.e., i) noncommittal religious tendencies (general support for religious groups, religion from a personal perspective and religious self-description); ii) ritualistic religious behaviour [performance of *Ibadah* (worship) requirement and surpassing of *Ibadah* (worship) requirement]; iii) religiosity and civil interaction (religion and lifestyle, display of religiosity and involvement with community); iv) religiosity and conventional political participation (support for political participation and support for non-violent action); and v) religiosity and violence (predisposition to violence).

Another study, Ali, Meyer and Lockler (1997) used six items to measure Muslim religiosity in his study. A scale was used in the study that was conducted in Kuwait. The items included; i) It is my duty to help those who are religiously misguided; ii) It is of utmost importance to enlighten people of their religion even if it develops unpleasant situations; iii) The world will be a better place for living if people follow my same religious ideas; iv) I believe that the world's problems are dangerously increasing because many people are misguided religiously; v) I believe that all

Muslims must make a firm stand against western challenges and sovereignty; and vi) I believe that Islam does not separate between religion and politics.

Wilde and Joseph (1997) have developed the Muslim Attitude towards Religiosity Scale (MARS) under the guidance and supervision of the University of Essex Islamic Society. The scale contained 14 items adapted from the Francis (1993) scale as presented in Table 3.1. Each item was rated by the respondents on a 5-point Likert scale ranging from 1 = strongly disagree to 5 = strongly agree.

Table 3.1
Wilde and Joseph (1997) Muslim Attitude towards Religiosity Scale

No.	Items
1	I find it inspiring to read the Qu'ran.
2	Allah helps me.
3	Saying my prayers helps me a lot.
4	Islam helps me lead a better life.
5	I like to learn about Allah very much.
6	I believe that Allah helps people.
7	The five prayers help me a lot.
8	The supplication (*dua*) helps me.
9	I think the Qu'ran is relevant and applicable to modern day.
10	I believe that Allah listens to prayers.
11	Mohammed (peace be upon him) provides a good mode of conduct for me.
12	I pray five times a day.
13	I fast the whole month of Ramadan.
14	I observe my daily prayers in the Mosque.

Rusnah (2005), in her study to measure Muslim religiosity and perceptions of unethical practice among Muslims, adapted the scale developed by Wilde and Joseph (1997). As presented in Table 3.2, in her study, item modification was applied in order to match the questions asked with the study itself. In addition to the 14 items in the original scale she added another two items making it 16 in total.

Table 3.2
Modified Muslim Attitude towards Religiosity Scale by Rusnah (2005)

No.	Items
1	Do you believe, beyond a shadow of doubt, that Islam is God's religion and that Prophet Muhammad is His Messenger?
2	Do you consider yourself religious?
3	I read the Qu'ran for inspiration and motivation.
4	I believe that Allah helps me.
5	Saying my prayers helps me a lot.
6	Islam helps me lead a better life.
7	I like to learn about Allah very much.
8	I believe that Allah helps people.
9	The five prayers help me a lot.
10	The supplication (*dua*) helps me.
11	Qu'ran is relevant and applicable to modern day.
12	Allah listens to prayers.
13	Muhammad (peace be upon him) provides a good conduct for me.
14	I pray five times a day.
15	I fast the whole month of Ramadan.
16	I perform my daily prayers in the Mosque.

Khairul Anwar (2001) in his study in Malaysia on the possible dimensions of religiosity in Islam developed a self-designed scale. Three questions were asked, i.e., i) What are the specific indications that one is religious?; ii) In your opinion, is the practice of Islam increasing, unchanged or decreasing?; iii) In your opinion, how do you rate yourself of being religious? As a result, fifteen major domain or indications were extracted by Khairul Anwar. Some of the domains were *solat* (prayer), *muamalah* (relation to others), fasting, Al-Quran, *akhlaq* (practice of virtue, morality and manners in Islam), thinking, knowledge, parental relationship, s*unnah* (normative for Muslims on the basis of the teachings and practices of Prophet Muhammad), hajj and eating habits.

In the current study, basically, the MARS scale proposed by Wilde and Joseph (1997) and Rusnah (2005) were adapted. With some modifications and additional items or questions, in total, 21 items were used to measure Muslim Religiosity among Malaysian Muslim consumers. The modifications of the items were deemed appropriate since the background of this and the previous studies were not the same. The items used in the construct measurement in this study are presented in Table 3.3.

Items 1 and 2 from the study conducted by Rusnah (2005) were excluded from this study because the researcher believes that the questions were considered to be leading questions and quite sensitive to the respondents. As explained in the earlier part of this chapter, people or respondents are quite reluctant to participate in research relating to religious issues. To avoid unanswered questions, the questions that are deemed inappropriate were excluded. Furthermore, in the exploratory factor

analysis, the items that are statistically not so important in capturing the construct will be excluded from further analysis.

Table 3.3
The Muslim Religiosity Scale Used in the Current Study

No.	Items	Source
1	I read the Quran for inspiration and motivation.	Wilde and Joseph (1997)
2	I believe that Allah helps me.	Wilde and Joseph (1997)
3	Saying my prayers helps me a lot.	Wilde and Joseph (1997)
4	Islam helps me lead a better life.	Wilde and Joseph (1997)
5	I will continuously seek to learn about Allah.	Wilde and Joseph (1997)
6	I believe that Allah helps people.	Wilde and Joseph (1997)
7	The five prayers help me a lot.	Wilde and Joseph (1997)
8	The supplication (*dua'*) helps me.	Wilde and Joseph (1997)
9	Quran is relevant and applicable to modern day.	Wilde and Joseph (1997)
10	I believe that Allah listens to prayers.	Wilde and Joseph (1997)
11	Muhammad (peace be upon him) provides a good conduct for me.	Wilde and Joseph (1997)
12	I pray five times a day.	Wilde and Joseph (1997)
13	I perform the obligation of *zakat fitrah* annually.	Self-developed
14	I perform the obligation of *zakat maal* (asset/income) annually.	Self-developed
15	I read the Quran every day.	Self-developed
16	I fast the whole month of Ramadan sincerely.	Wilde and Joseph (1997)
17	I perform my daily prayers in the Mosque / Muslim praying room regularly.	Wilde and Joseph (1997)
18	I always perform other optional prayer (i.e. *sunnat* prayer such as *Isra'*, *Dhuha* and others).	Self-developed
19	I do the optional fasting on Monday and Thursday regularly.	Self-developed
20	I will perform hajj after I fulfilled all the necessary conditions.	Self-developed
21	I cover my *aurat* properly.	Self-developed

Seven items were added because the researcher believes that these items are very significant issues to measure Muslim religiosity. As suggested by Khairul Anwar (2001), there are hundreds of indicators to indicate Muslim religiosity. So that, it is believe that including the important issues will further improve the measurement of Muslim religiosity construct. Additionally, the items used in Wilde and Joseph (1997) and Rusnah (2005) are lack of focus on optional *ibadah* (worship) such as optional prayer and optional fasting which are in practice, normally performed by devout Muslims. Finally, to maintain the consistency throughout the questionnaire, a 7-point Likert scale was employed to measure this construct.

3.5.2 Measuring the Consumer Animosity Construct

The construct of consumer animosity was first studied by Klein et al. (1998). In that study, the researchers developed a scale to measure the animosity among Chinese consumers in Nanjing, China towards Japan due to their historical war. It was divided into two types of animosity, i.e., war and economic animosity. The items they used to measure consumer animosity are presented in Table 3.4. The scale used a 7-point Likert scale ranging from 1 = strongly disagree and 7 = strongly agree. A higher score indicates a greater level of animosity.

Later research also adopted the original nine-item 7-point Likert scale [see Shoham et al. (2006); Hinck and Felix (2000); Klein and Ettenson (1999); Wittkowski (2000); and Shin (2001)]. Ettenson and Klein (2005) meanwhile used only four modified items to measure animosity towards France due to French nuclear testing in the South Pacific in 1995. The items are i) I feel angry towards France; ii)

France's recent nuclear testing was an act of aggression in the South Pacific; iii) France does not care what Australia or other nations think of its actions; and iv) I will never forgive France for its nuclear testing in the South Pacific.

Table 3.4
Consumer Animosity towards Japan Scale by Klein et al. (1998)

No.	Items
1	I dislike the Japanese.
2	I feel angry towards the Japanese.
3	I will never forgive Japan for the Nanjing massacre.
4	Japan should pay for what it did to Nanjing during the occupation.
5	Japan is not a reliable trading partner.
6	Japan wants to gain economic power over China.
7	Japan is taking advantage of China.
8	Japan has too much economic influence in China.
9	The Japanese are doing business unfairly with China.

Another study using the consumer animosity construct was developed by Hinck (2005). In his study, the focus was on the domestic animosity between the ex-East German consumers towards products from ex-West Germany. Because of the nature of his study all items referring to war animosity were taken off, so that only the economic animosity was measured. A four-item 5-point Likert scale was used to measure the economic animosity construct. The items were; i) The *Alte Länder* (West Germany) want to gain power over the *Neue Länder* (East Germany); ii) The *Alte Länder* are taking advantage of the *Neue Länder*; iii) The *Alte Länder* have too

much influence in the *Neue Länder*; and iv) The *Alte Länder* are unfair with the *Neue Länder*.

Researchers also try to explore the effects of animosity in business to business (B2B) market. Edwards et al. (2006), try to explore the effect of animosity in the industrial market instead of the consumer market. Russell and Russell (2006) used a six-item 5-point Likert scale in their study and they only focused on economic animosity items.

Jung et al. (2002) and Ang et al., (2004) classified the animosity construct into situational animosity, stable animosity, national animosity and personal animosity. In their study, the measurement of the constructs was generated from a pool of items and a review of prior research. The initial pool of items was reduced from 120 to 36, and then the remaining items were re-evaluated for their face validity. Finally, 15 items were selected as the measurement for animosity construct. A five-point Likert scale anchored on 1 = strongly agree and 5 = strongly disagree were employed for the items.

The scale of the consumer animosity construct developed by Klein et al., (1998), was adopted and modified by Nijssen et al. (1999) as well as Nijssen and Douglas (2004) when they tried to test the animosity model for consumers in the Netherlands towards German products'. In both studies, they modified the war and economic animosity to fit the Dutch – German situation. Instead of using nine original items, they modified the scale and their final consumer animosity construct measurement contains 11 items. Table 3.5 shows the items used by Nijssen et al. (1999) and

Nijssen and Douglas (2004). The first five items in the table were related to war animosity and the remaining six items measured the economic animosity.

Table 3.5
Modified Consumer Animosity Scale by Nijssen and Douglas (2004)

No.	Items
1	I feel anger because of the role that the Germans played in World War II.
2	I can still get angry over Germany's role in World War II.
3	I will never forgive the Germans for occupying our country and pursuing the Jews.
4	Germany is liable for the damage caused by the bombardment of Rotterdam in 1940.
5	I will never forgive the Germans for bombing of Rotterdam in 1940.
6	While doing business with Germans one should be careful.
7	German companies are unreliable trading partners (e.g., Fokker-Dasa).
8	Germany wants to gain economic power over the Netherlands.
9	German companies often outsmart Dutch companies in business deals.
10	Germany has too much influence on the Netherlands and the Dutch economy.
11	German companies are doing business unfairly with the Dutch.

For the current study, all the items used in the study conducted by Nijssen and Douglas (2004) and items by Klein et al. (1998) were all employed to measure Muslim consumer animosity towards the US. It was divided into two types of animosity (war and economic animosity) as suggested by previous researchers in the consumer animosity literature. All the items were modified to suit the current research situation and, in addition, to comply with the situation between Muslims and the US, another two additional items that are related to the main issue in this

study were included. The items are i) US policies are always unfair towards the Muslim world, and ii) US actions against Muslim prisoners in Guantanamo detention centre annoyed me.

These issues (US policies on Muslim countries and Guantanamo detention centre) are very significant issues among Muslims. As explained in earlier chapter, many Muslims believed that the US government always given negative treatment on Muslims countries, for example, Iraq, Afghanistan and Palestine, they also believed that their Muslim brothers in Guantanamo detention centre were not getting fair judgment by the US government and have been arrested without logical reasons. These two items are deemed relevant for the current study and therefore, they are included in the current study to measure the consumer animosity among Muslims towards the US.

Additionally, a seven-point Likert scale ranging from 1 = strongly disagree and 7 = strongly agree was employed in this study to maintain the uniformity with other scales used in the current study. In total, 15 items are used to measure consumers' war and economic animosity in this study.

The items used for the current study are presented in Table 3.6.

