Chapter 6
Data Gathering Methods

Rather than believing that one must choose to align with one paradigm or the other, I advocate a paradigm of choices. A paradigm of choices rejects methodological orthodoxy in favor of methodological appropriateness as the primary criterion for judging methodological quality. The issue then becomes not whether one has uniformly adhered to prescribed canons of either logical-positivism or phenomenology but whether one has made sensible methods decisions given the purpose of the inquiry, the questions being investigated, and the resources available. The paradigm of choices recognizes that different methods are appropriate for different situations. Situational responsiveness means designing a study that is appropriate for a specific inquiry (Patton 1990 p.39).

6. Introduction
Chapter 5 provided the research objectives and the hypothesised models. The models form the basis of the empirical component of the study and direct the development of an appropriate research design. This chapter discusses the design of the data gathering methods employed to facilitate the empirical assessment of the hypothesised models. The research design is the framework or plan for collecting and analysing the data and has been developed from consideration of the extant literature, the research objectives and the hypothesised models. This chapter focuses only on the design of the data gathering methods, methodological issues related to the analysis rationale and procedures are discussed in the following chapter (7). The following sections describe the procedures adopted to design and implement the data gathering stages of the study.
6.1 Introduction to the research plan & procedures

The hypothesised models developed from the literature review and theoretical development presented in chapters 3 and 4 will be empirically examined by surveying a sample of individuals for a selected product. The intent of this research is to hypothesise and test in detail the dimensions of the constructs and the nature of the interrelated effects operating among the variables/constructs under appropriate conditions. This process will enable a test of whether a relevant set of observed data are consistent with the proposed theoretical processes and relationships.

The data gathering methods employed in this study consisted of two empirical stages. Stage 1 was a preliminary empirical stage to generate survey items and select the most appropriate measures for the key individual difference variables hypothesised to relate to involvement and pretest the resulting questionnaire. The second empirical stage, stage 2, consisted of the administration of the final questionnaire and the full-scale testing of the hypothesised models. The two-stage empirical plan for the study is presented within the dashed lined box in Figure 6.1 and is placed within the context of the previous literature review and model development/hypothesis building stages.

Figure 6.1 Research Process Model
6.2 Instrument development: Design rationale & stages

As indicated by Cook and Campbell (1979), any social or behavioural research investigation must be designed with certain practical trade-offs between external and internal validity in mind. These trade-offs should be based on the purpose or goals of the research. As noted earlier, this research is concerned with assessing relationships between the four hypothesised forms of involvement and establishing nomological validity of these types of involvement with respect to consumer's values system, self-image product-image congruency, consumer product knowledge and expertise, consumer confidence and consumption consequences.

This study can be classified under what Caller, Phillips and Tybout (1981) called a 'theory application' as opposed to an 'effects classification'. Under the effects classification, concern focuses on the application, whereas with the theory application, the concern focuses on testing theoretical propositions or establishing relationships (Calder, Phillips & Tybout 1981 and Ratz & More 1989).

This study may be philosophically classified as pure research. The key feature of pure research is that it is intended to lead to theoretical developments and progress (Easterby-Smith, Thorpe and Lowe 1991). There are fundamentally, three forms that theoretical progress may take; 1) theoretical research or inquiry is about discovery (the popular view); 2) when an outcome of the research is an innovation (a new process, idea or technique, theory or measurement instrument that solves a particular problem); and 3) an outcome of reflection where an existing theory, technique or idea such as involvement is re-examined, possibly in a different context or different underlying content (Easterby-Smith, Thorpe and Lowe 1991).

The present study focuses in part on all 3 forms of theoretical progress. This study, focuses on discovering relationships between the four forms of involvement and values system, product knowledge and expertise, self-image product-image congruency and consumption consequences. It also focuses on a reconceptualisation and reconstruction of the construct of involvement.
This section will discuss the underlying rationale and issues covered in the process of designing and administering the instrument to gather the primary data for the study. This will be followed by step by step details of the actual construction and testing of the instrument.

The survey design process included the following considerations:

1. preliminary considerations
2. question content
3. question wording
4. response format
5. question sequence
6. physical layout and characteristics
7. pretest

These seven steps are adopted from Tull & Hawkins (1990) and are further discussed below using the model of questionnaire construction process.

The questionnaire was designed on the premise that a sound instrument requires applying appropriate principles, common sense, concern for the respondent, a clear concept of the required information and a pretesting process (Tull & Hawkins 1990). Figure 6.2 outlines the major decisions and questions that were addressed in the design process following the model recommended by Tull and Hawkins (1990).
Figure 6.2 Questionnaire Construction Decisions

1. Preliminary decisions
   . Exactly what information is required?
   . Exactly who are the target respondents?
   . What method of communication will be used to reach respondents?

2. Decisions about question content
   . Is the question really needed?
   . Is the question sufficient to generate the needed information?
   . Can the respondent answer the question correctly?
   . Will the respondent answer the question correctly?
   . Are there any external events that might bias the response to the question?

3. Decisions concerning question phrasing
   . Do the words used have but one meaning to all the respondents?
   . Are any of the words or phrases loaded or leading in any way?
   . Are there any implied alternatives in the question?
   . Are there any unstated assumptions related to the question?
   . Will the respondents approach the question from the same frame of reference desired by the research?

4. Decisions about the response format
   . Can this question best be asked as an open ended, multiple choice or dichotomous question?

5. Decisions concerning the question sequence
   . Are the questions organised in a logical manner that avoids introducing errors?

6. Decisions concerning layout of the questionnaire
   . Is the questionnaire designed in a manner to avoid confusion and minimize recording errors?

7. Pretest and revise
   . Has the final questionnaire been subjected to a thorough pretest?

(Source: Tull & Hawkins 1990 p.288)

The model of the design process recommended by Tull and Hawkins (1990) is reproduced in full to show the reader the design process followed to ensure a sound questionnaire was developed. The specific design process and the important methodological issues covered will be detailed here to provide the reader with an understanding of the development process for the final administered survey.
In developing and evaluating the survey instrument, a number of guidelines and procedures recommended by Bearden, Netemeyer and Mobley (1993) were followed in conjunction with Tull and Hawkins model to ensure that the instrument was as psychometrically sound as possible to meet the objectives of the study. Other sources were also consulted to provide a sound basis for the instrument development procedures employed. Such sources included psychometric properties of scale development by American Psychological Association (1985), Bohrnstedt and Borgatta (1981), Carmines and Zeller (1979), Churchill (1979), De Vellis (1991), Nunnally (1979), Peter (1979, 1981), Robinson, Shaver and Wrightsman (1991), Tull and Hawkins (1990) and Zikmund (1990).

The first procedure adopted was to base the measures on solid theoretical conceptualisations with the construct domains thoroughly delineated and outlined. Taking this issue into consideration, the theoretical propositions regarding construct definitions were fundamentally based on what was included in the domain of each construct and what was excluded. This provided clear boundaries for the domain of each construct.

Another important criterion was that any scales adopted from the existing literature as well as the scales developed specifically for this study should exhibit both content and face validity. The scales were required to be operationally consistent with the theoretical domain of the construct, particularly so with respect to the actual development of the involvement scale section of the instrument. A large number of items were generated to tap the domain of the involvement construct and these were further screened and culled using judges with expertise in the consumer behaviour literature (Churchill 1979; Robinson, Shaver & Wrightsman 1991 and Zaichkowsky 1984). An additional practical criterion was to develop items that were as short as possible to ensure ease of interpretation and response by respondents without undue sacrifice of reliability or validity (Carmines & Zeller 1979; Churchill 1979 and Churchill & Peter 1984).
6.2.1 Step 1: Preliminary decisions

Using the objectives of the study and the hypothesised models, the broad composition of the questionnaire was to include items which measure the following domains: consumer involvement, values, product knowledge, confidence, self-image product-image congruency and consumption consequences within the context of purchase and consumption of a product. From these concepts and dimensions, an extensive list of questions was developed and existing instruments evaluated. These questions and instruments were broken down and categorised under a number of dimensions applicable to each construct.

Also the target audience for the survey was identified. The target audience was broadly identified as undergraduate and postgraduate students. Since many of the intended respondents were off-campus students a Mail Survey was selected as the best method for administering the final survey.

6.2.2 Step 2: Question content: Generation of item pool

Empirical stage 1 was undertaken to accomplish three major objectives; 1) to generate items which operationalised the involvement conceptualisation based on the literature, from in-depth interviews with a small number of individuals from those who constituted the final sample and from the researcher and then refine and reduce if necessary, through a process of expert judging of the generated items, 2) to develop and test items and scales for the measurement of the constructs captured in the nomological net, namely involvement with self-image/product-image congruency, materialism, values, confidence, knowledge/expertise, consumption consequences and consumption values, and 3) to assemble and pilot test the preliminary version of the survey instrument that would be used in the full study to test the hypothesised models. The procedures for item pool generation and assessment are detailed below and the exact flow of the instrument development is depicted in Figure 6.3. The procedures adopted are in line with the recommendations of Converse and Presser (1986) regarding iterative development and pretesting and those of Tull and Hawkins (1990).
6.2.3 Generation of items from the literature and in-depth interviews

As identified in chapter 2, many measures of involvement have been proposed since the concept's introduction to marketing by Krugman in the mid 1960's. Unfortunately, most of the previous measures of involvement have met with significant criticism due to their underlying weakness in theory development and psychometric rigor and validation (e.g., comments by: Bloch 1981; Goldsmith, Emmert & Hofacker 1991; Jain & Srinivasan 1990; Seitz, Kappelman & Massey 1993; McQuarrie & Munson 1992; Mittal 1992 and 1995 and Poiesz & de Bont 1995). In reviewing the existing measures for the present research, it was concluded that no existing involvement measure taken as a whole was appropriate for tapping involvement as conceptualised in this study. This meant that the involvement section of the survey had to be designed from scratch.
As indicated in Figure 6.3 step 1 was the item generation process. The generation of the initial list of items was achieved by analysing the previously published measures on involvement to generate items, generating items from the researcher’s own perspective and conducting a small number of in-depth interviews with students to generate items that tapped involvement as conceptualised in this study. Seventy items were found in the literature that appeared to tap involvement as conceived in this study and were brought into an initial pool of ninety items that were developed by the researcher.

As part of the item generation process a small number (4) of in-depth interviews were conducted. These were semi-structured interviews and took approximately 25 to 30 minutes each. Each interviewee was asked to talk about fashion clothing, computers, music and their favorite product and least liked product. The approach adopted was that the introduction to the interview and questions should make the respondent aware of the purpose of the research and the interview and its format and also, importantly, make the respondent feel comfortable with the researcher.

The interviews began with broader questions to provide a background to the specific research objectives and establish the respondents views of the products. Probing was used in situations where the respondent provided answers that were not clearly understood and when they provided information that appeared contradictory to other responses given. The interview format and questions allowed the interviewer to probe for more detailed information if the respondent showed in-depth knowledge or revealed critical information, whilst not affecting the interview flow. The interviews were analysed for content and context that related to involvement. From these interviews a list of items were generated and submitted to the item pool. The in-depth interviews resulted in another ninety-nine items being developed to tap involvement as conceptualised in chapter 3.

Utilising the large number of items generated and grouping them into concepts and dimensions that tapped the four forms of involvement presented in chapter 3, a review process was undertaken to delete questions that were not necessary given the goals the study. Items that would not generate the required information, that were double barreled and/or that were not believed to be answerable by respondents from the target population were deleted.
Drawing from the literature on involvement and using the researcher's own items and the preliminary in-depth interviews, an initial pool of 259 items was developed for submission to expert judges as detailed in Figure 6.3.

