CHAPTER 5

RESEARCH DESIGN

5.1 INTRODUCTION

Most people have responded to so many questionnaires in their lives that they have little doubt about their ability to construct their own. However, very often, such confidence is misplaced. Frary (2000:1) of the Virginia Polytechnic Institute and State University believes that one reason for this phenomenon may be that many of the questionnaire designs in current use have deficiencies, which are consciously or unconsciously incorporated into new questionnaires by inexperienced developers. Another likely cause is inadequate consideration of aspects of the questionnaire process, which is separate from the instrument itself (such as how the responses will be analysed to answer the related research questions or how to account for non-returns from mailed questionnaires).

The design of a questionnaire is one of the most challenging elements for both students and professionals in research. Ambrose and Anstey (2001:1) believe that while there is a vast array of literature on the correct wording and sequencing of questions, the informational content of questions has been virtually ignored. The current level of research entails the use of very precise tools of analysis, but very ill-defined processes of research. According to Kinnear and Taylor, as quoted by Ambrose and Anstey (2001:1), questionnaire design is 'more an art form than a scientific undertaking. No steps, principles, or guidelines can guarantee an effective and efficient questionnaire'.

The aim of this chapter is to define and describe the research design pertaining to this study with particular emphasis on data collection and questionnaire/research design.

5.2 THE FRAMEWORK FOR QUESTIONNAIRE DESIGN

The overall framework for questionnaire design is depicted in Figure 5.1 (Gendall, 1998:1). A pyramid represents the framework, with the general principles at the top and specific principles at the bottom. At the apex of the pyramid is the concept of respondent orientation, and at the base, specific principles of question wording and graphic design.

The concept, which this representation is intended to convey, is that there are a small number of general principles of questionnaire design, which broaden out into a larger number of specific principles. Gendall (1998:3) purposely divides the pyramid into general and specific principles to illustrate the contention that much of what is written about questionnaire design starts at the level of specific principles. This concept does not exclude the notion that questionnaire design has a broader conceptual framework, but rather suggests that, if it has, that broader framework is generally assumed or implicit.