Table 3.6
The Consumer Animosity Scale used in the Current Study

No.	Items	Source
1	I dislike the Americans.	Klein et al. (1998)
2	I feel angry towards US involvement in the war against several Muslim countries.	Klein et al. (1998)
3	I can still get angry over the US role in the war in Iraq and Afghanistan.	Nijssen and Douglas (2004)
4	I will never forgive the US for occupying Muslim countries and killing the civilians in that country.	Klein et al. (1998)
5	US is liable for the damage cause by the bombardment of Muslim countries.	Nijssen and Douglas (2004)
6	US should pay for what it did during the occupation of Muslim countries.	Klein et al. (1998)
7	US actions against Muslim prisoners in Guantanamo detention centre annoyed me.	Self-developed
8	I will never forgive the US for bombing Muslim countries.	Nijssen and Douglas (2004)
9	US policies are always unfair towards the Muslim world.	Self-developed
10	When doing business with the US one should be careful.	Nijssen and Douglas (2004)
11	US companies are unreliable trading partners.	Klein et al. (1998)
12	US wants to gain economic power over the Muslim countries.	Klein et al. (1998)
13	US companies often outsmart Muslim companies in business deals.	Klein et al. (1998)
14	US have too much influence on the Muslims and their countries' economy.	Klein et al. (1998)
15	US companies are doing business unfairly with the Muslim companies.	Klein et al. (1998)

3.5.3 Measuring the Consumer Ethnocentrism Construct

Ethnocentrism is a socio-psychological concept that can be assessed like attitudes and it is a theoretical construct, unobservable in a direct manner, and, thus, has to be measured by socio-psychological instruments (Shimp and Sharma, 1987). Shimp and Sharma (1987) can be credited for having coined the term and for having done the first study that directly related the ethnocentrism construct to the consumer behaviour area. The construction of a unique scale is necessary because the classic measure of ethnocentrism, developed by Adorno et al. (1950) is not directly relevant to the study of marketing and consumer behaviour. Later versions of ethnocentrism scales are also available (e.g., Chang and Ritter, 1976; Warr, Faust and Harrison, 1967) but they also have little relevance to the study of consumer behaviour and marketing (Shimp and Sharma, 1987).

The measurement of consumer ethnocentrism was made possible with the development of the CETSCALE (Consumer Ethnocentric Tendencies Scale) by Shimp and Sharma (1987). In their study, they developed an instrument termed the CETSCALE to measure consumer ethnocentric tendencies for the case of consumers based in the US. The CETSCALE consists of items that measure the tendency of consumers to act consistently towards foreign and domestic products. Such tendencies may precede attitudes, but they are not the equivalent of attitudes, which tend to be object specific. In addition, to the extent that domestic products are viewed as superior, products from other countries (i.e. from out-groups) are objects of contempt to highly ethnocentric consumers (Shimp and Sharma, 1987). Through a series of purification studies and tests, the final CETSCALE consists of 17 items

using a 7-point Likert scale (1 = strongly disagree and 7 = strongly agree). The items used in the construct measurement for Shimp and Sharma (1987) are presented in Table 3.7.

Table 3.7

Original 17-item of Consumer Ethnocentric Tendencies Scale (CETSCALE) by Shimp and Sharma (1987)

No.	Items
1	American people should always buy American products instead of imports.
2	Only those products that are unavailable in the US should be imported.
3	Buy American-made products. Keep America working.
4	American products, first, last and foremost.
5	Purchasing foreign-made products is un-American.
6	It is not right to purchase foreign products.
7	A real American should always buy American-made products.
8	We should purchase products manufactured in America instead of letting other countries get rich off us.
9	It is always best to purchase American products.
10	There should be very little trading or purchasing of goods from other countries unless out of necessity.
11	Americans should not buy foreign products, because this hurts American business and causes unemployment.
12	Curbs should be put on all imports.
13	It may cost me in the long run but I prefer to support American products.
14	Foreigners should not be allowed to put their products on our markets.
15	Foreign products should be taxed heavily to reduce their entry to the US.
16	We should buy from foreign countries only those products that we cannot obtain within our own country.
17	American consumers who purchase products made in other countries are responsible for putting their fellow Americans out of work.

The CETSCALE is used to differentiate among consumers who perceive, to varying degrees, that buying foreign made goods or services is acceptable or unacceptable. The scale was constructed and initially validated in the English language using four samples in the US (Shimp and Sharma, 1987).

Previous studies have investigated the nature of the consumer ethnocentrism construct in the US (Shimp and Sharma, 1987; Durvasula et al., 1997), Germany, France, Japan (Netemeyer et al., 1991), Russia (Durvasula et al., 1997), New Zealand (Watson and Wright, 2000), and Australia (Acharya and Elliott, 2003). All these studies employed the 17 items of CETSCALE with a 7-point Likert scale from 1 = strongly disagree to 7 = strongly agree. Other studies also employed all the 17 original items on a 7-point Likert scale including Brodowsky et al., (2004); Kucukemiroglu (1999); Ang et al., (2004); Supphellen and Gronhaug (2003); Wang and Chen (2004); Martinez et al., (2000); Supphellen and Rittenburg (2001); Javalgi et al. (2005); and Hamin and Elliot (2006).

Hamin and Elliot (2006) argued that in order to have comparable results with the previous studies, the full 17 items of CETSCALE were also included in the survey instrument used in their study. Besides the seven-point scale that has been widely used, other researchers used the 5- point Likert scale from 1 = strongly disagree and 5 = strongly agree (Kaynak and Kara, 2001); (Lee et al., 2003); and (Yu and Albaum, 2002), as well as a 9-point Likert scale from 1 = strongly disagree to 9 = strongly agree (de Ruyter et al., 1998).

The concept of consumer ethnocentrism is a relatively straightforward construct, which may be measured with fewer than ten items. Hence, an iterative procedure was used to explore the feasibility of a reduced version of the ten-item CETSCALE (Klein et al., 2006). When doing a second study for the new scale, Shimp and Sharma (1987) encountered major questionnaire space limitations, and, therefore, applied a reduced ten-item version of the scale, which they found to be of similar validity. Similarly, Lindquist et al. (2001) used the ten items CETSCALE to a validation test in Czech Republic, Hungary and Poland. They found that the reduced ten-item version not found to be good fit in these three countries. A good fitting five-item model was found for Hungary, a six-item scale for Poland and a seven-item solution for the Czech Republic.The ten-item CETSCALE used in Shimp and Sharma (1987) and Lindquist et al. (2001) is shown in Table 3.8.

Steenkamp and Baumgartner (1998) successfully used this ten-item scale as the basis for assessing cross-national measurement variance, contrasting Belgium, Great Britain and Greece using 7-point Likert scale from 1 = strongly disagree to 7 = strongly agree. Other research using the ten-item version include studies by Durvasula et al., (1997); Netemeyer et al. (1991); Steenkamp and Baumgartner (1998); Lindquist et al. (2001); Douglas and Nijssen (2003); Balabanis and Diamantopoulos (2004); and Ettenson and Klein (2005). Besides the seven-point Likert scale, five-point Likert scales were also employed (Douglas and Nijssen, 2003). Most consumer ethnocentrism researchers employed the CETSCALE using a seven-point Likert scale, whilst some researchers do use the five-point Likert scale. Only in exceptional cases do the researchers find other decisions such as nine point scales (e.g. de Ruyter et al., 1998).

Table 3.8
Ten-item Modified CETSCALE by Shimp and Sharma (1987)

No.	Items
1	Only those products that are unavailable in the US should be imported (2).
2	American products, first, last and foremost (4).
3	Purchasing foreign-made products is un-American (5).
4	It is not right to purchase foreign products (6).
5	A real American should always buy American-made products (7).
6	We should purchase products manufactured in America instead of letting other countries get rich off us (8).
7	Americans should not buy foreign products, because this hurts American business and causes unemployment (11).
8	It may cost me in the long run but I prefer to support American products (13).
9	We should buy from foreign countries only those products that we cannot obtain within our own country (16).
10	American consumers who purchase products made in other countries are responsible for putting their fellow Americans out of work (17).

** In ( ) indicated original CETSCALE number.*

As most of the literature suggest that five-point to seven-point scales are adequate for the majority of surveys, the current research employed the seven-point Likert scale anchored with "1 = strongly disagree" and "7 = strongly agree" – like those normally used for consumer ethnocentrism research. For the number of items used to measure consumer ethnocentric tendencies, as mentioned by Hamin and Elliot (2006), in order to have comparable results with the previous studies, the full 17 items CETSCALE should be employed. Consequently, all the 17 items were chosen in the current research.

All the modified items used in the current study are shown in Table 3.9.

Table 3.9
The Modified CETSCALE Used for the Current Study

No.	Items	Source
1	Malaysian people should always buy Malaysian-made products instead of imports.	Shimp and Sharma (1987)
2	Only those products that are unavailable in Malaysia should be imported.	Shimp and Sharma (1987)
3	Buy Malaysian-made products. Keep Malaysians working.	Shimp and Sharma (1987)
4	Malaysian products, first, last and foremost.	Shimp and Sharma (1987)
5	Purchasing foreign-made products is un-Malaysian.	Shimp and Sharma (1987)
6	It is not right to purchase foreign products.	Shimp and Sharma (1987)
7	A real Malaysian should always buy Malaysian-made products.	Shimp and Sharma (1987)
8	We should purchase products manufactured in Malaysia instead of letting other countries get rich off us.	Shimp and Sharma (1987)
9	It is always best to purchase Malaysian products.	Shimp and Sharma (1987)
10	There should be very little trading or purchasing of goods from other countries unless out of necessity.	Shimp and Sharma (1987)
11	Malaysians should not buy foreign products, because this hurts Malaysian business and causes unemployment.	Shimp and Sharma (1987)
12	Curbs should be put on all imports.	Shimp and Sharma (1987)
13	It may cost me in the long run but I prefer to support Malaysian products.	Shimp and Sharma (1987)
14	Foreigners should not be allowed to put their products on our markets.	Shimp and Sharma (1987)
15	Foreign products should be taxed heavily to reduce their entry to the Malaysian market.	Shimp and Sharma (1987)
16	We should buy from foreign countries only those products that we cannot obtain within our own country.	Shimp and Sharma (1987)
17	Malaysian consumers who purchase products made in other countries are responsible for putting their fellow Malaysians out of work.	Shimp and Sharma (1987)

3.5.4 Measuring the Patriotism Construct

Patriotism is an attachment to and a sense of pride in one's own country, a desire to live there, a readiness to make sacrifices for it, and a respect for and loyalty towards its people (Barnes and Curlette, 1985). It is a willingness to love, support, and defend one's country against out-groups (Barnes and Curlette, 1985; Feshbach, 1987).

Patriotism should be distinguished from ethnocentrism that suggests an uncritical acceptance of one's own nation as superior to and deserving to be more powerful than other nations. The literature states that patriotic consumers are loyal to their countries but do not reject other countries. Han (1988) believes that patriotism has a positive influence on consumers' choice. However, in general, behavioural research involving patriotism has been contradictory in finding the influence of these attitudes (Heaven, Stones, and Bester, 1986; and Ray and Lovejoy, 1986).

Granzin and Olsen (1998) and Pullman et al. (1997) measured the patriotism construct using an eight-item 6-point Likert scale anchored by strongly disagree and strongly agree. They adapted the patriotism measurement from Levinson (1950) and used only two items, i.e., i) Patriotism and loyalty are the first and most important requirements of a good citizen; and ii) America may not be perfect, but the American way has brought us about as close as people can get to a perfect society. According to Pullman et al. (1997), because these items operationalized behavioural constructs rather than perceptions of fact, it was considered appropriate to use a six-

point format to avoid the mid-scale choices commonly obtained with an odd number of scale points (i.e. choices taken to represent "undecided" or "no opinion").

De Ruyter et al. (1998), in his study on the effect of patriotism on consumer ethnocentrism, used the measurement developed by Levinson (1950) by using five items to measure the patriotism construct. The study does not attach the complete scale and just presents two items as an example, i.e., i) Patriotism and loyalty are the first and most important requirements of a good citizen; and ii) People throughout the world may be in need, but it would be a mistake to lower our immigration quotas and allow them to flood into the country. Furthermore, instead of using a 6-point Likert scale as used by Granzin and Olsen (1998), they used a 9-point Likert scale.

Javalgi et al. (2005) measured patriotism in their study on consumer ethnocentrism in France. They investigated how the socio-psychological construct (patriotism) would influence the ethnocentric tendencies of French people. Their scale consisted of two items, which were obtained from Sharma et al. (1995). A sample item used in their research was “Devoting oneself to one’s country is worthwhile”. In addition, Han (1988) measured the patriotism construct using a four-item 5-point Likert scale anchored by “I strongly feel” and “I strongly do not feel” with “I am not sure as a midpoint”. The patriotic response was measured by the respondents’ emotional intensity.

Lee et al. (2003) in a study investigating the US consumers in the post 9/11 climate has included the construct of ethnocentrism, patriotism, nationalism and internationalism. In their study, patriotism was measured using the scale developed

by Kosterman and Feshbach (1989). The items for the scale were measured on a 5-point Likert scale type ranging from 1 = strongly agree to 5 = strongly disagree. Patriotism was measured on a 12-item scale, which included items such as 'I love my country'. The complete scale in Kosterman and Feshbach study is presented in Table 3.10.