6.2.4 Expert judges assessment number 1

The next step in the questionnaire development consisted of having the generated items assessed by expert judges for their content and face validity. The items generated through the literature analysis, by the researcher and in-depth interviews were pooled and submitted to expert judges for content validation and item reduction (Andrews 1985; French & Michael 1966 and Zaichkowsky 1984). This process involved an initial analysis by judges for deletion of unrepresentative items in line with the appropriate meaning of the terms and definitions and then a second for finer judging (Zaichkowsky 1984).

Evaluating the content validity of a test for a particular purpose is the same as subjectively recognising the adequacy of the definition (French & Michael 1966 and Zaichkowsky 1984). The content and face validity of the items was examined iteratively in two successive passes through the item list, first for initial deletion of recognisably poor items and then again for more rigorous refinement of the items which remained as recommended by Zaichkowsky (1984). Initially, three judges with expertise in the consumer behaviour literature rated each item as to whether the item tapped involvement as represented by the definition and overview of the concept provided by the researcher in written form.

According to the instructions given by the researcher, the three judges rated each item using a dichotomous categorisation process, as either representative of an involvement dimension or not representative of an involvement dimension, for involvement in products, purchase decisions, communications and consumption. As such the initial ratings were based on the item being either representative or not representative of the four forms of involvement as indicated in the first column of Table 6.1. The first assessment reduced the number of items from 259 to 177. The remaining items were then submitted to expert judging assessment number 2.
6.2.5 Expert judges assessment number 2

The second judging performed at step 3 of Figure 6.3 used five expert judges with expertise in the consumer behaviour literature to assess the items remaining from the first assessment. Judges were again provided with a written definition of the construct, a rating scale and instructions for completing the task. Judges rated each item by categorising items as being either clearly representative, somewhat representative or not representative of involvement as used by Zaichkowsky (1984) and indicated in column two of Table 6.1. To ensure a more refined judging in this second assessment the scale for evaluating items was increased from two point to three points. The criterion used for retaining an item was 80% agreement by the judging panel for item retention in the pretest instrument. That is, those items evaluated as being clearly representative or somewhat representative of involvement by at least four out of the five judges were included in the pretest instrument. This process reduced the item list from 177 retained after the initial judging to 139 at the conclusion of the second judging as indicated in the second column of Table 6.1.

<table>
<thead>
<tr>
<th>Table 6.1 Expert Judging Procedure and Stages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessment No. 1 by 3 Expert Judges</td>
</tr>
<tr>
<td>Category Rating Used</td>
</tr>
<tr>
<td>Representative of involvement</td>
</tr>
<tr>
<td>Not representative of involvement</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Number Administered 259</td>
</tr>
<tr>
<td>Number Retained 177</td>
</tr>
</tbody>
</table>

6.3 Step 3: Question phrasing

Following the step of reducing the number of items carried out in step 2, the retained items were examined for multiple meanings, inherent ambiguity, doubled barreled interpretation, phrasing bias and implicit assumptions about respondents knowledge. Those items identified as having one or more these problems were rewritten.
6.4 Step 4: Response format: Number of scale points chosen

The objective of this step was to determine the most appropriate type of response format for the questionnaire. Given that the questionnaire was to be self-administered via mail, the number of questions to be asked, the topic, phrasing and the purpose, a closed response format was chosen and an appropriate scaling format was selected. Thus, this step required the questions to be constructed and phrased with the response format in mind.

Given the outline of the previous steps, the most appropriate scale for the questionnaire items was the forced response Likert-type format (Emory & Cooper 1991 and Zikmund 1991). This scale response format was chosen after considering the views expressed by Hughes (1969), where he argued that the final selection of scale response formats needs to take into account the objectives of the study, the respondents’ familiarity with the product and data processing considerations. Further, the response format was also related to the product selection. As the respondents were believed to be familiar the product, the forced response format was considered not problematic (Hughes 1969; Parasuraman 1986 and Zikmund 1991).

In line with the views of Parasuraman (1986) and Zikmund (1991), a forced-response was considered more suitable to avoid the tendency by respondents to select mid-points on scales. The consideration of forced-response was also seen as appropriate for the product category being studied. Therefore the product category to be studied and scale type were interrelated decisions.

It was also believed that the benefits offered by the Likert-type format outweighed the disadvantages for this particular questionnaire instrument and its administration. The advantages sought were ease of construction and administration. Also, the instructions accompanying the scale were required to be relatively easily understood in terms of the mail survey. The Likert-type scale is known to offer such benefits and therefore the selection was based on information requirements and goals of the survey, the ease of development and administration for the respondents and cost (Emory & Cooper 1991 and Tull & Hawkins 1990).
A six-point Likert-type scale was chosen because it was believed that a five point and below would have been too coarse and a seven point or higher would have imposed on respondents the burden of making distinctions which were too fine (Mittal 1982). It was also decided to keep a six-point scale consistent throughout the instrument in order to reduce the respondent’s necessity of continually adjusting cognitive mindset and re-focusing their attention on changing number of scale points. The fact that respondents were required to make judgments on a large number of items made this imperative even more significant.

Churchill and Peter (1984, 1995) highlighted a further consideration in regard to scale response formats. They indicated that scales for which all rating points are labeled have higher reliability than scales where only the end rating points (poles) are labeled. Taking this into consideration, it was decided to label all scale points in the following fashion.

* Fashion Clothing is important to me.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Moderately Disagree</th>
<th>Slightly Disagree</th>
<th>Slightly Agree</th>
<th>Moderately Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Consideration was also given to the ability of the scale type to be able to be analysed using structural equation modeling techniques (Brom & Cooper 1991 and Tull & Hawkins 1990). Likert-type scales have also been used extensively in questionnaires for gathering data on constructs similar to those being investigated in this study and analysed via structural equation modeling procedures in consumer behaviour (e.g. Goldsmith, Emmert & Hofacker 1991; Hair et al 1995; Mittal & Lee 1989 and Uncles & Manaresi 1996).

6.5 Step 5: Question sequence

Given the objectives and nature of the survey, question sequence was considered important, not only to reduce possible errors, but encourage respondents to complete the survey and return it. Tull and Hawkins (1990) and Zikmund (1991) provided a number of general guidelines that aided in reducing the probability of generating measurement error caused by the sequence of the questions. The questions were listed from the more general or broader issues to specific product-focused questions. On this issue Tull and Hawkins (1990) argue that the first questions should be simple, objective and interesting. The overall objective for the questionnaire was that it should move from domain to domain in a logical manner. The general rules or guidelines set out by Tull and Hawkins (1990) were followed by starting
with the more general and basic questions that would give an understanding of respondents’
student profile and basic values and progressively focusing on the respondents’ views
regarding the product.

6.6 Step 6: Questionnaire layout

The issue of the questionnaire layout was of particular concern to the researcher. Firstly,
the issue of questionnaire length or actual number of questions asked was potentially a
problem. This problem was addressed by attempting to keep the number of pages to a
minimum in order to make the questionnaire look smaller (Zikmund 1991). This created the
problem of spacing the questions so that the questionnaire did not look overcrowded. This
was achieved by single-line spacing and reducing the font size of the scale and then leaving
a line between each question and question block (Zikmund 1991). The general instructions
for completing the questionnaire were given in the covering page and brief instructions to
focus the respondent’s attention were given at the beginning of each new section, so as to
avoid continually repeating the general instructions throughout the questionnaire.

To keep the instrument as attractive as possible and reduce confusion, the same scale and
layout was used throughout the questionnaire. The point of continuity in the layout was
also considered important to increase the response rate and also to reduce possible errors in
recording by respondents (Tull and Hawkins 1990).

6.7 Step 7: Pretest and revise: Final retained items for pretest

There is very little guidance about pretesting questionnaires. Converse and
Presser (1986, pp.52-75) recommend iterative pretesting with small sample
sizes. . . . After the questionnaire has been revised and evaluated by experts,
Converse and Presser recommend that a “polishing” pretest should be
administered to members of the target population (Bolton 1993 p281).

Items representing product involvement, purchase decision involvement, communications
involvement and consumption involvement were placed in the pretest questionnaire along
with the items contained in the materialism scale (Richins & Dawson 1992), the LOV scale (Kahle 1983) and items tapping product knowledge/expertise, consumption consequences by type and number, consumer confidence and self-image/product-image congruency.

The survey instrument was pretested to ensure that the content and format was clear and logical, the questions were not ambiguous and it could be completed in about 30 minutes. The pretest was conducted using a convenience sample of undergraduate students undertaking a unit in buyer behaviour. This tactic provided scope for debriefing pilot respondents at the completion of the pretest administration. The pretest sample consisted of 58 students who completed the questionnaire for the product running/sport shoes. It took, on average, approximately 30-40 minutes for each student to complete the pilot instrument.

6.7.1 Discussion of pretest: Pretest analysis outcomes

Responses for the pretest instrument were analysed using SPSS cluster analysis, bivariate Pearson inter-item correlation analysis, item-total correlation analysis and coefficient alpha reliability analysis. An initial hierarchical cluster analysis was performed on items focusing on each construct to determine if items within each construct clustered into theoretically and empirically meaningful clusters. Pearson correlation analysis and coefficient alpha reliability analysis were performed with the objective of identifying problem items in order to reduce the number of items to a more manageable and parsimonious set of items tapping involvement and the other constructs. Items within clusters were examined for substantive inter-item and item-total correlations. Items with average item correlations below 0.40 or above 0.80 (multicollinearity) were initially deleted. This criteria was adopted from Guildford (1956) and Zaichkowsky (1984) and others who argued these values represent meaningful guides for item retention or deletion.

As indicated by Hair, Anderson, Tatham and Black (1995), cluster analysis is a useful approach to scale development in the pretest stage. Cluster analysis was also valuable in guiding the deletion of items that did not cluster into theoretically meaningful groups, that is single item clusters. Another criteria for the pretest instrument when analysed via
correlations and Cronbach alpha analysis was to determine if this study's pretest instrument achieved similar results to that reported in the literature. Table 6.2 provides an overview of the number of items submitted to pretest and the number retained in the final survey instrument.

Table 6.2 Number of Items for Each Construct Represented in the Survey

<table>
<thead>
<tr>
<th>Concept</th>
<th>Number administered in Pretest</th>
<th>Contained in final questionnaire</th>
</tr>
</thead>
<tbody>
<tr>
<td>Product Involvement</td>
<td>48</td>
<td>25</td>
</tr>
<tr>
<td>Purchase Decision Involvement</td>
<td>34</td>
<td>24</td>
</tr>
<tr>
<td>Communications Involvement</td>
<td>28</td>
<td>18</td>
</tr>
<tr>
<td>Consumption Involvement</td>
<td>29</td>
<td>22</td>
</tr>
<tr>
<td>Global Values</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>Consumption Values</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>Materialistic Values</td>
<td>18</td>
<td>15</td>
</tr>
<tr>
<td>Confidence</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>Self/Product Image Congruency</td>
<td>20</td>
<td>6</td>
</tr>
<tr>
<td>Product knowledge/expertise</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Consequences by Type</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Consequences by Number</td>
<td>5</td>
<td>3</td>
</tr>
</tbody>
</table>

The final questionnaire, shown in Appendix A, was also used as a focal point within a focus group setting. The focus group consisted of five students undertaking a unit in marketing management. The focus group was used as a procedure for checking items and scale formats, order of sections, wording and presentation for the questionnaire. Following analysis of the focus group feedback, only minor alterations were made to the questionnaire. The alterations included issues related to grammar, cover page format and the movement of the self-image product-image section to last section of the questionnaire.