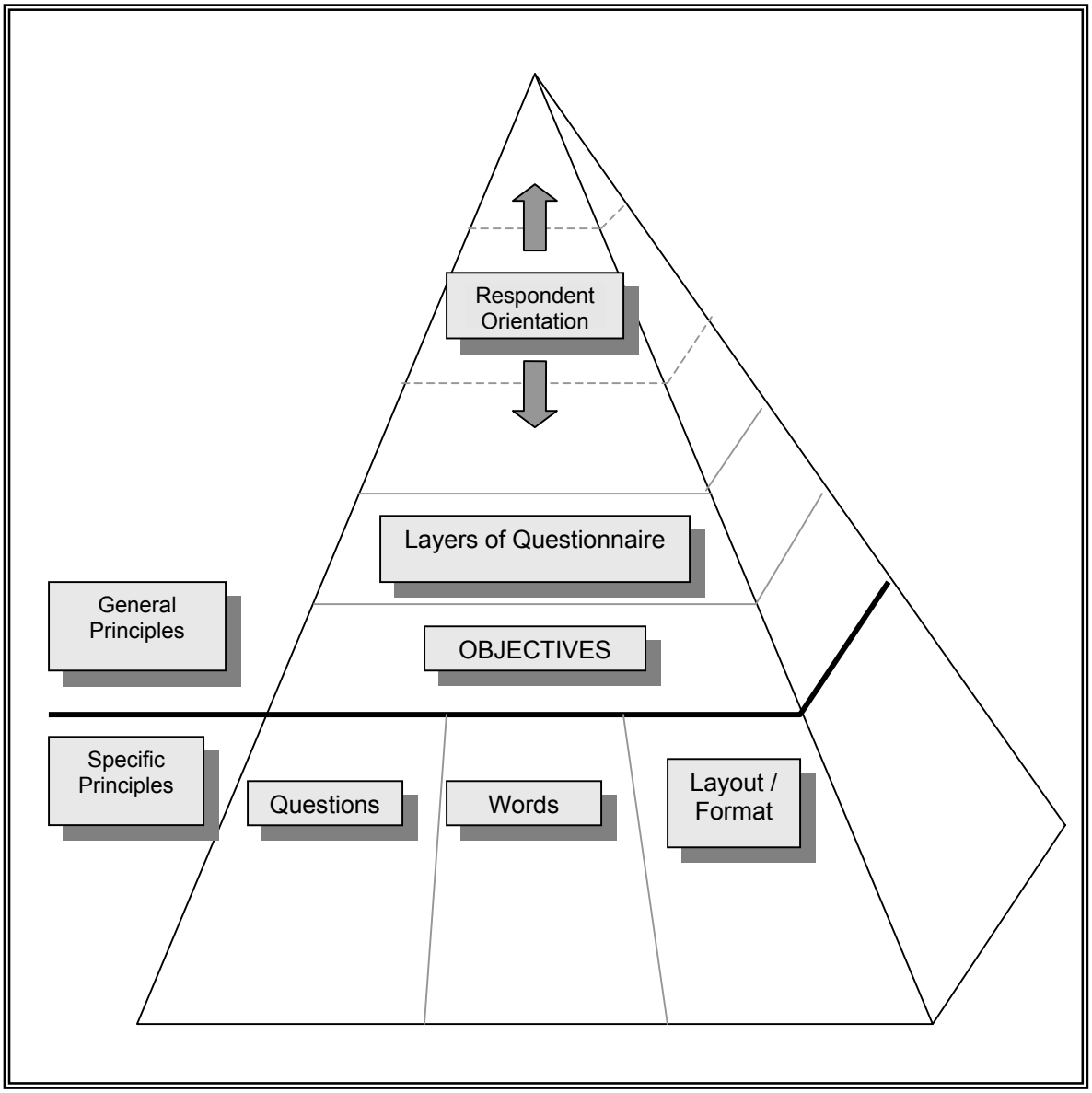
5.2.1 General principles

Gendall (1998:3) argues that the fundamental principle of questionnaire design is that the respondent defines what the researcher can accomplish. In other words, the target respondent will determine the type of questions a researcher can ask; the types of words the researcher can reasonably use; the concepts which may be explored; and the methodology that can be employed. For this reason, a survey aimed at aviators will be quite different from one that is aimed at the general public.

Gendall (1998:4) is of the opinion that in order to find out what is in respondents' minds, one needs to ask questions that can be truthfully answered about their physical environment, their consciousness, their knowledge, and their past behaviour. However, this is a contentious position as it means that attitude and opinion questions play only a minor role in questionnaire design. For the purposes of this study, it may be argued that Genadall's (1998) opinion is mostly true for questionnaire design aimed at marketing surveys and that opinion questions play a significant role in questionnaire design aimed specifically at gaining information with regard to respondents' attitudes and opinions.

Implicit in Gendall's contention is the assumption that the objective of most surveys is the prediction of human behaviour. However, a great deal of attitude and opinion research is conducted with no behavioural implications in mind; the measurement of attitudes and opinions is often simply regarded as an end in itself. Gendall (1998:4) believes that not all attitudes and opinions are necessarily of equal value and that information on respondents' environment, consciousness and knowledge can be used to weight their opinions to give a more realistic perspective on the views of the sampled population.

Figure 5.1: Framework for questionnaire design



Source: Gendall (1998:4)

The term 'environment' relates to the physical aspects of respondents' lives over which they have little control, but which impinge on their ability to act or respond in specific ways. These factors include age, gender, socio-economic status, race, locale and mobility. Respondent consciousness determines whether or not respondents can understand the implications of their answers; in other words, whether they fit the pieces together to form a coherent idea. The concept of the environment also emphasises the importance of past behaviour as a predictor of future behaviour, as respondents are often better able to discuss what they have experienced than what their actions might be.