Table 3.10
Patriotism Scale by Kosterman and Feshbach (1989)

No.	Items
1	I love my country.
2	I am proud to be an American.
3	In a sense, I am emotionally attached to my country and emotionally affected by its actions.
4	Although at times I may disagree with the government, my commitment to the US always remains strong.
5	I feel a great pride in that land that is our America.
6	It is not that important for me to serve my country*.
7	When I see the American flag flying, I feel great.
8	The fact that I am an American is an important part of my identity.
9	It is not constructive for one to develop an emotional attachment to one's country*.
10	In general, I have very little respect for the American people*.
11	It bothers me to see children made to pledge allegiance to the flag or sing the national anthem or otherwise induced to adopt such strong patriotic attitudes*.
12	The US is really just an institution, big and powerful yes, but just an institution*.

*Note: * – reverse score*

Meier-Pesti and Kirchler (2003) also adapted the scale developed by Kosterman and Feshbach (1989) to measure patriotism. Basically, the study tries to investigate the attitude of Austrians towards the European currency and the European identity,

which may be affected by their patriotic attitude. In their studies, they modified and used only 11 items and dropped the last items, i.e., the big and powerful institution. The items for the scale were measured on a 7-point Likert scale.

The items used to measure the patriotism construct in the current study are presented in Table 3.11.

Table 3.11
The Measurement Scale of Patriotism for Current Study

No.	Items	Source
1	I love my country.	Kosterman and Feshbach (1989)
2	I am proud to be a Malaysian.	Kosterman and Feshbach (1989)
3	In a sense, I am emotionally attached to my country and emotionally affected by its actions.	Kosterman and Feshbach (1989)
4	Although at times I may not agree with the government, my commitment to Malaysia always remains strong.	Kosterman and Feshbach (1989)
5	I feel a great pride in that land that is our Malaysia.	Kosterman and Feshbach (1989)
6	It is not that important for me to serve my country*.	Kosterman and Feshbach (1989)
7	When I see the Malaysian flag flying, I feel great.	Kosterman and Feshbach (1989)
8	The fact that I am a Malaysian is an important part of my identity.	Kosterman and Feshbach (1989)
9	It is not constructive for one to develop an emotional attachment to one's country*.	Kosterman and Feshbach (1989)
10	In general, I have very little respect for the Malaysian people*.	Kosterman and Feshbach (1989)
11	It bothers me to see children made to pledge allegiance to the flag or sing the national anthem or otherwise induced to adopt such strong patriotic attitudes*.	Kosterman and Feshbach (1989)

*Note: * – reverse score*

For the current study, patriotism was measured using an 11-item 7-point Likert scale. The scale is generally adapted from a study conducted by Kosterman and Feshbach (1989). From 12 items used in Kosterman and Feshbach study, only 11 items carried out for the current study. The last item from the original measurement is dropped due to the inappropriateness of the question from the Malaysian perspective and the nature of the question itself. The item, "the US is really just an institution, big and powerful yes, but just an institution" was deemed as not suitable in this study when the word "US" is changed into "Malaysia" where Malaysia can be considered as not as big and powerful as the US. The seven points were used in order to maintain consistency with the other constructs used in this study as well as to avoid the misunderstanding and confusion among the respondents.

3.5.5 Measuring the US Product Judgment Construct

Foreign product judgment is extensively used by researchers in studies related to international marketing and consumer behaviour such as the country of origin effects, consumer animosity and consumer ethnocentrism. Researchers argue that to study the effects of consumers' product evaluations and purchase decisions, respondents should be asked to evaluate products based on the available information and attributes of the products (Ettenson, 1993; Lin and Sternquist, 1994; Tse and Gorn, 1993; Malhotra and McCort, 2001). It shows that to measure foreign product judgment, researchers need to ensure the construct contains multiple attributes of the products and a single product cue is not sufficient.

During the end of the sixties and early seventies, Nagashima (1970, 1977) developed a popular measurement scale used to measure foreign product judgment in the country of origin study. The 20-item semantic differential scale shown in Table 3.12 was proposed and grouped under five dimensions: a) price and value, b) service and engineering, c) advertising and reputation, d) design and style and e) consumers' profile. This scale was later used by numerous researchers (for examples, Ulgado and Lee, 1998; Ahmed and d'Astous, 2004; and Ravi, Pascale and Ray, 2007).

Items used by Nagashima to measure the foreign product judgment were reduced by factor analysis from 20 to four by Han and Terpstra (1988). The authors did not give the details of the item-factor structure but only the following four underlying dimensions: a) technical advancement, b) prestige, c) workmanship and d) price. They also added "serviceability" and "overall evaluation". The items for the scale were measured on a 7-point Likert scale instead of semantic differential scale as suggested by Nagashima (1977).

In studies related to this construct, several scales were developed and modified by previous researchers depending on their nature and type of study. Lim and Darley (1997) measured the construct of foreign product judgment using a four-item, 7-point scale anchored by the following adjectives or phrases: inexpensive/expensive; of careful and meticulous workmanship/of not careful and meticulous workmanship; technically advanced/technically backward; of good serviceability/of poor serviceability.

Zhang (1996) used five instead of four items to measure foreign product judgment. A 7-point semantic differential scale (unreliable/reliable, common/exclusive, of not

careful and meticulous workmanship/of careful and meticulous workmanship, technically not advanced/advanced, of poor/good style) was used to measure foreign product judgment. In another study, Zhang (1997) used a three item nine-point semantic differential scale ranging from – 4 (very poor, very low quality, inferior) to 4 (very good, very high quality, superior) to measure subjects' overall product evaluation.

Table 3.12
Foreign Product Judgment Scale by Nagashima (1977)

Dimensions	**Bipolar adjectives (7-point scale)**
Price and value	Inexpensive vs. expensive
	Reasonably priced vs. unreasonably priced
	Reliable vs. unreliable
	Luxury items vs. necessary items
	Exclusive vs. common
	Heavy industry product vs. light manufacture product
Service and engineering	Careful and meticulous workmanship vs. not so careful and meticulous workmanship
	Technically advanced vs. technically backward
	Mass produced vs. hand made
	Worldwide distribution vs. mostly domestic distribution
	Inventive vs. imitative
Advertising and reputation	Pride of ownership vs. not much pride of ownership
	Much advertising vs. little advertising
	Recognizable brand name vs. unrecognizable brand name
Design and style	Large choice of size and model vs. limited choice of size and model
	More concerned with outward appearance vs. more concerned with performance
	Clever use of colour vs. not clever use of colour
Consumers' profile	More for young people vs. more for old people
	More for men vs. more for women
	Upper class vs. lower class

Darling and Arnold (1988), in their study, presented an analysis of general attitudes of consumers in Finland towards the products imported from the US, Japan and selected European countries. The questionnaire they developed contains 31 items and 13 of them were intended to measure attitudes of consumers towards the products from the selected countries. The details of the measurement scale for the foreign product judgment used by Darling and Arnold (1988) are presented in Table 3.13.

Table 3.13
Foreign Product Judgment Scale by Darling and Arnold (1988)

No.	Items
1	Products made in ____ are generally very well suited to needs of the Finnish consumers.
2	The suitability of products made in ____ to the Finnish consumers seems to have improved over the past several years.
3	Products made in ____ occupy very strong competitive position in comparison to the products of other countries.
4	Products made in ____ are carefully produced and have a fine workmanship.
5	Products made in ____ are generally of a lower quality than similar products available from other countries.
6	Over the past several years, the quality of most products made in ____ seems to have improved.
7	Products made in ____ show a very high degree of technological advancement.
8	Products made in ____ generally lack creativity and are very imitative of products made in other countries.
9	Products made in ____ are generally available in a wide range of sizes and models.
10	Products made in ____ are produced by firms that are more concerned with the outward appearance of the products than with the products performance.
11	Products made in ____ usually show a very clever use of colour and design.
12	Products made in ____ are usually quite reliable and seem to last the desired length of time.
13	Products made in ____ seem to be more in the nature of luxury items than necessary items.

Other studies that include products' judgment as a variable in their study have their own way to measure them. For example, Schaefer (1997) used ratings on a three-item 7-point semantic differential scale on the dimensions of "overall quality", "value for money", and "social acceptability/trendiness". Cerviño et al. (2005) used a 7-item 5-point Likert scale ranging from excellent to poor with such attributes as quality of products, competitive prices, design, innovation and technologically advanced, good service and prestige of brand names. Meanwhile, Watson and Wright (2000) employed a six-item 7-point semantic differential scale rating the technical advancement, prestige, workmanship, price, reliability, and value of the products. Additionally, Wang and Chen (2004) used a seven item 7 point Likert scale including workmanship, technological advancement, quality, reliability, design, and value for money.

Klein et al. (1998) used a six-item 7-point Likert scale ranging from 1 = strongly disagree to 7 = strongly agree to measure foreign product judgment adapted from Darling and Arnold (1988). It included items measuring the products' attributes of workmanship, technological advancement, quality, reliability, design, and value for money. The later studies on consumer animosity and product judgment constructs (for example, Shin, 2001; Nijssen and Douglas, 2004; Ettenson and Klien, 2005; Hinck, 2005; and Shoham et al., 2006), used either five or six modified items suggested by Klein et al. (1998).

Table 3.14 presents the items used to measure the foreign product judgment construct for the current study. The scale contains 13 items 7-point Likert scale adapted from Darling and Arnold (1998).

Table 3.14
Foreign Product Judgment Scale Used in the Current Study

No.	Items	Source
1	Products made in US are generally very well suited to needs of Malaysian consumers.	Darling and Arnold (1988)
2	The suitability of products made in US to the Malaysian consumers seems to have improved over the past several years.	Darling and Arnold (1988)
3	Products made in US occupy very strong competitive position in comparison to the products of other countries.	Darling and Arnold (1988)
4	Products made in US are carefully produced and have a fine workmanship.	Darling and Arnold (1988)
5	Products made in US are generally of a lower quality than similar products available from other countries*.	Darling and Arnold (1988)
6	Over the past several years, the quality of most products made in US seems to have improved.	Darling and Arnold (1988)
7	Products made in US show a very high degree of technological advancement.	Darling and Arnold (1988)
8	Products made in US generally lack creativity and are very imitative of products made in other countries*.	Darling and Arnold (1988)
9	Products made in US are generally available in different sizes and models.	Darling and Arnold (1988)
10	Products made in US are produced by firms that are more concerned with the outward appearance of the products than with the products performance*.	Darling and Arnold (1988)
11	Products made in US usually show a very clever use of colour and design.	Darling and Arnold (1988)
12	Products made in US are usually quite reliable and seem to last the desired length of time.	Darling and Arnold (1988)
13	Products made in US seem to be more in the nature of luxury items than necessary items*.	Darling and Arnold (1988)

*Note: * – reverse score*

To maintain consistency with earlier research and to get a clearer picture among Muslim consumers on products from the US, the same 13 items proposed by Darling and Arnold (1988) were adapted and used in this study. The items were then modified to suit the Malaysian environment and the nature of the study.

Furthermore, the scale used was Likert instead of semantic differential to be consistent with the other questions in the questionnaire as well as to ensure consistency with previous literature in the consumer animosity research.

3.5.6 Measuring the Purchase Willingness Construct

Purchase willingness has been widely used by researchers in international marketing and consumer behaviour research. Researchers have used different terms for this construct such as "purchase intention" (Han, 1988) and "willingness to buy foreign products" (Klein et al., 1998).

In some studies, purchase willingness was measured using a two-item, 7-point scale anchored by the following words: not probable/probable; impossible/possible (for example, Lim and Darley, 1997). Instead of using a 7-point scale, some research also used a 5-point rating scale ("definitely buy"/ "definitely not buy" in Elliot, Cameron and Acharya, 2000) and (1 = strongly agree, 5 = strongly disagree in (Uncles and Saurazas, 2000; and Zarkada-Fraser and Fraser, 2002), followed by a question about the respondents' likelihood of purchasing the products from foreign countries, which sought to determine the extent of the influence of country of origin on consumers' purchase intentions (Ahmed, Johnson, Yang, Fatt, Teng and Boon, 2004).

In other studies, several researchers using the term willingness to buy instead of purchase willingness. For example, in measuring the consumer willingness to buy foreign products, respondents were asked to indicate their agreement using a 7 point

Likert scale anchored by 1 = “strongly disagree” to 7 = “strongly agree” (Ettenson and Klein, 2005; Klein et al., 1998 and Wang and Chen, 2004). Willingness to buy construct was also measured in the study conducted by Darling and Arnold (1988), Darling and Wood (1990), Wood and Darling (1993), Jung (2001) and Nijssen and Douglas (2004).