A supplementary process carried out in step 4 to achieve a sound instrument was to use two expert judges for the examination of each step outlined above. Two academics, with skill and experience in questionnaire design, cooperated with the researcher by reviewing the instrument at crucial design stages and then the final draft was reviewed by them again. At each review stage, the researcher took into consideration the comments made by the academics regarding question content, phrasing, format, sequence and layout. The reviewers gave valuable contributions regarding question bias and wording, the scale format
and the number of questions and grouping and any possible misunderstanding that respondents may have with the instrument.

This process of testing, using an expert judgment system, was designed to detect conceptual or operational flaws (Jaworski & Kohli 1993 and Johnson, Black & Sakano 1993). This phase of the pretest also required the academics to critically evaluate the items from the standpoint of the domain representativeness, item specificity and clarity of construction. Based on these critiques, a very small number of items were revised to improve their specificity and precision. This process was adopted by Jaworski and Kohli (1993) in their research on market orientation and has been shown to be valuable in developing a sound instrument.

Table 6.3 provides a summary schemata for the evaluation of the pretest instrument following the outcomes of the pilot test data analysis and subsequent refinement of the final instrument using the procedures discussed above.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Cronbach Alpha Reported in Literature</th>
<th>Cronbach Alpha achieved in pretest</th>
<th>Expert judges Assessment of scale</th>
<th>Focus group assessment of scale</th>
<th>Revision or modification of scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Product Involvement</td>
<td>.67 to .97</td>
<td>.96</td>
<td>Acceptable</td>
<td>Appropriate</td>
<td>Nil</td>
</tr>
<tr>
<td>Purchase Decision</td>
<td>.77 to .93</td>
<td>.95</td>
<td>Acceptable</td>
<td>Appropriate</td>
<td>Minor</td>
</tr>
<tr>
<td>Communications</td>
<td>.68 to .95</td>
<td>.93</td>
<td>Acceptable</td>
<td>Appropriate</td>
<td>Very Minor</td>
</tr>
<tr>
<td>Consumption Involvement</td>
<td>.78 to .83</td>
<td>.92</td>
<td>Acceptable</td>
<td>Appropriate</td>
<td>Very Minor</td>
</tr>
<tr>
<td>Global Values</td>
<td>.58 &amp; above</td>
<td>.79</td>
<td>Acceptable</td>
<td>Appropriate</td>
<td>Nil</td>
</tr>
<tr>
<td>Consumption Values</td>
<td>Not applicable</td>
<td>.81</td>
<td>Acceptable</td>
<td>Appropriate</td>
<td>Nil</td>
</tr>
<tr>
<td>Materialistic Values</td>
<td>.71 to .83</td>
<td>.6 to .82</td>
<td>Acceptable</td>
<td>Appropriate</td>
<td>Minor</td>
</tr>
<tr>
<td>Consumer confidence</td>
<td>not applicable</td>
<td>.89</td>
<td>Acceptable</td>
<td>Appropriate</td>
<td>Nil</td>
</tr>
<tr>
<td>Product knowledge</td>
<td>not applicable</td>
<td>.95</td>
<td>Acceptable</td>
<td>Appropriate</td>
<td>Minor</td>
</tr>
<tr>
<td>Self-image product-image</td>
<td>.82 to .91</td>
<td>.88</td>
<td>Acceptable</td>
<td>Acceptable</td>
<td>Moderate</td>
</tr>
<tr>
<td>Consumption Consequences</td>
<td>not applicable</td>
<td>.77</td>
<td>Acceptable</td>
<td>Appropriate</td>
<td>Nil</td>
</tr>
</tbody>
</table>

As indicated above following analysis of the focus group feedback, only very minor or moderate alterations were made to some sections of the questionnaire. The minor to
moderate alterations included issues related to grammar, cover page format and the movement of the self-image product-image section to last section of the questionnaire.

The final instrument also contained items on the respondent’s age, gender and student status (part-time, full-time, undergraduate and post-graduate, internal (on-campus) and external (off-campus) and mode of study. Items 1 to 5 of the questionnaire reported in Appendix A relate to the respondents basic student and demographic characteristics.

6.7.2 Summary

An instrument or scale is valid if it measures what it is seeking to measure and not some other factor or situation (Zikmund 1991 and Tull & Hawkins 1990). Validity is concerned with the development of adequate operational measures for the concepts being investigated and whether the measure correctly represents the concept of study (Emory & Cooper 1991 and Hair et al 1995). To achieve content and face validity, this study adopted the process of developing the constructs to be investigated in chapter 3 and 4, using the processes advocated by Tull and Hawkins (1990), Bearden, Netemeyer and Mobley (1993) and Converse and Presser (1986). Reliability is concerned with whether a variable or set of variables in an instrument is consistent in what it is intended to measure (Emory & Cooper 1991 and Hair et al 1995). To ensure validity and reliability a systematic set of procedures were developed and carefully followed to ensure a sound data gathering and measurement process was developed to operationalise the constructs as presented in chapter 3 and 4.

Efforts to determine if the scales used in the research were measuring what they were supposed to be measuring acknowledges the potentially serious impact of measurement error on empirical results (Cote & Buckley 1988). However, it must be remembered that validity is a relative, transient, descriptive concept rather than a yes or no proposition (Nunnally 1967). The stages in the research design went some way to ensuring that the instrument validly measured what it purported to measure.

Pretesting is only one means of exploring issues which will impact on the development of the instrument (Sieber 1973). This study used an iterative process that included the use of
experts to alleviate any potential subjectivity or bias problems by the researcher in developing and testing the instrument. Often a good deal of exploratory work precedes even the pretest of a questionnaire, and as a rule, the more knowledgeable the researcher is about their population the more sophisticated and smoother the administration of the survey will be (Seiber 1973). This was a key guiding philosophy in the design of the instrument to gather the primary data.

Other key issues addressed in the design of the instrument were ways to increase the response rate. This was done through adhering to a number of general principles as set out by Tull and Hawkins (1990) and Zikmund (1991). The general process was that the mail questionnaire was most appropriate because of the cost, respondent convenience and geographic flexibility it offered. Therefore, in order to obtain a rate of response that was considered adequate, the following methods were employed. The length of the questionnaire was kept as short as practical (Zikmund 1991) and a covering letter from the researcher were used to outline the importance of the study, enlist the respondent’s assistance and detail the value of their input. Also, survey sponsorship was believed to be a key inducing factor and, in that light, a supporting letter from the Head of the Department of Marketing and Management accompanied each questionnaire (Sieber 1973 and Zikmund 1991). The use of university stationary to add weight to the research was adopted as a strategy to encourage respondents to complete the questionnaire and a return paid envelope was also included with each questionnaire (Sekaran 1991).

This section has dealt with the issues of validity and reliability of the survey instrument and how validity and reliability objectives were achieved in the design process. The above has detailed how the instrument was developed through following a systematic and methodologically sound process. The purposiveness and rigor of this stage, achieved through the design and testing procedures adopted, ensured the validity and reliability of the instrument. Further, the design and administration relied on systematic and replicable procedures.
6.8 Discussion of final questionnaire content

The following section contains a brief discussion of each section of the final questionnaire that forms the survey instrument used in empirical stage 2 of this study and as such fulfills the desires of Rothschild (1984) when he encouraged researchers to:

_Lets go collect data on interesting aspects of involvement_ (p.217).

**Global Values Measure**

Values were measured by the Kahle (1983) list of values (LOV) measure. The LOV approach to values was chosen to measure values, because much of consumer behaviour has been argued to be related strongly to values and because the Rokeach value survey has not been widely used to examine consumer behaviour issues. The LOV is argued to isolate values more applicable to marketing applications and consumer behaviour issues and is more parsimonious than other measures such as Rokeach (1973) or Schwartz (1987).

Following an approach used by McIntyre, Claxton and Jones (1994), the LOV scale was modified to a six-point unipolar Likert-type scale with scale anchors ranging from ‘Not at all important’ to ‘Extremely important’. This proved easier for respondents to process and respond to, given the feedback in the pretest stage. Munson and McIntyre (1979) have demonstrated, using Rokeach values survey (RVS), that values could be assessed using Likert-type scaling procedures, without affecting scale reliabilities. This provides a foundation for arguing that modification of the scale format and number of scale points is appropriate if it is logical and it improves the ease of administration of the instrument.

Kahle’s LOV, containing nine items and four items from Schwartz (1992) were included in the instrument. These items represented the four values dimensions of ‘influential’, ‘preserving my public image’, ‘curiosity’ and ‘independence’ and were believed to add conceptual breadth to the LOV. The relevant items for measuring global values are reported in Volume II Appendix A as items 6 to 18.

---

1 The instrument also included items on self-monitoring and locus of control. They do not form part of the study that is the foundation of this Ph.D. and are not discussed. The data on these constructs were gathered simultaneously because of the opportunity of accessing a large sample.

2 All questionnaire sections and items relevant to the following discussion can be found in Volume II Appendix A.
Materialistic Values Measure

The literature identifies various methods and instruments for the measurement of materialism. It has, for example, been measured by examining the importance of various social goals and by assessing attitudes toward possessions and associated behaviours. All the existing measures, except for Richins and Dawson (1992), seem to suffer from at least one or two important limitations. First, many of the measures do not possess adequate levels of reliability. Second, the construct validity of many of the measures has not been established. Since most scales except perhaps those of Belk’s (1984) and Richins and Dawson (1992), have not involved the psychometric procedures of construct definition, scale refinement and validity assessment, they are of limited use for this study. Further, given the problematic nature of wide-ranging reliability found in the literature related to Belk’s (1984) scale (0.09 to 0.81) this scale was not used. Therefore, the scale that appeared to have been developed using commonly accepted standards of psychometric scale development was that of Richins and Dawson (1992). This measure has proven to be a valid and reliable (.71 to .83) measure of an individual’s materialistic values (Bearden, Netemeyer and Mobley 1995). The scale consisted of 15 items tapping materialistic values using a six-point bipolar Likert-type format, anchored from ‘Strongly disagree’ to Strongly agree’. The items used to tap the materialistic values are reported in Appendix A from items 31 to 45.

Measure of Self-image / Product-image Congruency

Sirgy et al’s (1997) measure of global assessment of self-concept formed the theoretical basis of the measure developed to tap self-image product-image congruency for this study. Fundamentally, this method does not cue respondents to a specific image category or dimension. The method cues respondents to assess the product, typical user and their own image. The method then guides them to indicate their global perception of degree of match or mismatch between how they see themselves (self-image) and how they see the product (product-image). In other words, this method captures self-image congruence directly and globally.

The perceived congruity of the actual, ideal and social selves was assessed in this study. This method was considered to be more appropriate and parsimonious for this study’s purpose and, therefore, formed the basis of the measure of self-image product-image
The goal of modification was to make administration and interpretation easier for estimating the models and associations with involvement. The modified approach sought, from respondents, a direct rating of the degree to which they perceived the product’s image and their own self-image to be similar or not similar. This scale consisted of six items tapping the actual, ideal and social self-image product-image congruency using a six-point bipolar Likert-type format, anchored from ‘Strongly disagree’ to ‘Strongly agree’. The corresponding items tapping these congruencies are reported in Appendix A as items 174 to 179.