All questionnaires reflect the author's view of the world to some extent, regardless of how objective the researcher has attempted to remain. Intellectually, good questionnaire designers understand this and attempt to maintain a detached objectivity (Gendall, 1998:4).

Finally, the questionnaire is not only a series of questions, nor is a question simply a series of words; it is a structure consisting of several different layers that must be simultaneously integrated into an overall picture.

5.2.2 Specific principles

Specific questionnaire design principles are classified into three sections, namely question design, question wording and formatting or layout. As with all aspects of questionnaire design, these elements cannot be dealt with in isolation and each has a bearing on the others (Gendall, 1998:4).

- Questions: Good questions produce answers that are reliable and valid measures of the item of interest. Poor questions obscure, prohibit or distort the communication from the respondent to researcher, and *vice versa*.

- Words: Question wording variations generally have little impact on the stability of survey results. Variations become significant when they introduce or tap a different concept or reality or emotional level surrounding an issue.

- Layout: Questionnaires should be designed to make the task of following instructions, reading questions and recording answers as easy as possible for both interviewers and respondents.

5.3 PRELIMINARY CONSIDERATIONS

Many questionnaires give the impression that the creator of the questionnaire imagined every conceivable question that might be asked in respect of the topic in question. Alternatively, a committee may have incorporated all of the questions generated by its members. Frary (2000) is of the opinion that such approaches should be avoided, as they tend to yield very long questionnaires, often with many questions relevant to only small proportions of the sample, resulting mostly in annoyance on the part of many respondents.

The added time it takes to complete the questionnaires, as well as the belief that responses are unimportant if many of the questions are inapplicable, will result in incomplete and/or inaccurate responses, as well as the non-return of mail items. These difficulties can yield largely useless results.

Frary (2000:2) suggests the following to avoid these kinds of problems:

5.3.1 Exercise mental discipline

The investigator should define precisely what information is desired and should endeavour to write as few questions as possible to obtain it. Peripheral questions and questions that find out 'something nice to know' should be avoided. The compiler of the questionnaire should also consult colleagues about the results in this process.

To this end, in this study, the researcher identified four key areas of interest in the design of the AGAQ (Aviation Gender Attitude Questionnaire). They are the following: 'Learning Ability and Speed'; 'Piloting Skills'; 'Leadership and Decision-Making'; and 'General Prejudices and Stereotypes'. After the preliminary questionnaire had been reviewed several times by a panel of human factor experts, irrelevant and faulty items were eliminated (the researcher also invited several pilots to complete and comment on the questionnaire). The final questionnaire contains a total of 72 attitude questions – 18 questions in each of the above-mentioned categories.

5.3.2 Obtain feedback from a small but representative sample of potential respondents

This step may involve no more than informal, open-ended interviews with several potential respondents. However, it is better to ask such a group to criticise a preliminary version of the questionnaire. In this case, they should first answer the questions as if they were research subjects. The purpose of these activities is to determine relevance of the questions and the extent to which there may be problems in obtaining responses.

The above process should not be confused with a field trial of a tentative version of the questionnaire. Field trials are also desirable in many cases but have different purposes and should always follow the more formal review processes. A field trial is desirable and/or necessary if there is substantial uncertainty in areas such as the following:

5.3.2.1 Response rate

If a field trial of a mailed questionnaire yields an unsatisfactory response rate, design changes or different data gathering procedures must be undertaken.

5.3.2.2 Question applicability

Even though approved by reviewers, some questions may prove redundant. For example, the greatest majority of respondents may answer alike in a certain answer category, thus suggesting that they deem the question to be unnecessary. The process of designing the measurement instrument should be in accordance with the research problem, propositions, primary and secondary research objectives and the different measurement aspects.