Most of the studies that used the animosity construct tend to use the purchase willingness measurement developed by Darling and Arnold (1988) and later modified by Kline et al. (1998) to ensure the appropriateness and adequacy of the construct measurement. In a study conducted by Darling and Arnold (1998), they used a five item 5-point Likert scale (1 = strongly disagree to 5 = strongly agree) to measure the attitudes of the consumers regarding the willingness to purchase products made from selected countries. The items used by Darling and Arnold are presented in Table 3.15.

Table 3.15
Willingness to Purchase Scale by Darling and Arnold (1988)

No.	Items
1	During my shopping trip, I usually look for products made in _____.
2	Using products made in _____ for comparison purposes helps me to make more intelligent buying decisions.
3	Whenever available, I would prefer to buy products made in _____.
4	In comparison to the products from other countries, as a general rule I have not been very pleased with the products made in _____ that I have purchased.
5	I take great deal of my personal pride in the ownership of products made in ____.

As the study of consumer animosity model was developed by Klein et al. (1998), they adapted and modified the scale developed by Darling and Arnold (1988) to measure the purchase willingness of consumers towards foreign products. To measure the construct, they used a six-item 7-point Likert scale to indicate the respondent's agreement on the statements (on a 1 = strongly disagree to 7 = strongly agree). The details of the items used by Klein et al. (1998) to measure the purchase willingness construct are show in Table 3.16.

Table 3.16
Purchase Willingness Scale by Klein et al. (1998)

No.	Items
1	I would feel guilty if I bought a Japanese product.
2	I would never buy a Japanese car.
3	Whenever possible, I avoid buying Japanese products.
4	Whenever available, I would prefer to buy products made in Japan*.
5	I do not like the idea of owning Japanese products.
6	If two products were equal in quality, but one was from China and one was from Japan, I would pay 10% more for the product from China.

*Note: * – reverse score*

The current study primarily examined consumers' willingness to buy the products coming from the US in general terms and not specific product or brand. To be consistent with most of the studies in consumer animosity this study employed a Likert-type, 7-point scale to capture consumers' willingness to purchase the US made products, (e.g., Ettenson and Klein, 2005; Klein et al., 1998; Shimp et al., 2004; Shoham et al., 2006; Hinck, 2005; and Nijssen and Doughlas, 2004). The scale asked participants to select the range of their willingness to purchase products

made in the US. All six items used by Kline et al. (1998) were included for the current study. The items used to measure the purchase willingness construct for the current study are presented in Table 3.17.

Table 3.17
Purchase Willingness Scale Used in the Current Study

No.	Items	Source
1	I would feel guilty if I would buy a US product.	Klein et al. (1998)
2	I would never buy a US product.	Klein et al. (1998)
3	Whenever possible, I avoid buying US products.	Klein et al. (1998)
4	Whenever available, I would prefer to buy products made in the US*.	Klein et al. (1998)
5	I do not like the idea of owning US products.	Klein et al. (1998)
6	If two products were equal in quality, but one was from US and one was from Malaysia, I would pay 10% more for the product from Malaysia.	Klein et al. (1998)

*Note: * – reverse score*

3.5.7 Measuring the Purchase Action Construct

Purchase willingness construct is relatively well covered in the literature. However, not many researchers focus their study on an actual purchase behaviour construct. One study that tried to measure the purchase of local over foreign products is Granzin and Olsen (1998). In their study, the purchase behaviour construct was measured in terms of six different shopping activities. The construct was measured using 6-point, Likert type scale items anchored by 1 = strongly disagree and 7 = strongly agree. The items used by them are presented in Table 3.18.

Table 3.18
Purchase Action Scale by Granzin and Olsen (1998)

No.	Items
1	Mostly, I try to buy US-made products.
2	Mostly, I try to buy brands of U.S. companies.
3	I take the time to look on labels so I can buy more US-made products.
4	I take the time to look on labels so I can buy more brands of U.S. companies.
5	I shop first at retail stores that make a special effort to sell US-made products.
6	I shop first at retail stores that make a special effort to sell brands of U.S. companies.

Furthermore, in the Pullman et al. (1997) study, in their attempt to explore the effect of responsibility, shopping support and state empathy towards purchase action among US consumers, they specifically used 10 items to measure as the construct. It comprises a set of specific purchase and consumption related activities; e.g., buying a domestic product when a better quality foreign alternative is available. The measures of purchase action were selected to represent a range of available shopping, purchase, and consumption choices, rather than to provide interchangeable, alternative measures of the construct in question.

All the 10 items used by Pullman et al. (1997) are presented in Table 3.19, where, as can be seen, all the items used were specifically developed to measure the action taken by the respondents in making actual purchase decisions. This will allow them to directly measure the actual purchase action of the consumers.

Table 3.19
Purchase Action Construct by Pullman et al. (1997)

No.	Items
1	I chose an American made product when a similar foreign item was available.
2	I bought an American made product when a better quality foreign item was available.
3	I bought an American made product when a cheaper foreign item was available.
4	I specifically asked a salesperson to show me the American made product.
5	I asked a salesperson where an item I liked was made.
6	I explicitly recommended to someone else that he/she purchases only American made items.
7	I left a store because I was mad that they sold too many foreign items.
8	I criticized someone I know for buying foreign products.
9	I voted for a political candidate because they support American products over foreign products.
10	I am not using a gift item because it was foreign made.

For the current study, for the purchase action construct measurement, the items of scale measurement from the study conducted by Pullman et al. (1997) were used. From the original ten items proposed and used by them, only seven items are used for the current study. Three items, i.e., "I asked a salesperson where an item I liked was made", "I voted for a political candidate because they support American products over foreign products" and "I am not using a gift item because it was foreign made" were dropped due to the lack of suitability of the questions from the perspective of the current study.

The items selected were then adapted to fit the Malaysian consumers' perspectives, and a higher score indicates higher purchase of products made in the US. The 7-

point Likert scale type was applied for the respondent to indicate how they react to the questions asked. Table 3.20 shows the details of the items used. Seven points were used to ensure that the number of scale points was consistent with other constructs used in the questionnaire.

Table 3.20
Purchase Action Scale Used in the Current Study

No.	Items	Source
1	I chose US made product when similar products from other countries were available.	Pullman et al. (1997)
2	I bought products made in US when better quality items from other countries were available.	Pullman et al. (1997)
3	I bought US made products even though cheaper items made from other countries were available.	Pullman et al. (1997)
4	I specifically asked a salesperson to show me US made products.	Pullman et al. (1997)
5	I explicitly recommended to someone else that he/she purchased only US made items available in the market.	Pullman et al. (1997)
6	I left a store because I was mad that they sold too many products made in US*.	Pullman et al. (1997)
7	I criticized someone I know for buying US products*.	Pullman et al. (1997)

*Note: * – reverse score*

3.6 Questionnaire Design

This section will explain the issues involving the design of the questionnaire. The issues include the content, format and the structure. A structured questionnaire design was used in this study. The questionnaire will take approximately 20 minutes to complete. Some of the questions, especially on the Muslim religiosity construct,

are quite sensitive to be disclosed. The use of a self-administered questionnaire instead of personal interview may encourage people to participate in this study.

According to Malhotra (2003), the advantages of a questionnaire to collect data are: i) it is simple to administer; ii) the data obtained is reliable; and iii) the coding, analysis and interpretation of data are relatively simple and straightforward. In this questionnaire, basically, most of the questions are fixed alternative questions that require the respondents to select from a predetermined set of responses. A copy of the research instrument used in this study is shown in Appendix 1.

Itemized scales rating, i.e., Likert-type scale was applied to most of the questions in this questionnaire. The advantages of Likert scaling are that it is easy to construct and understand as well as flexible and economical in terms of space (Alreck and Settle, 1995). The 7-point Likert-type scale was applied in this study for all the items used to capture the attitudes of the respondents on the intended measured variables. It can provide the midpoint option for respondents if they are indifferent to the questions. Additionally, Malhotra (2003) mentioned that in order to apply the structural equation modelling or any other sophisticated statistical techniques, seven or nine point numerical scales are recommended. All statements and questions in section 1 use a 7-point Likert scale (1 = strongly disagree, 2 = disagree, 3 = slightly disagree, 4 = neutral, 5 = slightly agree and 6 = agree and 7 = strongly agree).

The questionnaire consisted of 13 pages excluding the cover page. The questionnaire is divided into two sections. Two pages were allocated for the cover page and a letter to the respondents with instructions on answering the questionnaire, and

assurance on the confidentiality of the information supplied. Instructions were clearly and precisely stated on the first page of each section. The instructions will hopefully guide the respondents to answer the questionnaire.

Section A of the questionnaire consisted of 90 statements intended to measure all the seven constructs used in this study. Specifically, the items in Section A are intended to measure the constructs of Muslim religiosity, consumer animosity, consumer ethnocentrism, patriotism, US products' judgment, willingness to purchase products from US and the actual purchase action of US products. The reason why all the questions were put and jumbled up in one section is to prevent bias when the respondents answer the questions. Additionally, it is hoped that it will also encourage the respondents to give more meaningful answers.

Specifically, for the Muslim religiosity construct, 21 questions were asked, 15 questions to measure the consumer animosity, 17 questions for the consumer ethnocentrism construct, 11 questions to measure patriotism construct, 13 questions used for the US product judgment construct, six questions to measure the willingness to purchase US made products and seven questions to measure the purchase action of US made products. As mentioned in the previous paragraph, all the questions in Section A were jumbled up to prevent the respondents from knowing the constructs measured.

Subsequently, the respondents' profiles were in Section B, which in total contained eight demographic questions. In this section, simple demographic questions such as the gender of the respondents, their occupation and their monthly household income

were asked. It will furnish information on the respondents' background, which is important for the findings and analysis part. It is useful in testing possible relationships between demographics factor and other constructs in this study (e.g., the relationship between gender or age group and consumer animosity and consumer ethnocentrism). Specifically, the questions in this section include gender, age, marital status, level of education, occupation, monthly household income, members in household, and in which state of Peninsular Malaysia the respondent lives.

Originally, this questionnaire was developed using the English language because all the adapted questions from previous literature used the English version. A graduate student fluent in English and Malay language translated the questionnaire into Malay, and was translated back into English by an instructor who is a teacher of the language course. In addition, the translation process was closely monitored by the research supervisor. The process of translation and back translation was done to avoid wrong translation and ambiguities of the statements. The translation of the questions into the Malay language was deemed appropriate since the levels of English language proficiency among Malaysians are different. Furthermore, the translation needs to be simple and easily understood by the respondents in order to get more meaningful answers. Both languages are presented in sequences in the questionnaire.

3.7 Pre-Testing

Before proceeding with the data collection, a pre-testing was carried out. The questionnaire was pre-tested on a convenience sample of consumers from multiple

backgrounds of education and holding different positions in organizations from clerical to managerial level, in order to assess the reliability of the main constructs used in this study, and to get feedback concerning understanding, phrasing and design of the questionnaire. Furthermore, the pre-test can also be used to check the face and content validity as well as assuring that the questions are understood and correctly translated into the Malay language.

According to Kaynak and Kara (2002), the pre-test is useful to check the clarity, comprehension and consistency of the questionnaire. In addition, it is also important for the respondents to understand and provide comments on the instructions of the questionnaire. The instructions must be easily understood by the respondents due to the fact that they come from different levels of education.

In January 2006, the questionnaires were distributed to 20 selected respondents. All the respondents in the pre-test were chosen from the same target population as the actual research. Each questionnaire was attached with a small gift as a token of appreciation for the respondents' participation in the pre-test. All the pre-test questionnaires were returned back and the respondents gave a good response to the questionnaire. The test was not used for statistical purposes, and responses from the pre-test were not included in the research findings. In fact, only an initial reliability assessment was conducted using Cronbach's coefficient alpha reliability test.

All the comments, feedback and suggestions from the respondents were taken into consideration. Based on the response of the pre-test data, only slight changes were made to the questionnaire and most of the amendments concerned the translation of

the questionnaire from English to the Malay language. Basically, all the constructs met the reliability requirements. As suggested by Nunnally (1967), in the early stage of the research, reliability in the range of 0.5 to 0.6 is sufficient. Table 3.21 shows a summary of the reliability test and the Cronbach's alpha values of the constructs. All the alpha values for the construct used in this study were above the suggested value of 0.5.

Table 3.21
Cronbach's Alpha Values for the Pre-Test

	Construct	No. of Items	Cronbach's Alpha
1	Muslim Religiosity	21	0.795
2	Consumer Animosity	15	0.871
3	Consumer Ethnocentrism	17	0.910
4	Patriotism	11	0.695
5	US Products Judgment	13	0.775
6	Purchase Willingness	6	0.601
7	Purchase Action	7	0.895

All the questions were retained to proceed with the actual data collection. As the established constructs, consumer ethnocentrism shows a very high internal consistency with a value of the coefficient alpha of above 0.9 value. For the consumer animosity construct, the inclusion of two self-developed items purposely for the current research, i.e., US actions against Muslim prisoners in Guantanamo detention centre annoyed me and US policies are always unfair towards the Muslim world, also show a high level of internal consistency. From the results of the

reliability tests of the pre-test, all the other constructs also exhibit high internal consistency and no items needed to be deleted to improve the coefficient alpha.