Consumption Values
Consumption values were assessed using a modified version of the scale used to measure global values. This scale focused on values engaged in the purchase and consumption of the focal object. The modification consisted of providing instructions to the respondents as to how important each value was in their purchase and consumption of the product. The scale consisted of the nine LOV items and four additional items from Schwartz (1992) using a six-point unipolar Likert-type format, anchored from ‘Not at all important’ to ‘Extremely important’. The items tapping the consumption values are reported in Appendix A as items 161 to 173.

Product Knowledge/Expertise
The product knowledge/expertise scale was developed specifically for this study and assessed respondent’s perception of their subjective knowledge and expertise relevant to the product and its use. The scale consisted of four using a six-point bipolar Likert-type format, anchored from ‘Strongly disagree’ to ‘Strongly agree’. The items tapping the knowledge dimension are reported in Appendix A as items 146 to 149.

Consumer Confidence
The items tapping consumer confidence measured the degree of confidence the respondent has in their ability to choose brands and make decisions in the product category. The scale consisted of three items using a six-point bipolar Likert-type format, anchored from ‘Strongly disagree’ to ‘Strongly agree’. The items corresponding to the confidence construct are reported in Appendix A as items 150 to 152.
**Consequences by Number**
The items tapping consequences by number sought to determine the number of positive consequences of owning or using the product and the number of connections between the respondent’s life and the product. Consequences by number also identifies the number of perceived benefits from owning and using the product. The scale consisted of three items using a six-point bipolar Likert-type format, anchored from ‘Strongly disagree’ to ‘Strongly agree’. The items tapping this construct are reported in Appendix A from items 153 to 155.

**Consequences by Type**
The items tapping the type of consequences provided underlying motive dimensions for owning and consuming the product. Consequences by type identified a limited but important set of underlying reasons or motives for purchase and consumption of the product. The scale consisted of six items using a six-point bipolar Likert-type format, anchored from ‘Strongly disagree’ to ‘Strongly agree’. The items corresponding to this construct are reported in Appendix A as item 71 and 156 to 160.

**Product Involvement**
The items tapping product involvement operationally defined the first type of involvement construct developed in chapter 3. The scale consisted of 25 items using a six-point bipolar Likert-type format, anchored from ‘Strongly disagree’ to ‘Strongly agree’. Items 57 to 81 relate to the product involvement measure in Appendix A.

**Purchase Decision Involvement**
The items tapping purchase decision involvement operationally defined the second type of involvement construct developed in chapter 3. The scale consisted of 24 items using a six-point bipolar Likert-type format, anchored from ‘Strongly disagree’ to ‘Strongly agree’. Items 82 to item 105 in Appendix A correspond to the purchase decision involvement measure.

**Consumption Involvement**
The items tapping consumption involvement operationally defined the third type of involvement construct developed in chapter 3. The scale consisted of 23 items using a six-
point bipolar Likert-type format, anchored from ‘Strongly disagree’ to ‘Strongly agree’. Items 106 to 128 in Appendix A relate to the consumption involvement measure.

Communications Involvement
The items tapping communications involvement operationally defined the fourth type of involvement developed in chapter 3. The scale consisted of 17 items using a six-point bipolar Likert-type format, anchored from ‘Strongly disagree’ to ‘Strongly agree’. Items 129 to 145 in Appendix A relate to the communications involvement measure.

6.9 Empirical stage 2: Data collection procedures for the main study
This stage embodied the main methodology used to gather the data necessary for testing the various models. The discussion below provides the rationale for the sample selection procedure, the survey administration method and the product selection criteria in detail.

6.9.1 Sample selection procedure
A population is defined for the purposes of this study as any complete group sharing some common set of characteristics (Zikmund 1991). The arguments expressed by Sieber (1973 p.1341) were seen as relevant to this study in that more and more, surveys are conducted among selected communities or organisations rather than among samples of isolated individuals. In these situations, a great deal of careful thought must be given to the selection of the collective (Sieber 1973). These arguments were seen as critical in deciding the philosophy that guided the definition of the population and the sampling frame for selecting the sample.

The sample for the survey were identified as students and consisted of the following groups:
1. Undergraduate On-campus Students
2. Undergraduate Off-campus Students
3. Postgraduate On-campus Students
4. Postgraduate Off-campus Students
Calder, Phillips and Tybout (1981) argued that student samples are appropriate if a study’s purpose is focused more on theory testing than on application, as was the case in this study. Another consideration was the demographic profile of the students. The part-time off-campus student characteristics meant a significant percentage of the sample work and thus were more in line with what we might term real consumers. That is, they did not possess the traditional student characteristics of being young and not working. Thus the overall sample consisted of both the traditional students used in much consumer behaviour research and those with characteristics more similar to more general consumers.

Students were randomly drawn from Student Administration Database of a major regional Australian University. This database contained both full-time and part-time undergraduate and post-graduate students studying in both on-campus and off-campus mode. The final sample consisted of 900 randomly selected students supplied to the researchers through the student administration office of the university. The target of 900 was based on the postage and handling costs and resources available to the researcher. Each questionnaire cost approximately $2 for printing and postage. The budget constraints allowed approximately 900 questionnaires to be printed and posted.

6.9.2 Survey administration method

The survey instrument used in empirical stage 2 of the data gathering for the study could have been administered in a number of ways. These would include telephone, in person or by mail. The researcher ruled out telephone because of the survey’s length and the types of questions contained in the instrument. Surveying respondents in person was also not possible because the geographical dispersment of the sample and prohibitive costs involved.

The only feasible method available was a mailed, self-administered questionnaire. The mail questionnaire could reach the geographically dispersed sample simultaneously and at relatively low cost (Zikmund 1991). The self-administered questionnaire could be widely distributed to the sample quickly and inexpensively. The cost of the questionnaire was relatively low when compared with either telephone or personal administration.
Respondent convenience was also considered in the decision of which method of survey administration to employ. Basically, the mail questionnaire could be filled out whenever the respondent had sufficient time. Thus, there was a better chance that the respondents would take a reasonable amount of time to think about their replies (Zikmund 1991). While these aspects were considered to be some of the advantages of mail administration there were also some disadvantages. These included investigator absence, which meant that the actual questioning process was beyond the control of the researcher. Even though the printed stimulus was the same for each respondent, they could potentially attach a different personal meaning to the questions in the instrument. It was believed that the rigorous process adopted to develop the instrument would reduce the significance of these aspects.

Also the mail questionnaires were highly standardised, therefore the questions and the instructions needed to be very clear. This problem was explicitly addressed in the design and testing stage to ensure that the final questionnaire was as clear in its instructions and questions as it could be. The length of the instrument could also have created potential problems; the general rule outlined by Zikmund (1991) was adopted and hence the instrument was kept to a minimum number of pages while maintaining acceptable page layout. Although the survey was lengthy, pretesting revealed no concerns related to respondent fatigue.

The final area considered in the administration of the survey was that of response rates. In an attempt to obtain an adequate response rate (see chapter 7 section 7.1.2 for sample size requirements), a number of strategies were employed (Ayal & Hornik 1986). The survey was sent via mail with an accompanying cover letter (see Appendix A) and reply paid envelope. The study objectives and importance of the respondents answers were explained and a final reply deadline was indicated in the cover letter. The objective of the cover letter was to highlight to the respondent the significance of the study and the importance of their participation. The respondents were identified by name and asked to participate. The reply deadline was included and anonymity offered to increase the probability of the respondents completing the questionnaire early. By including a definite reply date the probability of putting the questionnaire aside and forgetting was likely decreased.
Nonresponse issues were also considered in regard to meeting the required number of respondents to adequately fit the models. An examination of the structural equation modeling literature revealed an extensive discussion on sample size, very little discussion of response rate and nonresponse issues. To overcome the lack of literature regarding response rate issues discussion was undertaken with academics who had published articles in consumer behaviour using structural equation modeling. Consensus in these discussion was seen to revolve around the objectives of the study and how the models were being tested. In general it was believed by the academics that, as the study and models were not designed to examine differences among respondents or groups of respondents nonresponse was not considered critical as long as the requirements were met as detailed in chapter 7 section 7.1.2 regarding achieving an adequate response rate (per comm, Professor Richard Mizerski 6-3-1997).

The selection of subjects, survey administration method and product for use in this study were not independent considerations. The potential sample of products was limited to those products with which the entire sample would have the potential ability or opportunity to purchase and consume.

### 6.9.3 Product selection criteria and procedures

In order to test the models discussed in chapter 5, one or more products had to be selected to serve as the focal object. The inherent consideration which guided the product selection was that the product stimulus would elicit adequate variance in experiences and involvement across the respondents. The final product used was drawn from a sample of products such as durable, nondurable, convenience goods shopping goods and low and high-priced products. It was also important that the final product had been previously used in studies found in the involvement literature to allow for some comparison of findings. Table 6.4 contains the sample of products considered for this study.

<table>
<thead>
<tr>
<th>Table 6.4 Potential Product Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sport Shoes</td>
</tr>
<tr>
<td>Stereo Systems</td>
</tr>
<tr>
<td>Performing Arts</td>
</tr>
<tr>
<td>Sports Drinks</td>
</tr>
</tbody>
</table>
This sample comprised products which the respondents were generally expected by the researcher and principal supervisor to have bought and consumed and had appeared in the literature as indicated in chapter 2. This was done so that the final selection of the product would allow for a maximum number of completed responses. The list of products was presented to a number of academics in the researcher’s department and discussed as to whether respondents would potentially have different levels of involvement toward the products. After this discussion, the researcher discussed the proposed products and general feedback given in discussions with colleagues about the products with the principal supervisor. After considering the products and the objective of the study, fashion clothing was selected as the focal product for the study. The product offered several advantages for the research into involvement and its nomological network. The respondents potentially owned the product and were familiar with it. It is also assumed the respondents could exhibit a wide range of involvement levels with respect to the product. Further, the product type and semantic labeling in surveys of ‘fashion clothing’ had been used in previous studies on involvement by Bloch (1982), Fairhurst, Good and Gentry (1989), Flynn and Goldsmith (1993), Goldsmith and Emmert (1991), Ohanian (1989), Tigert, Ring and King (1976) and Tigert, King and Ring (1980) to provide the stimulus for respondents to focus on.

6.10 Summary

Although related to case studies, Parkhe (1993) makes a valid statement relevant to any study when he argued that strong measures can be taken to build rigor into studies at the design stage, data collection and data analysis stages. It is argued that this research design represents a logical set of statements and that the quality of the design can be judged according to certain logical tests (Yin 1989) related to the process used to design the instrument and gathering the data. The design of this study evolved following a method advocated by Tull and Hawkins (1990), Converse and Presser (1986) and Bearden, Netemeyer and Mobley (1993) and others which specified the tests, tactics and the phases to ensure a high level of validity and reliability.
This chapter has dealt with the data gathering methods of the empirical component of the study. It discussed the design of the survey instrument, the target population and sampling technique adopted and the administration of the instrument.

After detailing the data gathering methods it is necessary to discuss the underlying analysis and modeling procedures. Taking the following quote as a key rationale in adopting the processes used develop the data gathering methods:

As Torgerson (1958) suggests in discussing the ordering of the various sciences along a theoretical-correlational continuum (p.2): It is more than a mere coincidence that the sciences would order themselves in largely the same way if they were classified on the basis to which satisfactory measurement of their important variables has been achieved. The development of a theoretical science . . . would seem to be virtually impossible unless its variables can be measured adequately. Progress in the development of marketing as a science certainly will depend on the measures marketers develop (Churchill 1979 p.72-73).