According to Dillon, Madden and Firtle (1993:302), a researcher should translate the research problem into a set of research questions before formulating the questions. The research questions should identify:

- what information is required;
- who the appropriate target responses are; and
- what data collection method to use.

5.3.2.3 Question performance

The field trial response distributions for some questions may clearly indicate that they are defective. Also, pairs or sequences of questions may yield inconsistent responses from a number of trial respondents, thus indicating a need to change the response mode (Frary, 2000:2).

5.4 CLASSIFICATION MODEL

There are two types of data source, namely primary and secondary data sources (Cooper & Schindler, 1998:256). Primary data is original data collected specifically for the purpose of the research question. Researchers gather secondary data for their own purposes (which may be along the lines of the research in question). Secondary data may be obtained from internal organisational sources, or from external resources.

For the purposes of this study, primary data can be defined as the knowledge obtained from the attitudinal part of the questionnaire, *vis-à-vis* Questions 1 to 72 in Section II of the AGAQ (see Appendix F). Secondary data is identified as the biographical/demographic questions: Questions 1 to 13 of Section I of the AGAQ.

In Ambrose and Anstey's (2001) review of questionnaire study instruments at the University of Nebraska at Omaha, they found that data- and information-gathering techniques could be identified and classified in six distinct categories. These categories are inclusive. However, not all studies necessarily have elements of each of the six categories (Ambrose & Anstey, 2001:1).

5.4.1 Demographics

Demographics can be broadly defined and include measures that go beyond the components of age, gender and ethnic origins. The expanded definition of demographics includes the number of automobiles owned, the frequency of ATM usage in the family and other elements that might be argued as an extension of the concept of demographics. However, the issue is not the definition of demographics, but the inclusion of *appropriate* demographic measures in study instruments.

In the AGAQ, the researcher found it important to include in the demographic section questions eliciting data pertaining to the types and ratings of the pilots completing the questionnaire. These items are evident in Questions 5 to 13 of Section I (see Appendix F).

5.4.2 Attitudes, opinions, values and beliefs

Questionnaire design often includes an attempt to assess the attitudes and values of the respondents. One generally finds a scattering of attitude elements that probe the perceptions and preferences of respondents. Perception and preference assessments are challenging; defy verification and are somewhat vague. However, with careful composition of questions, attitude assessments provide insights that are extremely important and disclose critical information.

The AGAQ endeavours to identify and categorise any stereotypes, attitudes or prejudices that may or may not exist with regard to female aviators. To this end, the AGAQ relies heavily on the opinions, beliefs, values and attitudes of its respondents.

5.4.3 Behaviours and experiences

Understanding how individuals have behaved and currently behave provides a foundation that neither attitudes nor demographics disclose. Eliciting information about behaviour and experiences is particularly useful in marketing questionnaires. As previously noted, past behaviour is the best predictor of future behaviour.

With regard to this study, experience plays a significant role in determining respondents' attitudes towards female aviators. For example, an instructor pilot may hold valuable insights into positive or negative traits or patterns displayed by each of the genders during initial flight training.

5.4.4 Knowledge

Knowledge questions can provide a direct method of assessing the effectiveness of advertising or how the impact of an event might have damaged a product, service, person or organisation's image.

5.4.5 Predispositions and intentions

Research designs may, for example, include inquiries about brand loyalty or colour preferences. Research involving purchasing agents may measure the forecasted volume of purchases or the expansion of a product line. As compared to behaviours that document prior events, questionnaires eliciting information about predispositions and intentions are focused on assessing future events and behaviours.

Many of the questions in the AGAQ examine the comfort level of respondents with regard to ideas involving women in particular aviation roles, for example, issues addressing women in combat aviation.

5.4.6 Administrative codes and controls

For general purposes, codes and controls appear with some frequency in questionnaire design. They are subpopulation identification, administrative dates, surveyor, and survey respondent identification codes. Usually there is a requirement to include embedded coding in a questionnaire, interview form, or even the instrument used to collect

observations. For example, if one were surveying a known but different subpopulation, a code would provide information needed to distinguish the returns between the subpopulations.