3.8 Sampling Design

There are several guidelines that researchers need to take into consideration in order to accomplish the sampling design. This includes the relevant population, sample size, sampling procedure and sampling technique.

3.8.1 The Relevant Population

The target population is the complete group of specific population elements relevant to the research project (Zikmund, 2000). In other words, it is the collection of elements or objects that possess the information sought by the researcher and about which inferences are to be made. This process is very important in answering the question of who should and should not be included in the sample. The population of this study is Muslims in Malaysia, but they must also be those who are familiar with the US made products. If the target respondents cannot distinguish the products that come from the US, Japan, European countries or any other countries in the world, it will affect the answer they give in the questionnaire.

3.8.2 Sample Size

Determining the sample size is complex and involves several considerations. Malhotra (2004) suggests several criteria to be considered to determine the sample

size. The factors that should be considered when determining the sample size include the importance of the decision, the nature of the research, the number of variables, the nature of analysis, sample size used in similar studies, incidence rates, completion rates and resource constraints (Malhotra, 2004). Furthermore, he also suggests that if sophisticated analysis of the data using multivariate techniques is required, the sample size should be large.

As the sample size used by previous researchers varies from 73 (Okechuku, 1994) to 1,721 (Okechuku and Onyemah, 1999) in country-of-origin study; from 244 (Klein et al., 1998) to 477 (Ettenson and Klein, 2005) in consumer animosity study and from 357 (Reardon et al., 2006) to 1,535 (Shimp and Sharma, 1987) in the consumer ethnocentrism studies, there is no specific guideline as far as the selection of the sample size in the international marketing research. The types of respondents were also different from adults, students, shoppers, purchasing managers, farmers, etc., depending on the nature of the studies. Kline (1999) and Hair et al. (2006) argue that the sample of above 200 considered as large. However, Roscoe (1975) suggest that sample size of over 500 would be sufficient in the behavioural research. Similarly, Malhorta (2004) also suggest a minimum of 500 respondents. After considering the factors such as the sample size used by previous research and the data analysis techniques to be used, as suggested by Roscoe (1975) and Malhotra (2004), a minimum of 500 respondents would be employed for the current research.

For the above reason, a total of 1,000 questionnaires were distributed to collect the data. The minimum targeted sample size was set at 500 respondents. This sample size is considered to be feasible as well as being time and cost efficient for the

researcher. A sample size that is too small might affect the generalizability of the results, whereas a sample size that is too large will not be feasible for the researcher to complete the data collection due to time and cost constraints. The respondents must be familiar and know about the products' country of origin so that they can analyze and express their opinion and judgment about the product features and performance. The knowledge of consumers will affect the way they answer the questionnaire.

3.8.3 Sample Procedure

Basically, there are four regions in Peninsular Malaysia, namely, Central region, Southern region, Northern region and East Coast region. In each of the regions, two cities were chosen. These cities are the centre of commercial and business areas and are highly populated areas. Most of the economic activities and most of the major retailing centres are located in these areas. These areas generally represent the higher income group of the population. Additionally, these cities have been chosen because consumers in these areas have the knowledge of foreign made products, because, as the centre of economic and business, the imported products from many countries in the world are available. In this case, area sampling was used. In selecting the respondents in these cities, non-probability instead of probability was used and quota sampling was also used to select the target respondents.

3.8.4 Sampling Technique

In the consumer ethnocentrism studies by Keillor et al. (2001) and Hamin and Elliot (2006), the data in their study was collected through personal face-to-face interviews. In the Keillor study, the data collection was over a three-month period using a quota sampling method. In Hamin and Elliot's study, they focused on the upper socio-economic groups due to the nature of their study, i.e., focused on the use of international airlines as subject products. Reardon et al. (2005) also used personal interviews but instead of using self administered questionnaire. Generally, Reardon et al. (2005) is used a non-probability sample for the data collection.

In Klein et al. (2006), they used the self-administered technique to ensure the anonymity of the respondents. They also used a mix of non-probability snowball method in one area and a technique equivalent to mall-intercept in another area. In another study, Razak et al. (2002) used self-administered questionnaires distributed to undergraduate students using a convenience sampling technique due to the exploratory nature of the study and the easy availability of the sampling units.

Under the animosity construct, a multistage stratified equal probability random sampling method was adopted in the study conducted by Jung et al. (2002) and Ang et al. (2004). They conducted door-to-door personal interviews. Half of the respondents in each country (Korea, Singapore, Indonesia, Malaysia and Thailand) were asked to respond with regards to the US evaluation sample, and the other half to a Japanese evaluation sample. Klein and Ettenson (1999) used data from the 1992

National Election Study (NES) in the US. The target respondents were interviewed and selected using a random probability sampling technique.

Ettenson and Klien (2005), in their longitudinal research, examined an indirect boycott in which Australian consumers refused to purchase French goods because of France's nuclear testing in the South Pacific in 1995. The first study provides a theoretical and empirical analysis of an ongoing protest and the second study was carried out one year after the resolution of the conflict. They used a random sample method and the drop-off/pick-up technique; respondents were contacted in-person at home by trained interviewers and asked to complete the self-administered questionnaire.

After several considerations, a non-probability sampling technique using a quota sampling method was chosen for this study. The area sampling was also chosen for the current research. In the quota sampling method, two-stages are involved. The first stage consists of developing control categories of population elements and in the second stage, sample elements are selected based on convenience or judgment (Malhotra, 2004). The purpose of using quota sampling is to ensure that certain characteristics of the population sample will be represented. It is also to ensure that various subgroups in the population are included in the study.

Zikmund (2000) argues that the major advantages of quota sampling compared to probability sampling are speed of data collection, lower costs and convenience. Furthermore, the use of a quota sampling technique also has advantages that help to overcome problems associated with the generalization of the data collected. One of

the underlying objectives in most research is to obtain data that can be reasonably judged as generalizable. Therefore, study data should be as representative as possible. The use of quota sampling allows the researcher to obtain a database that is representative of the population as a whole along predetermined criteria such as age and gender.

In comparing quota sampling to random sampling, Marsh and Scarbrough (1990) found no significant differences that would represent substantial data biases between respondents. Further, these authors also found that no significant nonresponse biases existed when quota samples were compared to random samples. Previous researchers also recommend using quota sampling based on age and gender to reduce potential biases in the data collection (Sudman, 1980).

In the current study, the samples throughout Peninsular Malaysia were gathered with the use of the quota sampling procedure discussed above in order to ensure that the country samples were reasonably representative of their respective populations. Three criteria were selected as the basis of the quota, i.e., gender, income and geographical location/area. A four clustered area, i.e., Northern, Central, Southern and East Coast were identified and in each cluster, two cities selected. For the gender, approximately 50 percent of male and 50 percent of female respondents were targeted. For the income level, 50 percent of respondents earning a monthly household income of below RM3,000 and 50 percent of respondents earning a monthly income of above RM3,000 were targeted. This is due to the report provided by the Department of Statistics of Malaysia in 2004. In the report of Ninth Malaysia Plan 2006 – 2010, the average of monthly household income in Malaysia is

RM3,022. So that, the current study targeted to get approximately 50 percent of respondents that earned the monthly household income of above average and 50 percent of respondents that earned the monthly household income of below average. Finally, for the geographical location, 20 percent from the Northern area, 30 percent from the Central area, 25 percent from the Southern and East Coast area were targeted.

In each cluster/area, two cities were selected. This is due to the assumption that the consumers from urban areas have a basic knowledge of foreign products (for example, where are the products made). With that knowledge, they will have a clearer direction of how to evaluate the products and to answer the questionnaire. Previous research (Bhuian, 1997; Okechuku and Onyemah, 1999; Zain and Yasin, 1997; Kaynak et al, 2000; Vida and Dmitrovic, 2001; Supphellen and Rittenburg, 2001; Wang and Chen, 2004) also used consumers living in the countries' major cities for the reason that they were expected to be more familiar with foreign products. Consumers' knowledge about a product's origin has been shown to have a significant effect on subsequent product evaluations (e.g. Peterson and Jolibert, 1995; Bilkey and Nes, 1982; Parameswaran and Pisharodi, 1994; Piron, 2000; Reardon et al., 2005).

3.9 Data Collection Technique

The data gathered was mainly from survey. The process was carried out in several stages ranging from the development of the questionnaire, identification of states, areas and location for questionnaire distribution in Peninsular Malaysia, distribution

of questionnaire to selected respondents, reminding the research assistants/enumerators and respondents of questionnaire deadline, gathering of questionnaire and analyzing of data.

The enumerators were also hired to help the researcher collect the data. The enumerators passed out the questionnaire and gave oral instructions along with written instructions to the respondents. The data in this study was collected for about a three month period. It was a self-administered and drop-off method of survey where no personal interview was involved but short briefing to the respondents about the questionnaire was conducted. Enumerators would drop-off the questionnaires, give an explanation and verbal instruction and finally, pick them up later. This would give the respondents time to answer the questionnaire. This self-administered and drop-off survey method is more efficient for the researcher due to time and cost constraints. During the period, the researcher also reminded the enumerators about the deadline to return the questionnaires. In addition, the researcher also gave a contact number and e-mail address in the questionnaires in case the respondents needed further clarification on the questions and the study itself.

According to Sood and Nasu (1995), people or respondents are quite reluctant to participate in research relating to religious issues. This can contribute to the lowness of the response rate of the research. Another potential problem is the respondents might be biased in answering the question related to Muslim religiosity. To avoid the problem the questionnaire was attached with an envelope. The respondents were required to seal the envelope before returning the questionnaire (return the sealed

envelope containing the answered questionnaire). This will ensure the anonymity of the respondents.

The questionnaire takes approximately twenty minutes to complete. Respondents were assured of their anonymity and that all the responses would be kept strictly confidential. Training for the enumerators took place at the early stage of the data collection in each area (North, South, and East Coast) and was conducted by the researcher. For the data collection in Central area, it was conducted by the researcher himself.

Two enumerators were appointed in each area and each of them were given 125 sets of questionnaires. Data was collected for 12 consecutive weeks (three weeks for each area) from March to May 2006. The reason why the data collection process was not done concurrently in all areas was because the areas of the data collection were diverse and far away. Focusing on one area made it easier for the researcher to monitor and supervise the enumerators during the data collection process. The distribution of the questionnaires occurred for three days at each site. Then, the questionnaires were left with the respondents for approximately two weeks before they were collected by the enumerators.

The sample was drawn from residents of urban areas in each state selected for the data collection. The areas selected were Kota Bahru in Kelantan, Kuala Terengganu in Terengganu, Kangar in Perlis, Alor Setar in Kedah, Kuala Lumpur, Petaling Jaya in Selangor, Bandar Melaka in Melaka and Johor Bahru in Johor. The urban and suburban consumers were believed to be more knowledgeable about foreign-made

products and most retail centres are located in these towns. Most of the retail centres offer a wide variety of international products and brands where the potential respondents reside. The availability of products in the market will directly increase the knowledge of consumers and it will also increase a wide alternative of products for the consumers to choose. This is very important because if the target respondents do not have this knowledge, they will not be able to evaluate the products coming from a foreign country.

Based on the pre-test results presented earlier in this chapter, at the beginning of March 2006 the researcher began the data collection. It began with the East Coast region, followed by the Southern, Northern and lastly Central, to ensure that the activity of the enumerators could be observed during the distribution and the collection of the survey questionnaires. As to the sampling strategy, several considerations were taken into account to set the criteria that serve the research objectives and are manageable in the given theoretical framework. The next paragraphs describe the considerations leading to the sampling criteria:

- First of all, the selection of the regional area that was set in the study is a typical regional area that has been widely used in Malaysia. Furthermore, states from East Malaysia, i.e., Sabah and Sarawak were excluded from the study due to the cost and time constraints. If the scope of this research covered all the states in Malaysia, the data collection period would take a longer time and the costs incurred would be much higher.

- For the age criteria, the participants must be eighteen years or older, so that the research would not involve any minors, as could happen in some of the college-based student research.

- Furthermore, the respondents must be Muslims because the study focuses on the Muslim religiosity that includes the faith, the concept of *Wajib* (obligatory), *Musta'hab* (not obligatory but liked), *Mubah* (allowed; neither liked nor disliked), *Makrooh* (not prohibited but disliked) and *Haraam* (prohibited) which are not applicable to consumers from other religions. So that, all the respondents for the current study are Muslims.