The discussion of the data analysis rationale and processes follows in chapter 7.
Chapter 7

Rationale for Data Analysis and Model Testing

Testing structural equation models is viewed as a way of testing a specified theory about relationships between theoretical constructs. . . . this is not just a matter of reading the data into a structural equation program, obtaining a chi-square, and accepting or rejecting the model. Rather, as Tukey (1977) so elegantly puts it, “It is important to understand what you CAN DO before you learn to measure how WELL you seem to have DONE it” (p.v). There are many issues to be considered before an attempt of a statistical test should even be made (Jöreskog 1993 p.294).

7. Introduction
Chapter 5 provided the hypothesised model; developed out of chapters 3 and 4. These models formed the focus of the empirical component of the study and directed the development of an appropriate research design. Chapter 6 discussed the data gathering methods of the research design. This chapter discusses the rationale and basis for structural equation modeling (SEM) and procedures for estimating path coefficients and evaluating the fit of the hypothesised models.

The following discussion presents the rationale and theory behind the fitting of congeneric models in a two-stage process in this study. Specifically, this chapter discusses the initial preliminary analyses performed to assess the data and the use of congeneric SEM theory to test the structural models proposed here. The development of congeneric measurement models is described as a means of data reduction in order to obtain a manageable number of valid, more reliable, composite factors or variables. Composite scores are used in order to test the proposed models through a process of fixing the composite factor regression coefficients and measurement error variances (Raykov 1997).
Importantly, structural equation modeling is distinguished from other analytical approaches by two key characteristics. Firstly, it is distinguished by its estimation of multiple, interrelated and causally-directed dependence relationships between observed (measured) and unobserved (latent) constructs and, secondly, by its ability to account for measurement error in estimating such relationships (Hair, Anderson, Tatham & Black 1995). This chapter also discusses issues related to sample size, parameter estimation and model fit evaluation relevant to the structural models being tested.

Structural equation modeling has emerged in the consumer behaviour area as an integral research and analysis approach to help understand the relationships between observed measures and unobserved constructs. Synonymous terms such as analysis of moment structures, covariance structure analysis, latent variable analysis, confirmatory factor analysis and LISREL analysis (named after a very popular computer program) have been commonly employed in the marketing and consumer behaviour disciplines to refer to the analyses encompassed by structural equation modeling (Hair et al 1995).

7.1 An overview for Testing structural equation models

Many theories and models in consumer behaviour research are formulated in terms of latent or hypothetical constructs that are not directly observable or measurable. This is especially the case in studies of purchase and consumption behaviour that deal with constructs such as involvement, values, consumer confidence and self-image product-image congruency. To test such purchase and consumption related theories, the constructs involved are typically considered to be 'latent' variables which underlie the responses recorded in a larger set of observed or measured variables. Structural equation modeling (SEM) is a sophisticated analytical approach to theory testing that attempts to account for measurement error during the estimation and modeling of latent constructs. This process is accomplished by simultaneously estimating the measurement properties of latent constructs and structural relationships among latent constructs and/or other important variables.

Within the literature on structural equation modeling, there are arguments advocated by researchers for the adoption of a two-stage modeling process where the data is assessed for
its structure and reliability and the measurement model is first estimated in a similar manner to that of conventional factor analysis. The measurement model is then “fixed” in the second stage when the structural model is estimated (Anderson & Gerbing; 1988, 1992; Hair, Anderson, Tatham & Black 1995; Hu 1995; Kenny 1979; Manaresi & Uncles 1996 and Williams & Hazer 1986).

Fundamentally, the rationale for such an approach is that adequate assessment and accurate representation of the data and the reliability of the measured variables is best accomplished in two stages. Importantly, while the measurement and structural models cannot be evaluated in complete isolation, the potential for within-construct versus between-construct effects must be considered. Such effects in estimation may be significant and result in interpretational confounding (Burt 1976 and Hair, Anderson, Tatham & Black 1995).

Anderson and Gerbing (1988, 1992) argue much can be gained in theory testing and the assessment of construct validity from separate estimation of the measurement model prior to simultaneous estimation of the measurement and structural submodels. The measurement model tested in conjunction with the structural model enables a comprehensive confirmation assessing construct validity. Therefore, the measurement model provides an assessment of convergent and discriminant validity. Given acceptable convergent and discriminant validities, the test of the structural model then constitutes a confirmatory assessment of nomological validity (Anderson & Gerbing 1988, 1992; Campbell 1960 and Cronbach & Meehl 1955).

7.1.1 Stage 1 Preliminary Analyses
While this chapter is devoted largely to the second stage of the two stage process of constructing and estimating the structural equation models, it is important to cover the analyses that were performed in the first stage of the two stage process. Before evaluation of the structural equation model is undertaken, an assessment of the adequacy of the input data and the statistical assumptions underlying any estimation methods was performed. Bagozzi and Yi (1988) propose that one of the first activities in the evaluation of structural equation models should be an assessment of the input data and the statistical assumptions
underlying any estimation methods used in analyses. They recommend a number of univariate and multivariate statistics to accomplish such a task. Initial screening through estimates of skewness, kurtosis, variance and correlations (Bagozzi & Yi 1988).

There are two major kinds of summarising statistics that assist in assessing data. The first provides measures of the midpoint of the distribution of the data and are known as measures of central tendency. The second gives an indication of the amount of variation in the data comprising the distribution and are known as measures of dispersion (Parasuraman 1986 and Tull & Hawkins 1990).

Appropriate measures of central tendency and dispersion to assess the data are means, standard deviation, standard error of the mean, skewness and kurtosis estimates of the data. Estimates of the mean calculate the arithmetic average; the sum divided by the number of cases. The standard error of the mean is a measure of how much the value of the mean may vary from sample to sample taken from the same distribution. It is the standard deviation of the distribution of all possible means, if the sample of the same size were repeatedly taken. The standard deviation is a measure of dispersion around the mean, equal to the square root of the variance, measured in the same units as the variable itself. The variance is a measure of the dispersion around the mean, equal to the sum of the squared deviations from the mean divided by one less than the number of cases and is measured in units that are the square of those of the variable itself (Green, Tull & Albaum 1988 and Parasuraman 1986).

The most fundamental assumption in multivariate analysis is the normality of the data referring to the shape of the data distribution for an individual metric variable and its correspondence to the normal distribution, the benchmark for statistical methods. Common methods of estimating data-normality are through estimates of skewness and kurtosis. Kurtosis refers to the peakness or flatness of the distribution compared with the normal distribution (Hair et al 1995). Another common pattern is a simpler arc, either above or below the diagonal, indicating the skewness of the distribution. Skewness is a measure of the asymmetry of a distribution (Hair et al 1995 and Tull & Hawkins 1990).

Initial assessment of the data also included examination of the factor structure, and internal consistency of the data to assess the structure and quality of the measurement items.
(Hulland, Chow & Lam 1996). Further assessment of data was obtained through correlation analysis. Pearson correlation calculates the correlation between vectors of values. It is a measure of linear association ranging from -1 to +1, with a value of 0 indicating no linear association.

Exploratory factor analysis has been used widely as a technique to develop scales and subscales (Gorsuch 1997; Hair et al 1995 and Parasuraman 1986). The goal of item analysis is to determine the structure of the data and select those items that are most related to the construct (Stewart 1981). Factor analysis is used to analyse interrelationships among a large number of variables and to explain such variables in terms of their common underlying dimensions or structure. This goal is aided by evaluating how each item relates to its own construct, as well as how it relates to other associated or similar constructs.

There are a number of factor analytic techniques used to examine data structure, such as maximum likelihood, alpha factoring, principal axis, image, generalised least squares and unweighted least squares (Gorsuch 1983, 1997). A commonly applied factor analytic technique used in consumer behaviour is principal component analysis with varimax rotation (Green, Tull & Albaum 1988) and was a suitable method of factor analysis for the purposes of this study. Principal components analysis, uses the principal components model, in which variables are assumed to be exact linear combinations of factors. Performing factor analysis as a method of extracting factor from data, is generally combined with some form of factor rotation. Factor rotation is the process of manipulating or adjusting factor axes to achieve a simpler and more pragmatically meaningful factor solution (Hair et al 1995). A popular method of orthogonal rotation is the varimax method. To adequately apply factor analysis the sample size should exceed a minimum requirement of 5 times the number of observations per variable (Gorsuch 1997 and Hair et al 1995).

In summary the initial analyses performed were estimates of central tendency and dispersion, exploratory factor analysis, Pearson correlation and reliability estimates using SPSS and then confirmatory factor analysis of the factors for each of the constructs using AMOS. To judge the fit of the confirmed factors, goodness-of-fit indices (GFI), adjusted-goodness-of-fit (AGFI), root mean square residual (RMSR) and root mean square error of approximation (RMSEA) fit indices were estimated. The reliability of the constructs was
also estimated using the Werts, Rock, Linn and Jöreskog (1978) index for the composite reliability ($R_{cc}$).

The exploratory assessment of the data was also valuable because it helped determine the most appropriate fitting function. There are various protocols of structural estimation which impose different assumptions about data theory and the ties between unobservable variables and indicators. Commonly applied fitting functions in structural equation modeling are maximum likelihood, generalised least squares, ordinary least squares and asymptotic distribution free estimation. All fitting functions except for ordinary least squares are not scale dependent. All fitting functions except asymptotic are based on multivariate normality of the data. Asymptotic distribution free estimation does not depend upon data normality. Consumer behaviour data do not always satisfy the requirements of multinormality and interval scaling or attain the sample size required by maximum likelihood estimation. Estimation problems are introduced into structural equation modeling when the distribution of the observed variables depart substantially from multivariate normality. If multivariate normality is violated, the variation of the observed variables will not be completely summarised by the sample covariances. In this case information from higher order moments is needed. Asymptotic distribution free estimation procedures provides such higher order information by estimating fourth order moments.

Therefore not only does the preliminary analyses provide an assessment of the data quality it also provide guidance in determining the most suitable fitting function to estimate the hypothesised models. Considering the issues related to data potentially containing skewness and kurtosis, estimation based on asymptotic distribution free or weighted procedures should be considered (Lomax 1989).

These preliminary analyses procedures allowed for evaluation of the soundness of each measurement indicator in accordance with the two-step approach to structural equation modeling (Anderson & Gerbing 1988, 1992).

It is important to indicate that the underlying assumption in this process is, that this structural modeling analysis rests on two conceptually distinct models; measurement and structural (Gerbing 1979 and Jöreskog & Sörbom 1978). The measurement model specifies
the causal relations between the observed variables or indicators and the underlying latent variables or constructs, which are presumed to determine responses to the observed measures. The structural model specifies the causal relations among the theoretical constructs. The reason for the distinction drawn, is proper specification of the measurement model is necessary before meaning can be assigned to the analysis of the structural model. Good measurement of the latent variables is a prerequisite to the analysis of the causal relations among the latent variables (Anderson & Gerbing 1982). The adequacy of the measurement model will be established in chapter 8.