The above classifications are not formally considered in the design of the research instrument, but are unconsciously developed. The framing of questions is still demanding, but the classifications prompt the author to be more inclusive regarding categories (Ambrose & Anstey, 2001:2).

5.5 DATA COLLECTION METHODS

The nature of research can be either qualitative or quantitative. According to Malhotra (1996:164), qualitative research is an unstructured, exploratory research method based on small samples intended to provide insight and understanding about the problem statement. Quantitative research involves the collection of primary data from a large number of individuals, frequently with the intention of projecting the results onto the larger population (Martins *et al*., 1996:125).

There are various methods of collecting primary research data, namely mail-based self-administered questionnaires, telephone interviews, personal interviews, and focus groups. This study has also relied heavily on the use of electronic mailing (email) and internet-based submissions.

Dillon *et al*. (1993:158) lists the following factors that should be considered in the selection of the best survey method (these factors are also depicted in Table 5.1):

- Versatility: Versatility refers to the extent to which the survey method can handle question formats and scenarios.
- Quantity of data: This refers to the amount of data that can be collected.
- Sample control: Sample control is the ease or difficulty of ensuring that desired respondents are contacted.
- Quality of data: Quality of data refers to the accuracy of the data collected using a particular data-collection method.
- Response rate: The number of responses, divided by the sample size, calculates the response rate.

- Speed: Speed refers to the total time that it takes to complete the study by using a particular data-collection method.
- Cost: This refers to the cost per completed questionnaire.
- Uses: Uses refer to how the collected data will be used.

Table 5.1: Summary of data collection methods

CRITERIA	DIRECT MAIL/EMAIL	TELEPHONIC	PERSONAL
Versatility	Not much	Substantial but complex or lengthy scales are difficult to use	Highly flexible
Quantity of data	Substantial	Short, lasting typically between 15 and 30 minutes.	Greatest quantity
Sample control	Little	Good, but non-listed respondents can pose a problem	In theory, provides greatest control
Quality of data	Better for sensitive or personal questions, but no interviewer present to clarify question	Interviewers can clear up ambiguities, but their presence may lead to socially accepted answers.	Possibility of cheating
Response rate	In general low (± 10%) Email: 60 – 70% *	60 – 80%	Greater than 80%
Speed	Several weeks Email: Relatively fast*	Large studies can be completed in 3 to 4 weeks	Faster than mail, slower than telephone interviews
Cost	Inexpensive	Depends on incidence rate and length of survey	Expensive, but considerable variability
Uses	Executive, industrial, medical, etc.	Ineffective in studies that require national samples	Prevalent in studies requiring visual cues, etc.

* This reflects the opinion of the researcher with regard to this study and has not been scientifically tested.

Adapted from Dillon *et al*. (1993:173)

5.6 MEASUREMENT AND MEASUREMENT SCALES

Measurement is the process of assigning numbers to objects to represent quantities of attributes (Dillon *et al*., 1993:302). Measurement relates to the procedure used to assign numbers that reflect the amount of an owned attribute.

5.6.1 Level of measurement

Measurement can be undertaken at different levels. The levels reflect the correspondence of numbers assigned to the characteristics in question and the meaningfulness of performing mathematical operations on the numbers assigned. Levels of measurement include:

5.6.1.1 Nominal measurement

Nominal measurement is the process whereby the numbers assigned allow the researcher to place an object in one set of mutually exclusive and collectively exhaustive classes with no implied ordering (Dillon *et al*., 1993:273).

5.6.1.2 Ordinal measurement

Ordinal measurement is the process in which the response alternatives define an ordered sequence so that the choice listed first, is less (or greater) than the second, the second less (or greater) than the third, and so on (Dillon *et al*., 1993:274). The number assigned does not reflect the magnitude of an attributed possess by an object.