- The study was also intended to look at the different level of socio-economic status of the respondents. In this case, the respondents from the Southern and Klang Valley area are expected to have a higher income, be more urbanized and be greatly influenced by the western lifestyle and their religiosity might be less. As a comparison, respondents from the Northern and East Coast are expected to be more traditional, conservative and basically more religious. The majority of the population in those areas are Malays, compared to the South and Central area where the population is mixed between Malay, Chinese and Indian where the Chinese and Indians are basically non-Muslims. When there is a mix of Muslims and Non-Muslims, the things that are prohibited in Islam (such as liquor and night clubs) are also widely available in that particular area.

Typical for research in this region as well as many other parts of Asia (Ahmed and d'Astous, 1999; Malhotra et al., 1996), attached with all questionnaires distributed – as a token of appreciation for the time spent – the participants were handed a small present. It was planned to provide a gift of very limited commercial value, yet of some relevance for the respondents. The token of appreciation given to the participants was actually a handkerchief or face towel for the male respondents and a face towel or potpourri for female respondents. The value of these presents was clearly more of a gesture and honorarium for time spent than an inducement to participate.

3.10 Sampling Result

As discussed in the previous chapter on methodology, the sample for the study was conducted among Malaysian Muslim consumers in four areas in Peninsular Malaysia, i.e., Northern, Central, Southern and East Coast. There are two basic sampling techniques, i.e., probability and non-probability sampling. For the current research, non-probability sampling using the quota and area sampling were employed. The next paragraph is the discussion on the sampling results, which includes the response rate.

Data were collected for 12 consecutive weeks (three weeks for each area) from March to May 2006. The questionnaire distribution was done stage by stage. First, the questionnaires were distributed to the East Coast of Peninsular Malaysia, which contains the state of Kelantan and Terengganu, followed by Southern region (Johor

and Melaka), Northern region (Kedah and Perlis) and lastly, Klang Valley or Central region (Kuala Lumpur and Selangor).

Table 3.22 briefly summarized the total of questionnaires distributed and the total of usable questionnaires.

Table 3.22
Response Rate

Item	Descriptions	Percent (%)
Total questionnaires distributed	1000	100.0
Total questionnaires returned	710	71.0
Less: Non-usable		
i: Incomplete	31	3.1
ii: Inconsistent	16	1.6
Total usable questionnaires	663	66.3
Non-response	290	29.0

By the end of May 2006, the data collection process was completed. As shown in Table 3.22, out of the 1,000 questionnaires distributed, 710 were received back within the period of twelve weeks, making it about 71 percent response rate. Only 663 of the respondents completed the entire questionnaire. A total of 31 of the returned responses were grossly incomplete while another 16 had the indication of inconsistencies in the responses. In many cases, the inconsistencies can be clearly seen in the way they answered the negatively worded questions. They might strongly

agree with the positive worded questions and at the same time, they also strongly agreed with the negative worded questions. For the incomplete questionnaires, most of the incomplete sections come from the questions about the religiosity and the demographic profile. After deducting the unusable questionnaires, only 663 questionnaires were coded and could be used for further data analysis. As such, the usable rate for the questionnaire was 66.3 percent.

3.11 Data Screening and Cleaning

Before proceeding further, it is essential to check the data set for errors. Mistakes might occur during data entry and this can affect the analysis. Besides, some analysis can also be very sensitive to what is known as outliers, the values that are well below or above the other scores (Pallant, 2005). According to Hair et al., (2006), outliers are observations with unique combinations of characteristics that are identifiable as distinctly different from other observations. It can be either an unusually high or low value on a variable. In this case, it could arise from procedural error, such as data entry error or a mistake in coding. These outliers should be identified during the data cleaning stage and they should be eliminated or recorded as missing values (Hair et al., 2006).

3.11.1 Checking for Errors and Missing Data

It is crucial to spend some time for checking for mistakes at the early stage of data analysis rather than facing bigger problems later. In this research, while entering the data collected and answered by the respondents in the questionnaires, errors in

entering the value of the data could occur. The values might fall outside the range of possible values for a variable. In this case, after entering the data in the database using the SPSS software, detection of the errors for the categorical variable (demographic variables) and continuous variables (e.g. consumer ethnocentrism and consumer animosity) was made.

The categorical variable, from the SPSS menu, can be checked by using the "analyze menu", "descriptive statistics", "checking for the frequencies" and "show the minimum and maximum value" for each categorical variable. If there is a score lying outside the possible range for that particular variable, then it must be cleaned. For the current study, after checking for error, all the entered values were within the range for all variables. For the continuous variables, in this study, the variables were measured using a seven-point Likert scale. The procedure is almost the same with the categorical variables except the researcher adds the mean value for each variable. After running the analysis, the results showed that the values of the data all fell inside the intended range.

It is also important to detect the missing data in the questionnaires. Missing data is a situation where the valid values on one or more variables are not available. For the missing data, since all of the data is keyed into SPSS manually, it is detected during that particular process. As explained earlier, 31 questionnaires were detected as incomplete and had missing data, and all of these questionnaires were excluded from further analysis. According to Malhotra (2004), for the treatment of the unsatisfactory responses, there are several ways that can be used by the researchers; a) returning to the field to get better data; b) assigning missing values; and c)

discarding the unsatisfactory responses. In this study, as explained earlier, the researcher has chosen the third method, which is to remove them from further analysis.

3.11.2 Checking for Outliers

There are four classes of outliers; a) data entry error or mistake in coding; b) outliers due to extraordinary event; c) extraordinary observations for which the researcher has no explanation; and d) observations that fall out of the ordinary range of values on each of the variables (Hair et al., 2006). In this research, the kind of outliers the researcher intended to examine is the first type of outliers, which arise from a procedural error, i.e., data entry error or mistake in coding.

For this study the extreme points used in measuring the variables, other than demographic variables, are one and seven. Any numbers outside this range are considered as the outliers. After running the necessary tool in checking for the outliers, the results showed that no outlier was detected from the data entered in the database. So, no action was taken for the treatment of outliers as it is not applicable in this case.

3.11.3 Data Manipulation

After all the data had been entered and the data file has been checked for accuracy, the next step involved manipulating the raw data into a form that could be used to conduct analysis. The negatively worded items are presented in Table 3.23.

Table 3.23
Negatively Worded Items

Construct	Item No.	Items
Patriotism	34	In general, I have very little respect for the Malaysian people.
	41	It is not that important for me to serve my country.
	51	It is not constructive for one to develop an emotional attachment to one's country.
	59	It bothers me to see children made to pledge allegiance to the flag or sing the national anthem or otherwise induced to adopt such strong patriotic attitudes.
Product Judgment	20	Products made in US are generally of a lower quality than similar products available from other countries.
	28	Products made in US are produced by firms that are more concerned with the outward appearance of the products than with the products performance.
	53	Products made in US seem to be more in the nature of luxury items than necessary items.
	67	Products made in US generally lack creativity and are very imitative of products made in other countries.
Purchase Willingness	6	If two products were equal in quality, but one was from US and one was from Malaysia, I would pay 10% more for the product from Malaysia.
	70	I do not like the idea of owning US products.
	82	Whenever possible, I avoid buying US products.
	86	I would never buy a US product.
	90	I would feel guilty if I would buy a US product.
Purchase Action	24	I criticized someone I know for buying US products.
	29	I left a store because I was mad that they sold too many products made in the US.

For the interval scale (seven-point Likert scale in this case), all the negatively worded items or questions were reversed. The negatively worded items were reversed before calculating the total scores for the variables. Therefore, the range of seven-point Likert scale for the negatively worded items was transformed from 1 (Strongly Disagree) – 7 (Strongly Agree) to 1 (Strongly Agree) – 7 (Strongly

Disagree). Then the data was explored using the descriptive method once again to check for normality, validity, reliability and so forth. Then, the total score for each variable was calculated.

3.12 Data Analysis Technique

This section describes in detail the specific procedures for analyzing the data collected from the respondents. Computer software – Statistical Package for the Social Sciences (SPSS) Version 12.0.1 and Analysis of Moment Structure (AMOS) Version 5.0.1 were used for data analyses. This software will be used in this study for data management and analysis. They are powerful software that allows the analysis of the most widely used types of statistical technique. Data analysis included two types of analyses. The first part describes the methods used for any initial analyses performed to investigate and prepare for the main analyses. The second part describes analysis methods used to investigate the main research questions, which were stated in testable hypotheses.

In terms of the hypotheses testing, 12 main hypotheses were tested in this study. All the hypotheses were tested using structural equation modelling (SEM). The first three hypotheses (H1, H2 and H3) are intended to examine the effects of Muslim religiosity on the consumer animosity, consumer ethnocentrism and patriotism among Muslim consumers in Malaysia. Hypothesis 4, Hypothesis 5 and Hypothesis 6 are intended to examine the impact of consumer animosity, consumer ethnocentrism and patriotism on the judgment of products from the US. Furthermore, Hypotheses 7 and 8 are intended to examine the impact of the US

product judgment on the purchase willingness and purchase action of US made products, respectively.

In addition, for Hypothesis 9, it will examine the effects of purchase willingness of US made products on the consumers' purchase action of US made products. Hypotheses 10a, 10b and 10c, test the mediating effect of the US product judgment in the relationship between consumer animosity, consumer ethnocentrism and patriotism and purchase willingness of the US made products. Meanwhile, Hypotheses 11a, 11b and 11c test the mediating effect of the US product judgment in the relationship between consumer animosity, consumer ethnocentrism and patriotism and the purchase action of the US made products. Finally, Hypothesis 12 was developed to test the mediating effect of purchase willingness in the relationship between product judgment and purchase action (please refer Figure 3.1).

3.12.1 Internal Consistency Reliability Analysis

The initial analyses included the calculation of descriptive statistics (such as means, standard deviations, and distribution frequency) for each of the variables measured. Assessments of the scales' internal consistency reliability and validity were conducted using Cronbach's alpha. Internal consistency reliability is used to assess the reliability of the summated scale, where several items are summed to form a total score and each item measures some aspect of the construct measured by the entire scale, and the items should be consistent in what they indicate about the characteristic (Malhotra, 2004). For a scale to be internally consistent, the items of

that scale must be highly inter-correlated. The most widely used measure of internal consistency reliability is Cronbach's coefficient alpha.

Cronbach's coefficient alpha values vary between 0.00 and 1.00. However, there is no general agreement as to what constitutes good or very good levels of Cronbach's alpha. Gabel (1986) suggests that alpha coefficients in the high 0.80 or above should be considered good, and Cortina (1993) suggests that alpha coefficients 0.85 or above are quite good. But, it is also important to know that Nunnally (1967) suggests that the alpha values that are above 0.5 can be considered as adequate.

3.12.2 Independent Sample T-Test and One-way ANOVA

For the test of mean difference among the groups of respondents in the studied variables, the statistical analyses to be applied will be the mean comparison analysis, which falls under the category of bivariate analysis. The independent sample t-test and one-way ANOVA will determine the correlations between the demographic profile and other continuous variables used in this study, i.e., Muslim religiosity, consumer animosity, consumer ethnocentrism, patriotism, US products judgment, purchase willingness of US made products, and purchase action of US made products. Furthermore, the independent sample t-test and one-way analysis of variance (ANOVA) were used to determine if the perception means vary among different demographic characteristics, at least at the 95 percent confidence level.

The independent sample t-test is used to compare the mean score of two different groups of people or conditions, or in other words to compare the mean score on

some continuous variables for two different groups of subjects (Pallant, 2005). In this study only one demographic variable, i.e., gender have two different groups in one variable, therefore, the test of mean differences used was the independent sample t-test.

The one-way ANOVA is used when there is one independent (grouping) variable with three or more levels (groups) and one dependent continuous variable. The significant value was equal to or less than 0.05. The post-hoc test will tell you exactly where the differences among the groups occur (Pallant, 2005). One-way ANOVA tests were conducted on other demographic variables, which were geographical area, age, education level, marital status, occupations, income level and number of household members.

3.12.3 Factor Analysis

According to Sekaran (2000), factor analysis is a multivariate technique, which would confirm the dimensions of the concept that have been operationally defined, as well as indicate which of the items are most appropriate for each dimension. Factor analysis will basically help the researcher to reduce a vast number of variables to a meaningful and manageable set of factors. Similarly, Zikmund (2003) explains that the general purpose of factor analysis is to summarize the information contained in a large number of variables into a smaller number of factors. He added that the general goal of factor analysis is data parsimony, i.e., to reduce a large number of variables to as few dimensions or constructs as possible. With the small number of factors, further analyses are easier to perform and the patterns of

associations between variables assist the researcher to establish the interrelationship of variables that belong together (Hair, Black, Babin, Anderson and Tatham, 2006).