7.2 Sample Size

The issue of sample size for adequate estimation has been the focus of an extensive debate in the structural equation literature for some time (Chou & Bentler 1995; Carmines & McIver 1981; Hoelter 1983; Marsh, Balla & McDonald 1988; Tabachnick & Fidell 1989 and Tanaka 1987). The simple rule of thumb, applied to multivariate analytical techniques such as MANOVA or multiple regression, of ten research participants for each measured or latent variable is generally not applicable in structural equation modeling, for two fundamental reasons. Firstly, in structural equation modeling it is more appropriate to consider the ratio of the number of research participants to the number of parameters being estimated. For example, in multiple regression, the regression coefficients are estimated for each independent variable and represent the unique contribution made by variables in the prediction of the dependent variable. However, in structural equation modeling, the measurement model focuses on the contribution of a latent characteristic towards the responses observed on a number of variables. This requires estimation of both the regression coefficients of the latent constructs on the observed variables and the error variances associated with the observed variables. Consequently, more information (or research participants) are needed because there are more parameters to be estimated (Chou & Bentler 1995 and Holmes-Smith & Rowe 1994). Secondly, the statistical theory underlying parameter estimation in SEM is asymptotic in nature. This means that statistics such as standard errors for parameter estimates increase in precision as the total number of cases approaches infinity (Chou & Bentler 1995 and Holmes-Smith & Rowe 1994).
The work of Jöreskog and Sörbom (1988, 1989) and Holmes-Smith and Rowe (1994) provided guidance in determining sample size adequacy for this study. In cases where there are fewer than 12 variables ($k$'s) a sample size of 200 is appropriate (Jöreskog & Sörbom 1988). When the number of variables ($k$) is greater than 12, the sample size must be at least $1.5k(k+1)$. Table 7.1 below indicates adequate sample sizes required for asymptotic distribution free (ADF) estimation of structural equation models.

### Table 7.1 Minimum Sample Size Required for ADF Estimation

<table>
<thead>
<tr>
<th>Number of variables [$k$]</th>
<th>Sample size required (1.5$k$(k+1) or 200 if $k &lt; 12$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>200</td>
</tr>
<tr>
<td>20</td>
<td>630</td>
</tr>
<tr>
<td>30</td>
<td>1,395</td>
</tr>
<tr>
<td>40</td>
<td>2,460</td>
</tr>
</tbody>
</table>

(Source: adapted from Holmes-Smith & Rowe 1994)

In moderate to large studies, such as that carried out here, where a large number of potentially non-normal indicator variables is obtained, it is possible to develop congeneric models as a means of data reduction. Sample size and congeneric theory are inextricably linked. Congeneric modeling allows the reduction of data to obtain a more manageable number of composite variables that are used to estimate the model parameters (Baumgartner & Homburg 1996 and Hau 1995). Congeneric models generally permit reliance on smaller samples as being adequate for stable parameter estimation and full model fitting. This procedure also allows for final models to be fitted with improved scale validity and reliability of the composite constructs. This procedure will be further elaborated on in the following sections.

### 7.2.1 Congeneric theory

In areas such as psychology and consumer behaviour, responses to a number of items or questions of indicators are usually collected in order to develop composite scores that represent some underlying latent construct or trait (e.g., Bacon, Sauer & Young 1995 and Hau 1995). These composite scores are determined through the fitting of explicit
measurement models for each construct. There are essentially three forms of such measurement models. These models are detailed below in Figure 7.1 and are labeled Parallel, Tau-equivalent and Congeneric models. Figure 7.1 provides a diagrammatic representation of the three different forms of structural equation measurement models.

Figure 7.1 Forms of Structural Equation Models

PARALLEL Model
Where: $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda$ & $\text{var} \quad \theta_1 = \text{var} \quad \theta_2 = \text{var} \quad \theta_3 = \text{var} \quad \theta_4 = \theta$

TAU-EQUIVALENT Model
Where: $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda$ however $\text{var} \quad \theta_1 \neq \text{var} \quad \theta_2 \neq \text{var} \quad \theta_3 \neq \text{var} \quad \theta_4$

CONGENERIC Model
Where: $\lambda_1 \neq \lambda_2 \neq \lambda_3 \neq \lambda_4$ and $\text{var} \quad \theta_1 \neq \text{var} \quad \theta_2 \neq \text{var} \quad \theta_3 \neq \text{var} \quad \theta_4$

(Source: adapted from Holmes-Smith & Rowe 1994)
The Parallel measurement model is based on the assumption that each item or measure is treated as a parallel measure and that each measure provides an equally accurate reflection of the true underlying trait. Parallel measurement also assumes that errors of measurement for items of composite scores have the same variance. This model is thus based on the twin assumptions of equal error variance (theta) parameter and regression coefficients (lambda) parameter. A Tau-equivalent model is based on the assumption of equality of accuracy of measurement the indicators (lambda parameters), as in the Parallel models; however, the Tau-equivalent model error variances (theta parameters) are allowed to differ. The third type of model, the Congeneric model (Hau 1995 and Jöreskog 1971), is based on the assumption that each indicator or item reflects the same true score as in the Parallel and Tau-equivalent models; however each indicator reflects the underlying trait or construct in differing degrees and with differing error variances. Thus, within congeneric models, both the error variances (theta parameters) and the regression coefficients (lambda parameters) are allowed to differ. That is, Congeneric error variance and regression coefficients are fixed at a 'true' estimated value as opposed to the Parallel and Tau-equivalent whose values are arbitrarily assigned a value (Raykov 1997).

7.2.2 Calculation of lambda and theta values

Latent constructs have been traditionally computed in social psychology, psychology and educational research as factor scores or unit-weighted composite scores of their measured indicators. Such derived indices for the composite variables are then treated as continuous variables in omnibus general linear modeling techniques such as ANOVA or multiple regression which assume that such indices are measured without error. However, a more appropriate procedure is based on congeneric modeling. Congeneric models are those models whose error variances and regression coefficients are calculated and fixed at estimated values as indicated in Figure 7.1. That is, the error variance (theta) and the regression coefficient (lambda) are not given arbitrarily assigned values as in Tau and Parallel models (Hau 1995 and Jöreskog 1971).

This study investigates a large number of latent constructs that are each represented by a larger number of measured or observed variables. In order to deal with the problems
associated with estimating large numbers of observed indicators for latent constructs, a
defensible method of coping was sought. One way of overcoming such problems is to
assume that each of the measures for the latent constructs is parallel. It would then be
possible to simply sum all the related measures (items within each factor) to form a reduced
number of manifest or composite measures. This reduced number of composite variables or
factors could then be used as the observed variables in the structural equation models for
this study. If all items associated with any one of the constructs being investigated here are
parallel, then the reliability of the composite score will not be severely degraded.

However, if the measures associated with the various latent constructs studied here are
congeneric then a simple arbitrary unit weighting applied to the summation of the indicator
variables would result in the composite variables having a less than optimal reliability.
Additionally, if the measures for each latent construct do not reflect the same generic true
score, when they are added together the resulting composite score will lack validity (Hau
1995; Holmes-Smith & Rowe 1994 and Raykov 1997). This issue will be further
elaborated on in the following sections.

### 7.2.3 Congeneric modeling benefits

There are a number of key benefits from utilising congeneric modeling in structural equation
modeling. Firstly, a fitted congeneric model allows a large number of equivalent observed
variables or items to be reduced to a smaller number of composite scores, thus reducing the
number of variables in the proposed structural equation models. Secondly, as indicated by
Jöreskog (1971), Fleishman and Benson (1987) and Raykov (1997), fitting a congeneric
measurement model allows for differences in the degree to which each individual measure
(item) contributes to the overall composite scale thus providing a more realistic
representation of the data, a feature not associated with parallel or tau equivalent models.
Thirdly, the fit statistic for the congeneric model is a quasi-test of validity. For a model to
be accepted, the indicator variables contributing to the overall measurement of the manifest
or composite variable must all represent the same generic true score, meaning they must all
be valid measures of the underlying latent construct (Hau 1995; Jöreskog & Sörbom 1989
and Raykov 1997).
Fundamentally, a composite factor congeneric model is the simplest form of a measurement model and represents the regression of a set of observed indicators on a single latent construct. The congeneric model depicted in Figure 7.1 is a generalised form of a composite factor congeneric model, where the measured variables are represented as composite factors Test 1, Test 2, Test 3 and Test 4.

In the example provided by Figure 7.1 a generalised congeneric model is represented where, ξ is the single latent variable; Test 1, Test 2, Test 3 and Test 4 are the observed indicator variables which are the composite factor scores; θ₁, θ₂, θ₃ and θ₄ are the estimated errors associated with the measurement of Test 1, Test 2, Test 3 and Test 4; and λ₁, λ₂, λ₃, and λ₄ are the estimated regression coefficients in the regression of Test 1, Test 2, Test 3 and Test 4 on ξ respectively.

The example of the congeneric model provided in Figure 7.1 implies that the matrix of covariances (Σ) amongst the observed indicators can be expressed as:

$$\Sigma = \Lambda\Phi\Lambda' + \Theta_5$$

Where: Λ is the vector of regression coefficients λ₁, λ₂, λ₃, and λ₄

Φ is the covariance of ξ and

θ₅ is the matrix of variances and covariances (θ₁, θ₂, θ₃, and θ₄) amongst the measurement error terms δ₁, δ₂, δ₃, and δ₄

Jöreskog and Sörbom (1989) and Hau (1995) detail a procedure for fitting and accepting a single factor congeneric model where it is possible to compute an estimated composite score (ξ̂) for each respondent by using the formula:

$$\hat{\xi} = \omega X$$

They also argue that the vector of factor score regressions (ω) could be computed by the formula:

$$\omega = \Phi\Lambda'\Sigma^{-1}$$
For example, the model shown in Figure 7.1 indicates that the factor score regression weights $\omega_1$, $\omega_2$, $\omega_3$, and $\omega_4$ (where $\omega$'s elements are $\omega_1 = \lambda_1/\theta_1$, $\omega_2 = \lambda_2/\theta_2$, $\omega_3 = \lambda_3/\theta_3$, and $\omega_4 = \lambda_4/\theta_4$) would be calculated and the composite score for the $i$th respondent ($\xi_i$) would be calculated as:

$$\hat{\xi}_i = \omega_1 x_{1i} + \omega_2 x_{2i} + \omega_3 x_{3i} + \omega_4 x_{4i}$$

Importantly, these factors are not only used in computing composite scales, they can and are also used to determine the composite scale reliability. Werts, Rock, Linn and Jöreskog (1978) showed that the reliability of a factorial complex composite ($R_{cc}$) can be calculated by the formula:

$$R_{cc} = \omega_1 (\sum - \theta^2)^\omega \omega^T \Sigma \omega$$

Werts, Rock, Linn and Jöreskog’s (1978) $R_{cc}$ can also be used to supplement the traditional Cronbach alpha reliability measure. Reliability of any weighted subset of $y$ can be obtained by summing the corresponding elements of the estimated $\Sigma$ and $\theta^2$ (Werts, Rock, Linn and Jöreskog 1978).

Using this procedure, it is then possible to use composite scale reliabilities to fix the composite variables or factor regression coefficients (lambdas) and measurement error (thetas) in the hypothesised models detailed in chapter 5. Following this procedure of estimating both error variance and regression coefficients, it is possible to develop the structural equation models to examine the relationships amongst the latent constructs underlying the composite factors (Baumgartner & Homburg 1996 and Raykov 1997).

More importantly, because the reliabilities of the composite factors are known, it is possible to build this information into the models and account for the known amount of error associated with the congeneric measurement of each latent construct. Munck (1979) and Raykov (1997) showed that it is possible to fix both the regression coefficients (lambdas) which reflect the regression of each composite factor on its latent construct, and the measurement error variances (theta’s) associated with each composite factor. Munck
(1979) also showed that the regression coefficients (lambdas) and r (where r = the reliability) are given by the formula:

$$\lambda = \sigma(x) \sqrt{r}$$

and the measurement error variances ($\theta_i$) are given by the formula:

$$\theta = \sigma^2(x)(1-r)$$

This procedure is fundamental to the congeneric theory and was used to estimate the error variances and regression coefficients for the models estimated and presented in chapter 9. A congeneric model was developed for each of the constructs being examined. The actual structural equation model allowed fixed and free paths to be calculated using this procedure. This procedure corresponds to step 3, section 7.3.3, of the steps outlined by Hair et al (1995) which were followed in the modeling process.