5.6.1.3 Interval measurement

Interval measurement allows the researcher to indicate how far apart two or more objects with respect to the attribute, and consequently to compare the differences between the assigned numbers (Dillon *et al*., 1993:275). As the interval lacks natural or absolute origin, the absolute magnitude of the numbers cannot be compared.

5.6.1.4 Ratio measurement

Ratio measurement has the same properties as interval scales, but also has a natural and absolute origin (Dillon *et al*., 1993:277).

5.6.2 Scale types

Measurement scales fall into two broad categories: comparative and non-comparative scales.

5.6.2.1 Non-comparative scaling

Non-comparative scaling is a method whereby the respondent is required to evaluate each object on a scale independently of the other objects being investigated. According to Dillon *et al*. (1993:277), the following types of non-comparative scaling can be used:

- Line marking/continuous rating scales

  This is a procedure that instructs the respondent to assign a rating by placing a marker at the appropriate position on a line that best describes the object that is being investigated. There is no explicit standard for comparison.

- Itemised rating scales

  With itemised rating scales, the respondent is provided with a scale with numbers and/or brief descriptions associated with each category and is asked to select one of the limited number of categories, ordered in terms of scale position, that best describes the object under investigation.

  Dillon *et al*. (1993:278) believe that when itemised rating scales are utilised, the researcher must have clarity on the following issues:

  - The number of categories
    The researcher may choose to include any number of response categories, provided that the respondents have to discriminate among alternatives. The researcher may include between five and nine response categories.

- The number of favourable and unfavourable categories
  When a balanced scale is used, the scale has an equal number of favourable and unfavourable categories. When unbalanced scales are used, the scales have unequal numbers of favourable and unfavourable categories.

- The nature and degree of verbal description
  Verbal category descriptors help to ensure that each respondent is operating in the same paradigm. Pictures and other forms of graphic representations can be successfully utilised when the respondents are children, or when illiteracy levels are high among the respondents.

- The presence of a neutral position
  In odd numbered scale items, the middle scale usually becomes the neutral point.

- Forced and unforced itemised rating scales
  With forced itemised rating scales, the respondent is required to indicate answers even if he/she has no opinion or knowledge about the subject. For this kind of rating scale to be successful, it is of great value for the respondent to have knowledge of or an opinion on the topic.

5.6.2.2 Comparative scaling

Comparative scaling is a process whereby the respondent is asked to compare a set of stimulus objects directly against one another. According to Dillon *et al*. (1993:281), the following types of scaling can be used:

- Paired comparisons scale

  This is a scale where the respondents are provided with two objects at a time and the respondents are asked to select one of the two according to some criterion.

- Geared paired comparisons

  This form of scale is an extension of the paired comparison method. Respondents are asked for their preference and the extent to which they prefer their choice.

- Rank order scales

  Rank order scales are scales where respondents are presented with several objects simultaneously and requested to order or rank them. Conditional rank order scaling is the process whereby respondents consider each object in turn as a standard for comparisons. Respondents assign ranks to other objects according to this standard (Dillon *et al*., 1993:282).

- Constant sum scales

  Respondents are asked to allocate a number of points by choosing between alternatives according to the same criterion, for example, importance or preference.

- Line marking/continuous rating comparative scales

  This is the process whereby respondents are presented with object pairs and the respondents are asked to judge their similarity by placing a mark on a continuum.

5.6.3 Single-item versus multiple-item scales

After deciding on a scale type (or a combination thereof), the researcher should decide whether to use single-item or multiple-item scales (or a combination thereof). A multiple-item scale usually consists of a number of statements that the respondent must react to, for example, how favourable or unfavourable their opinion of an item is. Multiple-item scales are usually utilised in the measurement of attitude surveys (Dillon *et al*., 1993:288). Three multiple-item scales can be identified.