Generally, there are two types of factor analysis, i.e., exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). EFA explores the data and provides the researcher with information about how many factors are needed to best represent the data (Hair et al., 2006). On the other hand, CFA is similar to EFA in some respects, but philosophically it is quite different as in CFA the researcher must specify both the number of factors that exist within a set of variables and which factor each variable will load highly on before results can be computed (Hair et al., 2006). In other words, CFA procedures do not assign variables to factors and researchers must be able to assign variables into factors before the results can be obtained. According to Pallant (2005), EFA is often used in the early stages of research to gather information about the interrelationships among a set of variables, while CFA, on the other hand, is a more complex and sophisticated set of techniques used later in the research process to test (confirm) specific hypotheses or theories concerning the structure underlying a set of variables.

For the current study, both EFA and CFA will be used. The EFA was used to reduce the number of items in the factors, as well as to remove all the cross loaded items and to ensure that the items are loaded into intended variables. Subsequently, the reliability test will be conducted for all the factors extracted from the factor analysis to examine the internal consistency of the variables. After that, the CFA procedures will be applied before proceeding to the hypotheses testing. The next subsection will

discuss the EFA. An explanation of CFA will be discussed in detail in the SEM section.

a. Exploratory Factor Analysis

As explained earlier, EFA is a technique for data exploration and to determine the structure of factors to be analyzed. It is used to establish dimensionality of the relationship between items and constructs. In addition, it is also used in this study to identify the structure among the set of variables. The EFA was performed for all the constructs included in this study, i.e., Muslim religiosity, consumer animosity, consumer ethnocentrism, patriotism, product judgment, purchase willingness and purchase action. Besides determining the dimensionality, the objective of doing factor analysis in this study is to identify representative variables and to create new variables, if any are to be used in the subsequent analysis.

The first step in applying factor analysis is the assessment of the suitability of the data where two main issues are considered, i.e., sample size and the strength of relationship among the variables or items (Pallant, 2001). Generally, a larger sample size is better. For the strength of inter-correlations among the items, coefficient values of above 0.3 are preferable (Tabachnick and Fidell, 2001). If the majority of the coefficients are below 0.3, factor analysis may not be appropriate. Furthermore, the measure of sampling adequacy uses the Bartlett's Test of Sphericity (Bartlett's Test) and Kaiser-Mayer-Olkin (KMO), which are also important to assess the factorability of the data. The Bartlett's Test should be significant ($p<0.05$) for the factor analysis to be considered appropriate and the measure of sampling adequacy

produces the KMO index that ranges from 0 to 1, and indicates that KMO more than 0.60 are considered appropriate for factor analysis (Pallant, 2001).

Furthermore, the rotation called orthogonal rotation method was applied in this study, and the Varimax with Kaiser Normalization was used to analyze the scales. Varimax orthogonal rotation was favoured since it minimizes correlation across factors and maximizes within the factors (Loehlin, 1998). The Varimax orthogonal rotation also minimizes the number of variables with high loadings on a factor, thereby enhancing the interpretability of the factors (Malhotra, 2004). Orthogonal rotation is also the preferred method when the research goal is data reduction to either a smaller number of variables or a set of uncorrelated measures for subsequent use in other multivariate techniques (Hair et al., 2006).

Another important consideration of factor analysis is the factor loading. Factor loading indicates the degree of correspondence between the variable and the factor i.e., the correlation of each variable and the factor (Hair et al., 2006). In other words, factor loading indicates the strength of the relationship between the items and the factor with higher loading making the items representative of the factor. Nunnally (1978) posits that items with loadings higher than 0.50 indicate a reasonable loading. Hair et al. (2006) asserts that factor loadings of above 0.5 are considered practically significant. Therefore, in this study, only items with loadings higher than 0.50 on one factor were retained for further analysis.

3.12.4 Correlation Analysis

The most popular technique that indicates the relationship of one variable to another is simple correlation analysis (Zikmund, 2000). Zikmund (2000) added, a correlation indicates both the magnitude of the linear relationship and the direction of the relationship. According to Malhotra (2004), correlation analysis is the most widely used statistic in summarizing the strength of association between two metric (interval or ratio scaled) variables. It indicates the degree to which the variation in one variable, for example X, is related to the variation in another, for example, Y. It is also known as the Pearson correlation coefficient, simple correlation, bivariate correlation, or merely the correlation coefficient (Malhotra, 2004). Basically, correlation analysis was carried out to determine the degree of association and indication of multicollinearity between all variables used in this study. Additionally, it not only indicates the degree of association of the variables but the direction of such association as well. Further explanation will be presented in the Pearson correlation analysis in the next chapter.

3.12.5 Structural Equation Modelling

As explained earlier, all the hypotheses were tested using SEM via AMOS software. SEM is a commonly used statistical method for quantifying the relationships among variables that cannot be observed directly. It 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 and

Mentzer, 1999). In this study, the main purpose of using SEM is to test the proposed theoretical model and to investigate relationships between variables in the model.

Furthermore, SEM is a comprehensive statistical approach to testing hypotheses about relations among observed and latent variables. Latent variable is an unobserved concept that can be represented by observable or measurable variables (Hair et al., 2006). Measured variables are sometimes referred to as manifest variables or indicators. The SEM has the advantage over standard regression analysis of explicitly considering the measurement error in the observed variables while simultaneously estimating a system of structural equations (Hoyle and Panter, 1995). Furthermore, unlike the traditional statistical methods that can only examine a single relationship at a time, the SEM method greatly expanded the researchers' capability to study a set of interrelated relationships simultaneously.

Effect size is commonly used to compliment SEM because it is a large-sample technique. It is generally understood among statisticians that SEM requires large sample sizes (Kline, 1999). According to Hair et al. (2006), SEM in general, requires a larger sample relative to other multivariate approaches and some of the statistical algorithms using SEM programmes are unreliable with a small sample. For example, Hair et al. (2006) explained that SEM has been found to provide valid results with sample sizes as small as 50, but the recommended minimum sample size is 200, which provides a sound basis for estimation. When the test of a relationship deals with a large sample size, effect size helps researchers to differentiate between statistical significance and practical significance. However, Kline (1998) argues that with less than 100 cases, almost any type of SEM analysis is untenable unless a very

simple model is evaluated. A sample size less than 100 could be considered small, 100 – 200 subjects as a medium sample size and sample sizes that exceed 200 cases could be considered large.

Furthermore, MacCallum et al. (2001) argue that the opinion among researchers regarding the minimum sample size varies depending on the analysis procedure and the model characteristics. Hair et al. (2006), propose several suggestions in determining the sample size needed to perform SEM analysis based on the model complexity and the basic measurement model characteristics. In the current study, based on the model complexity and sample size, it can be concluded that the SEM technique can be used for further analysis. It will be a two-step procedure, i.e., the measurement model relating the observed indicators to the latent variables, and the underlying structural model expressing a relationship among the unobserved variables (Hair et al., 2006).

a. Confirmatory Factor Analysis / Measurement Model

The majority of SEM researchers advocate the two step approach (Garver and Mentzer, 1999). In the first step, the researcher validates the measurement models through confirmatory factor analysis, which includes the test for construct validity by testing construct unidimensionality, reliability, convergent validity and discriminant validity (Garver and Mentzer, 1999). Once the measurement model is validated, the second step, estimating the structural relationships between latent variables, is conducted, to test the hypotheses between the variables included in the

study. Further explanation of the structural model will be presented in the next chapter.

As recommended by Hoyle and Panter (1995), Anderson and Gerbing (1988), Medsker, Williams and Holahan (1994), Garver and Mentzer (1999) and Novak, Hoffman and Yung (2000), the current study adopted a two-step analytic procedure. First, the measurement models were evaluated prior to the estimation of the structural models, so as to prevent measurement misspecifications from being misinterpreted as misspecifications of the structural models (Burt, 1976). Second, a structural model, a set of one or more dependence relationships linking the hypothesized model's constructs, are most useful in representing the interrelationships of variables between constructs (Hair et al., 2006).

A measurement model is a sub-model of structural equation modelling that specifies the manifest variables (i.e., statements, items, indicators or questions) for each latent variable (i.e., construct, factor, or dimension) and assesses the reliability of each construct for estimating the casual relationship (Hair et al., 2006). Since the underlying structure of the measurements was clearly established and most of the items of each scale were adapted from previous research, confirmatory factor analysis (CFA) was employed for these analyses. Basically, for the current study, CFA was used to confirm that the indicators sort themselves into factors corresponding to how the researcher has linked the indicators to latent variables as well as to examine convergent and discriminant validity.

During the CFA process, researchers may wish to examine possible modifications to improve the theoretical explanation or to improve the goodness-of-fit of the model. If the measurement model possesses an unacceptable fit, standardized residual and modification indices can help the researcher determine why the model is unacceptable. Nevertheless, when examining standardized residuals and modification indices, theoretical considerations should always be used as the primary consideration in making model modifications (Garver and Mentzer, 1999). Generally, removal of problematic items and re-specification may result in a better fit of a model (Bollen, 1989).

Standardized residuals are important in diagnosing problems with a measurement model. Residuals refer to the individual differences between observed covariance terms and the fitted covariance terms. Standardized residuals are calculated through the raw residuals divided by the standard error of the residuals. Typically, a value less than 2.5 does not suggest a problem, conversely, a value greater than 4.0 raises a red flag and suggests a potentially unacceptable degree of error (Hair et al., 2006). The most likely response is to drop one of the items associated.

According to Garver and Mentzer (1999), in examining standardized residuals, patterns of large residuals should be taken into consideration. A large residual will be over 2.00 and 2.58, and is considered as statistically significant at the 0.05 level (Garver and Mentzer, 1999). A significant residual indicates a substantial prediction error for a pair of indicators. Those items with cross-loading or corresponding to more than one factor will show large residuals with different items from different factors and should be deleted from the model. If the modification is implemented,

the model should then be re-specified and re-evaluated after each modification (Garver and Mentzer, 1999).

For the modification indices (MI), it is calculated for every possible relationship that is not free to be estimated. MI are very helpful in determining how to modify the measurement model. A substantial modification index value of 7.88 is considered to be a significant model improvement (Garver and Mentzer, 1999), but Hair et al., (2006) recommended that modification indices of approximately 4 or greater will improve the model significantly by freeing that particular corresponding path. The largest MI indicates the greatest improvement in fit and these items should be assessed for modification first, if and only if, the modification is consistent with a priori theory or can be interpreted substantively (Bryne, 2001). Similar to standardized residual modification, the model should be re-evaluated after each re-specification through MI (Garver and Mentzer, 1999).

The re-evaluation of such items includes the assessment of their standardized loadings (i.e., standardized regression weights). In AMOS software, the standardized structural coefficients are labelled standardized regression weights. The standardized regression weights are similar to the coefficients used to test the strength of relationships. If the standardized loading of an item is below 0.50, the cut off value for factor loadings recommended by Stevens (1996), the item might be omitted.

According to Hair et al. (2006), the size of factor loadings is one of the important considerations; high loadings on a factor would indicate that they converge on some common point. At a minimum, all factor loadings should be statistically significant

(Anderson and Gerbing, 1988). Because a significant could still be fairly weak in strength, Hair et al. (2006) proposed that a good rule of thumb is that standardized loading estimates should be 0.5 or higher, and ideally 0.7 or higher.

The relative strength of the model's effect was tested through a number of fit measurements. There is no single statistical test that best describes the strength of the model, therefore, researchers have developed a number of goodness-of-fit (GOF) measures that, when used in combination, assess the results from various perspectives (Hair et al., 2006). The analysis of fit provides standards to determine the degree to which the presumed relationships are similar to the observed data (Hair et al., 1998). There are three different types of fit measurements, i.e., absolute fit measure, incremental fit measure, and parsimonious fit measure (Hair et al., 2006).

In general there are two strategies to evaluate overall model fit: (i) selecting fit indices that represent different families of fit indices and (ii) specifying a stringent criteria and selecting fit indices that best represent this criteria (Garver and Mentzer, 1999). Although a number of fit indices are available to evaluate the overall model fit, there is little consensus regarding the best index to be used or which index performs better under different conditions. According to Hair et al. (2006) and Bentler (1990), the proposed model has to illustrate a satisfactory fit in terms of absolute fit, incremental fit and model parsimony. Therefore, the model tested needs to show satisfactory fit for all three different groups of fit measurement, and then the model can be considered as fit. Model fit means that the hypothesized model fits the data well.

Absolute fit measures determine the overall goodness-of-fit for the measurement model, although, there is no distinction as to whether the model is preferred or inadequate (Hair et al., 1998). However, incremental fit measures determine the goodness-of-fit of the actual model to a null model (Hair et al., 1998). Finally, parsimony fit indices refer to the application of parameters or the coefficient of hypothesized model. The fewer the estimated parameters used in the model, the more parsimonious the model (Hair et al., 1998). In determining which fit measurements are going to be used to assess the fitness of the model, the current study adhered to Smith and McMillan's (2001) suggestion to avoid using only textbook criteria (i.e., numerical cutoffs) for evaluating fit measurements and involved multiple sources for evaluating the model.