### 7.3 Processes adopted in this study

Fundamentally, the true value of structural equation modeling lies in its capacity to simultaneously estimate the structural and measurement models in the analysis. In order to ensure the procedures are correctly specified in this study, a process similar to that outlined by Hair et al (1995) was followed. The stages of the process are listed in Table 7.2 followed by a discussion of each step that formed the basis of the modeling and estimation procedures employed.

**Table 7.2 Steps in Structural Equation Modeling**

<table>
<thead>
<tr>
<th>Step 1</th>
<th>Develop a theoretically-based model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 2</td>
<td>Construct a path diagram</td>
</tr>
<tr>
<td>Step 3</td>
<td>Convert path diagram into structural equations and specify measurement models. Construct the congeneric models.</td>
</tr>
<tr>
<td>Step 4</td>
<td>Choose the input matrix type &amp; estimate the proposed model</td>
</tr>
<tr>
<td>Step 5</td>
<td>Evaluate the fit of the model</td>
</tr>
<tr>
<td>Step 6</td>
<td>Interpret and modify the model if necessary</td>
</tr>
</tbody>
</table>
The following discussion focuses on the procedures adopted to ensure that rigorous model development and sound estimation procedures and evaluation were used in this study and is based on the six steps depicted in Table 7.2.

7.3.1 Step 1: Develop a theoretically based model

Structural equation modeling is based on causal relationships, in which the change in specific independent variables are proposed to cause changes in one or more dependent variables. However, the strength and conviction with which causation can be assumed to exist between the constructs in the model lies not with the analytical methods chosen, but in the theoretical foundation developed and justification provided to support the analyses (Hair et al. 1995). The theoretically-based structural equation models presented in chapter 5 represent a series of hypothesised models which specify how the variables in the analysis are generated and related to each other.

The basic goal of the present study is to understand involvement and its relationship to the other constructs. However, the desire to include all the possible variables needed to understand or meet this goal must be balanced against the practical limitations of structural equation modeling (Hair et al. 1995) and the limited resources available to the researcher. Though no theoretical limit exists on the number of variables to include in structural equation models, practical concerns regarding interpretation and meaningfulness are raised even before the technological limits of the currently available computers and software programs are reached (Holmes-Smith & Rowe 1994). The interpretation of the results also becomes more difficult as the number of constructs increases. Within the domain of consumer behaviour/marketing research using structural equation modeling, Baumgartner and Homburg (1996) argued that studies using more than 7 constructs are rare and on average, the number of constructs used is 7. Hulland, Chow and Lam (1996) indicated the number of latent variables ranged between 3 and 8 in a review of the literature they conducted. The present study employs 15 higher-order constructs to model involvement and to place involvement within a broader nomological network.
7.3.2 Step 2: Construct a path diagram
Path diagrams are very helpful in depicting a series of causal relationships. Path diagrams are based on graphically depicting relationships of constructs. A construct is a theoretical based concept that acts as a building block used to define relationships. Constructs such as involvement, values, product knowledge and self-image/product-image congruency form the theoretical building blocks of this study. This study was based on developing the path diagrams presented in chapter 5 based on the constructs as detailed in chapters 3 and 4 and then developing variables and measures that tap the domain of each of the constructs as discussed in chapter 6.

7.3.3 Step 3: Convert path diagram into structural equations and specify measurement model. Constructing the congeneric models
After developing the formal theoretically-based models and developing the path diagrams, the next stage was to develop or specify relationships in terms of the structural equations. The structural equations link the constructs being examined, the measurement model specifying which variables measure which constructs and a set of matrices indicating hypothesised associations among constructs or variables (Hair et al 1995). This process is accomplished automatically by the structural equation modeling software program (AMOS-Analysis of Moment Structures) used in this study. This capability easily allows for estimation of both structural and measurement models simultaneously in AMOS. AMOS automatically converts the path diagrams into a series of structural equations (AMOS users’ guide 1997) through its graphical input interface. This step corresponds to procedures discussed sections 7.2.1 and 7.2.3 where the calculation the error variances and regression coefficients were discussed.

7.3.4 Step 4: Choose the input matrix type & estimate the proposed model
The focus of structural equation modeling is not on individual observations, but on the pattern of multivariate relationships across a sample of respondents. The input data matrix most appropriate for such analysis was the variance/covariance matrix relating all variables specified in a particular hypothesised model. The use of the variance/covariance matrix is
recommended by Hair et al (1995) anytime a true test of theory is being performed as is the case in the present study. The variance/covariance matrix satisfies the assumptions of the methodology because it is the appropriate form of data for validating causal relationships.

Another issue in the estimation and testing of SEM's arises when the observed variables used in the measurement component of the model do not meet distributional assumptions. For example, non-normality in the data warrants particular adjustments in the estimation procedures used to fit such models. Defensible methods have been devised to handle situations where data do not meet the criteria of normality (usually when distributions are either highly skewed or show substantial kurtosis as previously discussed in section 7.1.1). To allow for fitting models with non-normal data, Browne (1984) developed the asymptotic distribution free (ADF) estimator which uses a weight matrix in the function for fitting covariance structures. This weight matrix was based on fourth-order central moments. AMOS also has the capability of using ADF in its fitting function for estimating models. ADF estimation was employed in the present study because many of the survey items reflected substantial kurtosis in their response distributions.

7.3.5 Step 5: Evaluate the global fit of the model

Model evaluation is one of the most controversial and difficult issues connected with Structural Equation Modeling (SEM). Mulaik, James, Van Alstine, Bennett, Lind and Stilwell (1989), MacCallum (1990), Steiger (1990) and Bollen and Long (1993) present a number of recommendations and methods for evaluating models. Baumgartner and Homburg (1996) also present a good review of assessment of model fit and this review provided guidelines for the evaluation criteria used here.

A wide variety of statistics have been proposed as measures of the merits of a proposed model. The AMOS program calculates all the generally applied fit indices found in the literature. The measures discussed here are used throughout the presentation of the individual factor measurement models in chapter 8, as well as for intermediate and full structural models in chapter 9. Fit measures are reported for each specified intermediate and full model. AMOS also presents the results for two additional models called the
"saturated" model and the "independence" model. In the saturated model, no constraints are placed on the population moments. The saturated model is the most general model possible. The independence model is the extreme or opposite of the saturated model. In the independence model, all observed variables are assumed to be uncorrelated with each other.

Thus, the proposed models are constrained versions of the saturated models. The key fit indices for the independence models calculated by AMOS will be presented for the full models in chapter 9 also. This will allow a comparison of key indices for the proposed model and the independence model estimated by AMOS.

7.3.5.1 Discussion of Model Fit Measures

The following is a discussion of the measures used in this study to assess the proposed models and is intended to provide the reader with a review of key fit measures and what are deemed to be acceptable values for each of the fit indices used.

7.3.5.2 Chi-square Test

The likelihood-ratio Chi-square statistic ($\chi^2$) is the most fundamental measure of overall fit of a model. It is argued that a large value of Chi-square relative to the degrees of freedom in the model signifies that the observed and estimated variance/covariance matrices differ considerably. Low Chi-square values, which result in probability levels larger than .05 or .01, indicate that the actual and predicted input matrices are not statistically different.

However, an important caveat in the use of Chi-square statistic for this study is the substantive criticism raised in the literature regarding its sensitivity to large sample sizes (Bagozzi 1982 and Hair et al 1995). As sample size increases so too does the chance of rejecting the model. This point is adequately covered by marketing researchers such as Bagozzi (1982) and Hair et al (1995) and there are generally accepted guidelines for alleviating such concern. Hair et al (1995) indicate that Chi-square becomes very sensitive to sample sizes, especially when sample size exceeds 200 with the results that even models with very good fit will show a significant $\chi^2$. Thus researchers have relied on the use of $\chi^2$/df ratio as a method of compensating for the sensitivity of Chi-square statistics to sample size.
7.3.5.3 Chi-square/degrees of Freedom Ratio ($\chi^2/df$)

A number of researchers have suggested the $\chi^2/df$ ratio as a measure of fit. Wheaton, Muthén, Alwin and Summers (1977) suggested a $\chi^2/df$ ratio of approximately 5.0 or less as a criterion for reasonable fit. Further, Marsh and Hocevar (1985) have indicated that various researchers have recommended using ratios as low as 2.0 or as high as 5.0 to indicate a reasonable fit (Hair et al 1995).

7.3.5.4 Root Mean Square Error of Approximation (RMSEA)

Another measure of model fit that is used is the RMSEA. Literature relating to the RMSEA indicates that a value of .08 or less indicates an acceptable root mean square error of approximation and thus good model fit (Browne & Cudeck 1993).

7.3.5.5 Root Mean Square Residual (RMSR)

The RMSR is the square root of the mean of the squared residuals. That is, the mean of residuals between observed and estimated input matrices. Because the input in this study is the variance/covariance matrix RMSR is the average residual covariance (Hair et al 1995). Although no threshold level is consistently found in the literature an acceptable level of model fit for this study is .140 or less. This value was determined by the researcher based on the input matrix type (Bagozzi & Yi 1988), and other studies found in the consumer behaviour literature provided a basis for setting an adequate RMSR for this study.

7.3.5.6 Goodness-of-fit index (GFI)

The GFI was devised by Jöreskog and Sörbom (1984) and represents the overall degree of fit, unadjusted for model degrees of freedom. The literature indicates a value of 0 (zero) indicates poor fit while a value of 1 indicates perfect fit (Hair et al 1995). Thus, values close to .9 or above are deemed acceptable in this study (Hair et al 1995).

7.3.5.7 Adjusted Goodness-of-fit Index (AGFI)

The AGFI is an extension of the GFI, which is adjusted for the degrees of freedom in the proposed model with respect to the degrees of freedom for the independence model. The AGFI thus takes into account the degrees of freedom available for testing the model. Again, values range between 0 indicating poor fit and 1 indicating perfect fit. Values
deemed acceptable here are those above .8 and approaching .9 (Baumgartner & Homburg 1996).

7.3.5.8 Tucker-Lewis Index (TLI)
The TLI was discussed by Bentler and Bonett (1980) in the context of analysis of moment structures. The TLI is also referred to in the literature as the nonnormed fit index (NNFI). This index combines a measure of parsimony (adjusted for degrees of freedom) into a comparative ratio between the proposed model and the independence model and results in values ranging between 0 indicating poor fit and 1 indicating excellent fit. Recommended values of greater than .8 or above are deemed acceptable for this study (Hair et al 1995 & Baumgartner & Homburg 1996).

7.3.5.9 Normed Fit Index (NFI)
The NFI is a relative comparison of the proposed model to the independence model. The NFI sees values ranging from 0 indicating poor fit to 1 indicating perfect fit. Recommended values of .8 or greater are deemed acceptable and indicate good model fit (Baumgartner & Homburg 1996).

7.3.5.10 Parsimonious Normed Fit Index (PNFI)
The PNFI is a modification of the NFI and takes into account the number of degrees of freedom used to achieve a level of fit. Parsimony is defined here as achieving higher degrees of fit per degree of freedom used. Again, higher values are more desirable, however, there is no generally recommended level of acceptable fit for this measure. However, values of differences between models of .06 and .09 appear to be indicative of substantial differences in models (Hair et al 1995).