5.6.3.1 Semantic differential scales

This is a technique where a measure of the person's attitude is obtained by rating the object or behaviour in a question on a set of bipolar adjective scales (Dillon *et al*., 1993:289). The semantic differential scale measures the psychological meanings of an attitude object.

5.6.3.2 Staple scales

A staple scale is a procedure using a single criterion or key word(s) and instructing the respondents to rate the object on a scale. A staple scale is used as an alternative to the semantic differential scale, especially when it is difficult to find bipolar adjectives that match the investigation item (Dillon *et al*., 1993:290).

5.6.3.3 Summated scales

The Likert scale is the most frequently used variation of the summated rating scale and the most popular choice for surveys. The Likert scale is a scale consisting of a number of evaluative statements (Dillon *et al*., 1993:292).

Summated scales consist of statements that express either a favourable or an unfavourable attitude toward the item in question. The respondent is required to agree or disagree with each statement. Each response is given a numerical score to reflect its degree of attitudinal approving. Likert scales aid researchers in comparing individuals' scores with the distribution of scores from a well-defined group.

A five-point Likert (interval) scale was utilised in this study in order to determine respondents' opinions on a variety of items.

5.7 WRITING EFFECTIVE QUESTIONS

Accurate and complete feedback from respondents is what ensures the success of any research, but ensuring that a study returns valid, unbiased results is often easier said than done. Leading phrases, inappropriate questions and skewed designs can result in pre-empted and inaccurate results. In order to achieve outcomes that can be confidently applied, care needs to be taken when structuring the questions, as well as in administering them (the manner in which they are asked) and how they are measured once they have been received.

5.7.1 Formulating questions

Dillon *et al*. (1993:303) provide two general guidelines for devising effective questionnaires:

- A researcher should write specific questions only after thoroughly researching the objectives and research propositions.
- For each question posed, consideration should be given to how the information obtained from the responses will assist in answering the research propositions.

There are a number of specific considerations than need to be borne in mind when developing questions. Dillon *et al*. (1993:304) suggest the following basic principles:

- Principle 1: Be clear and concise.
- Principle 2: Response choices should not overlap.
- Principle 3: Use natural and familiar language.
- Principle 4: Do not use words or phrases that show bias.
- Principle 5: Avoid double-barrelled questions.
- Principle 6: State explicit alternatives.
- Principle 7: Questions should meet the criteria of validity and reliability.

The most important issue is whether or not a researcher can truly measure what he/she is attempting to measure and whether or not the responses can be replicated at a later stage.

5.7.2 Asking a good question

The foundations of any questionnaire are good, clear unambiguous questions. These will be easier to formulate if the questioner can answer the following:

- Will the respondent be able to understand the questions?
- Having understood the question, will respondents be willing to answer it?
- Provided he/she has understood the question and is willing to answer the question, will the respondent be able to answer it in a way that accurately reflects his/her feelings?

5.7.3 Understanding the question

Obscure technical terms that confuse respondents, the use of imprecise words, abstract concepts or trying to ask two questions at the same time serves to create misunderstanding (Sheward, 2002:1). Often respondents who do not understand the questions are unwilling to ask for clarification and avoid answering the question. Many

questionnaires present a range of possible legitimate answers and encourage responses even where the question is meaningless to the respondent.

5.7.4 Willingness to answer the question honestly

Sheward (2002:1) believes that more embarrassing questions yield more accurate responses if they are administered remotely (for example, by means of direct mail as opposed to face-to-face). In addition, respondents are becoming more sophisticated in their ability to detect leading or biased questions that seem to be trying to answer for the respondents. Respondents may refuse to answer these questions.

The desire or pressure to give 'socially acceptable' answers often plays a part in some less than honest responses. In some subject areas (such as politics) and especially in face-to-face interviews where an answer may be overheard by others, questions on political or moral issues might elicit a response more in keeping with what are perceived to be acceptable norms than the respondent's true opinion.