The most fundamental measure of overall fit is the chi-square statistic ($\chi 2$). Low values, which result in significance levels greater than 0.05, indicate that the actual and predicted input matrices are not statistically different, hence, a good fit. However, the $\chi 2$ measure is often criticized for its over-sensitivity to sample size, especially in cases where the sample size exceeds 250 respondents (Hair et. al., 2006). Furthermore, they argue that when your sample size is more than 250 and the variables you use are more or equal to 12, significant p-values can be expected. As sample size increases, this measure has a greater tendency to indicate significant differences for equivalent models.

The specific fit measurements and criteria used in this study are:

1. Absolute fit indices:

a. Goodness of Fit (GFI) is a non-statistical measure ranging in value from 0 (poor model fit) to 1 (perfect model fit) and it represents the overall degree of fit (the squared residuals from prediction compared with the actual data) but it is not adjusted for the degrees of freedom (Hair et al., 2006). Smith and McMillan (2001) recommend a value of 0.90 or greater as an appropriate cutoff for an adequate model, however cutoff values of 0.92 and 0.95 have been considered to produce a more desirable model fit (Bollen and Long, 1993) and 0.98 is considered to be an excellent model fit (Fassinger, 1987). GFI is considered to be one of the important measures of absolute fit (Kline 1998; Hair et al., 2006).

b. Root Mean Square Error of Approximation (RMSEA) is the discrepancy per degree of freedom of the average of the residuals between the observed and estimated input matrices (Hair et al., 2006). In other words, it is also known as an index that measures the discrepancy between the observed and estimated covariance matrices per degree of freedom (Garver and Mentzer, 1999). The values run on a continuum from 0 to 1. Regardless of the sample size, a cut off value of < 0.08 is recommended by Funk, Ridinger, and Moorman, (2004), to reduce the number of errors. Meanwhile, Garver and Mentzer (1999) suggest that values falling between 0.05 and 0.08 are deemed acceptable. Browne and Cudeck (1993), imply that the RMSEA value for a model less 0.05 is considered to be an indicator of a “good” fit, while values from 0.05 to 0.08 suggest a “fair” fit. Furthermore, values ranging from 0.08

to 0.10 indicate a mediocre fit, and those greater than 0.10 indicate poor fit (MacCallum, Roznowski, and Necowitz, 1992).

2. Incremental fit indices:

a. Comparative Fit Index (CFI) is an incremental fit index that is an improved version of Normed Fit Index (NFI), according to Hu and Bentler (1999). The NFI is not used in this study in determining the fitness of the model. The CFI was developed as a non-centrality parameter-based index to overcome the limitation of sample size effect (Garver and Mentzer, 1999). The CFI is normed so that values range between 0 to 1, with higher values indicating a better fit and it is among the most widely used indices (Hair et. al., 2006). CFI values less than 0.90 are not usually associated with a model that fit well. Garver and Mentzer (1999) also suggest that a CFI value above 0.9 represents an acceptable fit.

b. Tucker-Lewis Index (TLI) predates the CFI and is conceptually similar in that it also involves a mathematical comparison of a specified theoretical measurement model and a baseline null model (Hair et al., 2006). It is not normed and thus its value can fall below 0 or above 1. The TLI is also known as Nonnormed Fit Index (NNFI), (Garver and Mentzer, 1999). TLI also seems resilient against variations in sample size and, thus, is highly recommended (Marsh et. al., 1988). Garver and Mentzer recommend a TLI value of 0.90 or greater, with a good fit having values that approach 1.

Higher values reflect that the models are a better fit compared to the models with a lower value.

3. Parsimony fit indices:

a. The most widely used measure of parsimony fit is the Adjusted Goodness-of-Fit Index (AGFI) provided by AMOS. AGFI is an extension of GFI but adjusted by the ratio of degrees of freedom for the proposed model to the degrees of freedom for the null model. A recommended acceptance value of AGFI is 0.90 or greater (Hair et al., 2006; and Segars and Grover, 1993).

Table 3.24 summarizes all the fit indices used in the measurement model of the current study. The reason why those indices were chosen for this study is due to the fact that those indices have been used in numerous studies and have proven to be stable across various situations and have always been suggested by previous researchers and literature. According to Garver and Mentzer (1999), TLI, CFI and RMSEA are the indices scaled on a pre-set continuum (0 to 1) for easy interpretation and are all relatively independent of sample size effects. The other two indices (GFI and AGFI) were used because the original study in the consumer animosity model conducted by Klein et al. (1998) used these indices to check on the model fitness.

After achieving the desired fitness requirement of the goodness-of-fit of the measurement model and fulfil the validity (i.e., convergent and discriminant validity) of the measurement model, and the model is now validated, the next step is to estimate the structural relationship (regression or path analysis) between latent

variables (Medsker et al., 1994) and in this stage the theoretical model can be tested (Anderson and Gerbing, 1988) and hypotheses can be tested as well.

Table 3.24
Summary of Fit Indices

Indices	Abbrev.	Acceptable Level	Comments
Chi-Square	(χ2) (df, p)	p > 0.05 at α = 0.05	P > 0.05 reflects acceptable fit; 0.1 reflects a good fit.
Normed Chi-Square	(χ2) /df	1.0 < (χ2) /df < 3.0	Values close to 1 indicate good fit but values less than 1 may indicate overfit.
Goodness of Fit index	GFI	GFI > 0.90	Values between 0.90 – 0.95 indicate satisfactory fit and values higher than 0.95 indicate good fit.
Adjusted Goodness of Fit index	AGFI	AGFI > 0.90	Values between 0.90 – 0.95 indicate satisfactory fit and values higher than 0.95 indicate good fit.
Tucker-Lewis Index	TLI	TLI > 0.90	Values between 0.90 – 0.95 indicate satisfactory fit and values higher than 0.95 indicate good fit. Values greater than 1 indicate overfit.
Comparative Fit Index	CFI	CFI > 0.90	Values between 0.90 – 0.95 indicate satisfactory fit and values higher than 0.95 indicate good fit. Values close to 0 indicate poor fit, CFI =1 indicates perfect fit.
Root Mean Square of Approximation	RMSEA	RMSEA < 0.08	Values between 0.05 – 0.08 indicates satisfactory fit. Value 0 indicates a perfect fit.

Source: Adapted from Schumacker and Lomax (1996); Kline (1998); Garver and Mentzer (1999); and Hair et al. (2006).

3.13 Research Hypotheses

After discussing the previous results and literature in the preceding chapters as well as developing several research propositions based on the constructs included in this study, a number of hypotheses were developed in order to investigate the relationship among the constructs. These hypotheses will then be tested using the statistical techniques mentioned earlier. Specifically, in testing the hypotheses, the SEM technique was utilized using the AMOS software.

The hypotheses of the study are as below:

Hypothesis 1 – The higher the Muslim religiosity, the higher will be the Muslim consumer animosity towards the US.

Hypothesis 2 – The higher the Muslim religiosity, the higher will be the Muslim consumer ethnocentrism.

Hypothesis 3 – The higher the Muslim religiosity, the higher will be the level of patriotism.

Hypothesis 4 – There is a negative relationship between consumer animosity and US product judgment.

Hypothesis 5 – There is a negative relationship between consumer ethnocentrism and US product judgment.

Hypothesis 6 – There is a negative relationship between patriotism and US product judgment.

Hypothesis 7 – There is a positive relationship between US product judgment and purchase willingness of US made products.

Hypothesis 8 – There is a positive relationship between US product judgment and purchase action of US made products.

Hypothesis 9 – There is a positive relationship between purchase willingness and purchase action of US made products.

Hypothesis 10a – US product judgment mediates the relationship between consumer animosity and purchase willingness of US made products.

Hypothesis 10b – US product judgment mediates the relationship between consumer ethnocentrism and purchase willingness of US made products.

Hypothesis 10c – US product judgment mediates the relationship between patriotism and purchase willingness of US made products.

Hypothesis 11a – US product judgment mediates the relationship between consumer animosity and purchase action of US made products.

Hypothesis 11b – US product judgment mediates the relationship between consumer ethnocentrism and purchase action of US made products.

Hypothesis 11c – US product judgment mediates the relationship between patriotism and purchase action of US made products.

Hypothesis 12 – Purchase willingness mediates the relationship between US product judgment and purchase action of US made products.

Figure 3.1 shows the proposed conceptual framework of the current study with the designated hypotheses.

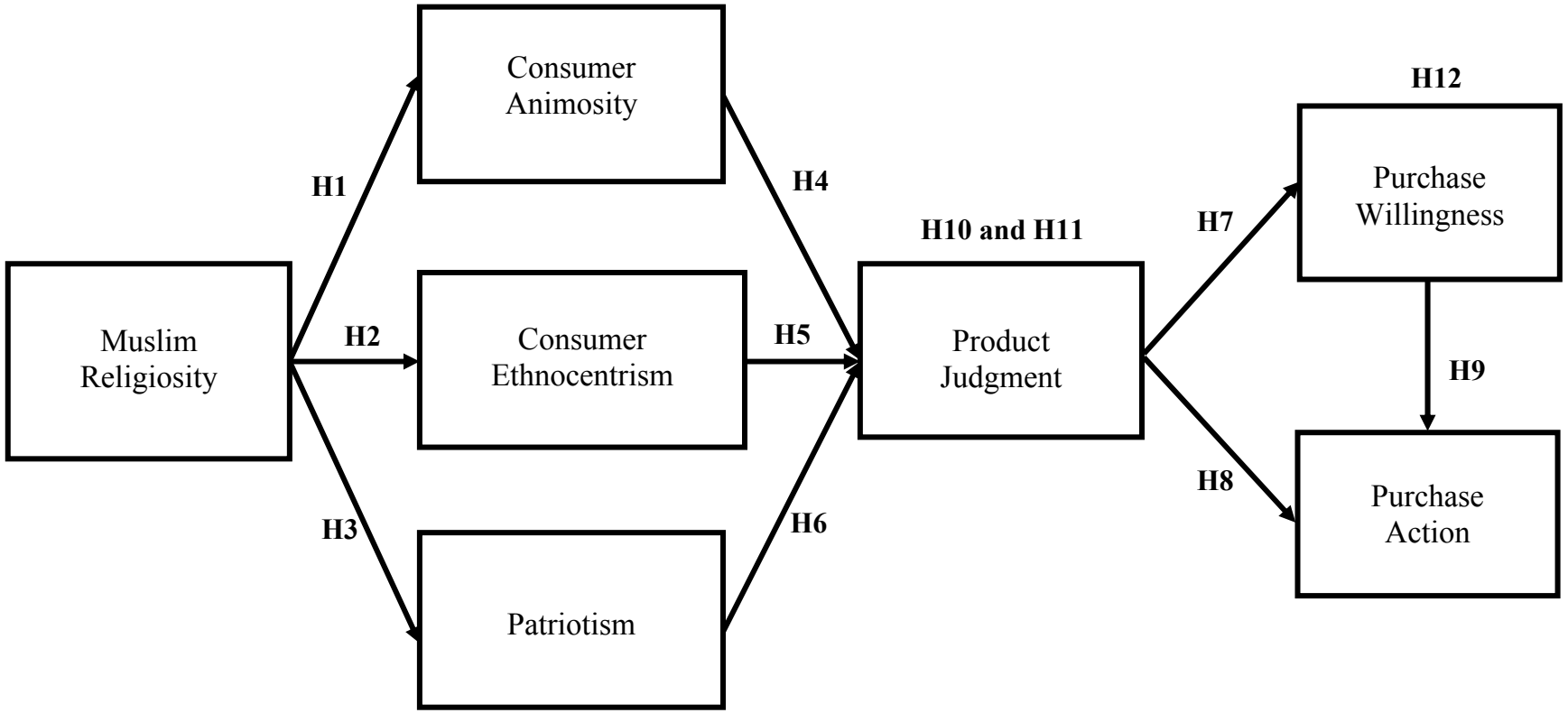

Figure 3.1: Framework of Study with the Designated Hypotheses

3.14 Conclusion

Basically, this chapter outlines the research methodology for the current study. The discussion includes several aspects such as the research design, product and country selection, measurement of construct, questionnaire design, sampling technique, data collection technique as well as the data analysis techniques. The study focuses on a survey of the attitudes of Malaysian Muslim consumers towards US made products based on the animosity model of foreign products purchase. In total, seven main variables, i.e., Muslim religiosity, consumer animosity, consumer ethnocentrism, patriotism, product judgment, purchase willingness and purchase action were discussed, especially on the measurement items of the construct.

Furthermore, this chapter also discussed the validity and reliability assessment to ensure the validity and reliability of the scale used in the research. It details the tests used to examine the validity and reliability of each construct in which the methods of assessment including the exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) were explained in detail. For reliability assessment, Cronbach's coefficient alpha was used. Finally, the discussion on the structural equation modelling was also presented, as it is the primary technique used in the hypothesis testing.