7.3.5.11 Parsimonious Goodness-of-fit Index (PGFI)
The PGFI is based on the parsimony of the proposed model. Parsimony is indicated through the degrees of freedom and the number of manifest variables. That is, PGFI takes into account the number of degrees of freedom used to achieve a given level of fit. The values of the PGFI varies between 0 and 1 with higher values indicating acceptable model fit (Baumgartner & Homburg 1996).
7.3.5.12 Akaike Information Criterion (AIC)

The AIC is similar to the PNFI in that it is a comparative measure between models with differing numbers of constructs. The AIC is a reverse-scaled measure of fit where values closer to 0 are indicative of better fitting models with greater parsimony. This index not only indicates a good fit between the observed and predicted matrices, but also a model not prone to over-fitting when values approximate zero (Hair et al 1995).

7.3.5.13 Relative Fit Index (RFI)

Baumgartner and Homburg (1996) provided some indications of fit achieved for RFI in their literature review of structural equation modeling in marketing. Their review indicated that values for RFI ranged between .82 and .85 in the studies examined, values above .8 are deemed acceptable in this study.

7.3.5.14 Incremental Fit Index (IFI)

Baumgartner and Homburg (1996) provided some indications of fit achieved for the IFI in their literature review of structural equation modeling in marketing. Their review indicated that values for IFI ranged between .95 and .98 in the studies they examined and values approaching .9 are deemed acceptable in this study.

7.3.5.15 Comparative Fit Index (CFI)

Baumgartner and Homburg (1996) provided some indications of fit achieved for the CFI in their literature review of structural equation modeling in marketing. Values for CFI in studies examined for this research ranged between .8 and .97, and values above .8 are deemed acceptable in this study.

7.3.5.16 Noncentrality Parameter (NCP)

Hair et al (1995) identify the NCP as an alternative to the likelihood-ratio Chi-square statistic that is relatively independent of sample size. The noncentrality parameter equals the $\chi^2$ minus the degrees of freedom (Hair et al 1995). The NCP is used to compare alternative models and the objective is to minimise the values. Thus lower values are deemed acceptable.
7.3.5.17 Scaled Noncentrality Parameter (SNCP)
To standardize the NCP it can be divided by the sample size to obtain the SNCP. The scale noncentrality parameter equals the $\chi^2$ minus the degrees of freedom divided by the sample size (Hair et al 1995). The SNCP is used to compare alternative models and the objective is to minimise the values.

7.3.5.18 Expected Cross Validation Index (ECVI)
The ECVI is an approximation of the goodness-of-fit an estimated model would achieve in another sample of the same size. The ECVI takes into account the actual sample size and the difference that could be expected in another sample. It also takes into account the number of estimated parameters for both the structural and measurement models (Hair et al 1995). The ECVI is used to compare alternative models and the objective is to minimise the values.

Table 7.3 provides a summary overview of the fit measures, categorised by type, and used to assess the models in chapter 9 to judge the final models.

<table>
<thead>
<tr>
<th>Absolute Fit Measures</th>
<th>Levels of Acceptable Fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-square ($\chi^2$) significance level (p value):</td>
<td>Statistical test of significance. Significance level. (Good fit if p-value for $\chi^2$ is greater than .05 or .01).</td>
</tr>
<tr>
<td>Noncentrality parameter (NCP)</td>
<td>Stated in terms of respecified $\chi^2$, judged in comparison to alternative model. Lower values in alternate of models.</td>
</tr>
<tr>
<td>Scaled Noncentrality parameter (SNCP)</td>
<td>NCP stated in terms of average difference per observation for comparison between models. Lower values in alternate models</td>
</tr>
<tr>
<td>Goodness-of-Fit index (GFI)</td>
<td>Higher values indicate better fit: close to .9 or above acceptable.</td>
</tr>
<tr>
<td>Root mean square residual (RMSR)</td>
<td>Stated in terms of input matrix, with acceptable level set by researcher of .140</td>
</tr>
<tr>
<td>Root mean square error of approximation (RMSEA)</td>
<td>Average difference per degree of freedom expected to occur in the population, not the sample. Acceptable values under .08.</td>
</tr>
<tr>
<td>Expected cross-validation index (ECVI)</td>
<td>The goodness-of-fit expected in another sample of the same size. used in comparing models. Lower values in comparing alternate models.</td>
</tr>
</tbody>
</table>
Table 7.3 Continued

<table>
<thead>
<tr>
<th>Absolute Fit Measures</th>
<th>Levels of Acceptable Fit</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Incremental Fit Measures</strong></td>
<td></td>
</tr>
<tr>
<td>Tucker-Lewis index (TLI or NNFI)</td>
<td>Recommended level .9 or higher</td>
</tr>
<tr>
<td>Normed-fit-index (NFI)</td>
<td>Recommended level .9 or higher</td>
</tr>
<tr>
<td>Adjusted Goodness-of-fit index (AGFI)</td>
<td>Recommended level .9 or higher</td>
</tr>
<tr>
<td><strong>Parsimonious Fit Measures</strong></td>
<td></td>
</tr>
<tr>
<td>Parsimonious Goodness-of-Fit index (PGFI)</td>
<td>A respecification of the GFI with higher</td>
</tr>
<tr>
<td></td>
<td>values reflecting greater model parsimony.</td>
</tr>
<tr>
<td>Chi-square/degrees of freedom ratio ((\chi^2/df))</td>
<td>(\chi^2) divided by the degrees of freedom, lower</td>
</tr>
<tr>
<td></td>
<td>limit 1; upper limit 5.</td>
</tr>
<tr>
<td>Degrees of freedom (df)</td>
<td>The number of nonredundant covariances in</td>
</tr>
<tr>
<td></td>
<td>the input matrix minus the number of</td>
</tr>
<tr>
<td></td>
<td>estimated coefficients.</td>
</tr>
<tr>
<td>Parsimonious normed fit index (PNFI)</td>
<td>Higher values close to 1 indicate better fit,</td>
</tr>
<tr>
<td></td>
<td>used in comparing alternate models.</td>
</tr>
<tr>
<td>Akaike information criterion (AIC)</td>
<td>Smaller positive values indicate greater</td>
</tr>
<tr>
<td></td>
<td>parsimony, used to compare alternate models.</td>
</tr>
<tr>
<td>Incremental Fit Index (IFI)</td>
<td>Values .8 and above indicate good fit</td>
</tr>
<tr>
<td>Relative Fit Index (RFI)</td>
<td>Values .8 and above indicate good fit</td>
</tr>
<tr>
<td>Comparative Fit Index (CFI)</td>
<td>Values .8 and above indicate good fit</td>
</tr>
</tbody>
</table>

7.3.5.19 Evaluation of the specific paths & robustness of the models

Apart from using basic fit indices to evaluate the fit of a model, a number of other approaches are available to test paths and the robustness of a model. Modern, computationally intensive statistical methods provide a completely different approach to tests of goodness-of-fit and parameter estimates. One such approach is to bootstrap models, which treats a random sample of data as a substitute for the population and resamples from it a specified number of times to generate sample bootstrap estimates and standard errors and confidence intervals. The bootstrapped estimates and standard errors are averaged and used to obtain a confidence interval around the average of the bootstrap estimates. The bootstrap estimator and the associated confidence interval are used to determine how stable or good the sample statistic is as an estimate of the population parameter (Mooney & Duval 1993; Schumacker & Lomax 1996 and West, Finch & Curran 1995). Measures appropriate within bootstrapping include confidence interval estimates,
standard error estimates and the Bollen and Stine (1992) test of the correctness of a model. These will be discussed next.

7.3.5.20 Bootstrapped standard errors

Bootstrapping in AMOS produces estimates of standard errors for the parameter estimates via the bootstrap algorithm of Efron (1982). For example, the bootstrapped estimate of the standard error (S.E) of a path coefficient is simply the standard deviation of the parameter estimate across a large (500 or more) number of bootstrapped samples (AMOS User’s guide 1997).

7.3.5.21 Bootstrapped confidence intervals

The development of confidence intervals around population parameters is the most common method of using an estimate of a statistic’s sampling distribution to make inferences. The traditional interpretation of confidence intervals are that they contain most of the values of θ that could have generated the random sample of size n, given the actual data (Mooney & Duval 1993). AMOS can produce bootstrapped confidence intervals for all parameter estimates. An appropriate method appears to be bootstrapped confidence intervals using the bias corrected method at a 95% level (Efron 1987).

7.3.5.22 Bollen-Stine bootstrapping

The Bollen and Stine (1992) procedure provides a test of the hypothesis that the proposed model is correct. This is the same null hypothesis that is conventionally tested via the chi-square test of fit. The objective of the Bollen and Stine procedure is to ascertain the probability that the discrepancy function would be as large as it actually turned out to be in the current sample, under the hypothesis that the proposed model is correct. A transformation of the sample data is carried out, so as to make the proposed model fit the transformed data exactly. Bootstrap samples are drawn from the transformed sample data and the distribution of the discrepancy function across the samples is then taken as an estimate of its distribution under the hypothesis that the proposed model is correct. The Bollen and Stine (1992) procedure provides a value that permits one to determine whether the data depart significantly from the model at a conventional significance levels (AMOS User’s guide 1997).
The Bollen-Stine estimate will be used to fit the final and best-fitting model only. The selection of the best-fitting model to run bootstrapping and the Bollen-Stine procedures is chosen because it is a very time-intensive and computer intensive procedure, particularly for large models like the model developed in this study (AMOS User’s guide 1997 and Bollen & Stine 1992).

7.3.6 Step 6: Interpret and modify if necessary the model
The final step proposed by Hair et al (1995) is interpreting and modifying the proposed models. This step involves examination of the diagnostic tools available in AMOS that may indicate potential ways to modify the model to improve fit. In this study, this issue will be examined in the conclusion to chapter 9 and will only be undertaken if there is theoretical justification for modification of the proposed model. Modification of relationships can be performed by analysing the modification indices computed by AMOS and examining the calculation of critical ratios for paths.

7.4 Conclusion to rationale and background to analysis
In structural equation modeling, estimation of a series of separate, but interdependent, multiple regression equations is achieved simultaneously by specification of a structural model. This process requires formal derivation through theoretical arguments in order to produce a testable structural model which distinguishes which independent variables predict/cause the dependent variables.

Structural equation modeling also has the ability to incorporate hypothetical or latent constructs as part of the modeling and analysis process. For example, within this study, involvement is a latent construct because it is not directly observed. It is reflected in responses to the observed variables that are directly measured in the survey instrument.

This chapter has sought to develop a sound basis on which to develop and estimate the structural equation models. It has covered the fundamental procedures for assessing the adequacy and quality of the measures and for developing and fitting congeneric models and discussed the key measures of fit to be used in evaluating the hypothesised models in a two-
stage modeling process. The use of structural equation modeling is appropriate for this study and the views espoused by Hulland, Chou and Lam reinforce this when they stated that:

Causal modeling in marketing . . . is now recognized as a general approach for integrating the theory-construction phase of research with the empirical and hypothesis-testing stages (Bagozzi 1982). . . . and . . . By combining and confronting theory with data (Fornell, 1982), and by forcing researchers to be explicit about both their measurement and theoretical assumptions (Bagozzi, 1980; Bagozzi, 1984), causal models can yield new insights (Hulland, Chow and Lam 1996 p.196)

The next chapter 8 uses the staged process as outlined by Hair, Anderson, Tatham and Black (1995) and Hulland, Chow and Lam (1996) to assess the measures and to estimate the initial factor structure and reliabilities of the constructs to assesses the quality of the measurement models before proceeding to stage 2 of estimating the causal models (Hulland, Chow and Lam 1996) that were hypothesized in chapter 5.