In compiling a questionnaire in this study, the author of this research was aware that some respondents might favour 'politically correct' answers with regard to women in aviation. In order to overcome this predicament, the questionnaire clearly states that all respondents may remain anonymous.

5.7.5 Ability to answer accurately

Many complex questions can best be answered by inviting an open-ended statement, accurately recording the exact words used by the respondent. The problem with this type of answer is the virtual impossibility of analysing large numbers of such responses. In the majority of cases, researchers utilise scaling or a multiple-choice system. This is achieved by presenting respondents with a list of statements which the questionnaire designer feels adequately represent the range of legitimate answers. This type of questionnaire runs the risk of oversimplifying the issues involved and many respondents may find it difficult to choose an answer that accurately reflects their true opinions.

More often, rating scales are used that allow a full range of opinions to be applied to statements. For instance, respondents may be invited to choose from the following options:

- Agree a lot
- Agree a little
- Neither agree nor disagree
- Disagree a little
- Disagree a lot

While this is a valid scale, questionnaire designers should ensure that they include a 'don't know' option. It is not advisable to use a scale with more than seven points as it poses too many choices and causes confusion (Sheward, 2002:2).

5.7.6 Open-ended and close-ended questions

Two options are available to researchers in terms of question formats: open-ended and close-ended questions.

5.7.6.1 Open-ended question formats

With open-ended questions, the respondent is able to choose any response deemed appropriate. This occurs within the limits implied by the question. According to Dillon *et al*. (1993:310), there are several good reasons for asking open-ended questions:

- Open-ended questions are useful to check and/or corroborate the results of quantitative or close-ended questions.
- Open-ended questions may be used to obtain direct comparisons and to specify particular causes for preference or rejection when two or more stimuli are involved in a test.
- Open-ended questions are useful in determining whether a particular communication vehicle conveys its intended objectives.
- Open-ended questions are able to elicit a respondent's general reaction to or feelings toward a certain subject.

Open-ended questions are not well suited for self-administered questionnaires and answers to open-ended questions may be more of an indication of the respondents' knowledge about or interest in the issue being investigated.

Interview bias can be a serious problem with the use of open-ended questions and open-ended questions must be coded or categorised for analysis, which can be a tedious task, laden with ambiguities.

5.7.6.2 Close-ended question formats

With close-ended questions the respondents are provided with numbers and/or predetermined descriptions and is required to select the one that best describes their feelings. There are several issues related to the success of itemised question formats (Dillon *et al*. 1993:310):

- the number of response alternatives;
- the nature and degree of verbal description;
- the number of favourable and unfavourable categories;
- the statement of a neutral position; and
- the forced or unforced nature of the scale.

The advantages of close-ended question formats are:

- their ease of use in the field;
- their ability to reduce interview bias; and
- their ability to reduce bias based on differences in how articulate respondents are.

5.8 POTENTIAL SOURCES OF ERRORS IN RESEARCH DESIGN

The usefulness of the collected data and the data analysis depends on the overall quality of the research design. However, errors may occur in the research design that can influence the research process. Figure 5.2 depicts the types of errors that can affect research design. A discussion of total errors, random sampling errors and non-sampling errors follows.

5.8.1 Total error

Malhotra (1996:100) defines a total error as the total variation between the true mean value in the population of the variable of interest and the observed mean value obtained in a research project. A total error can be sub-divided into a random sampling error and a non-sampling error.

5.8.1.1 Random sampling error

A random sampling error occurs when a particular selected sample is an imperfect representation of the population of interest. A random sampling error may be defined as the variation between the true mean value for the sample and the true mean value of the population (Malhotra, 1996:102).

5.8.1.2 Non-sampling errors

Malhotra (1996:102) describes a non-sampling error as one that can be attributed to sources other than sampling and explains that it can be random or non-random. Non-sampling errors consist of response errors and non-response errors.

5.8.1.3 Response errors

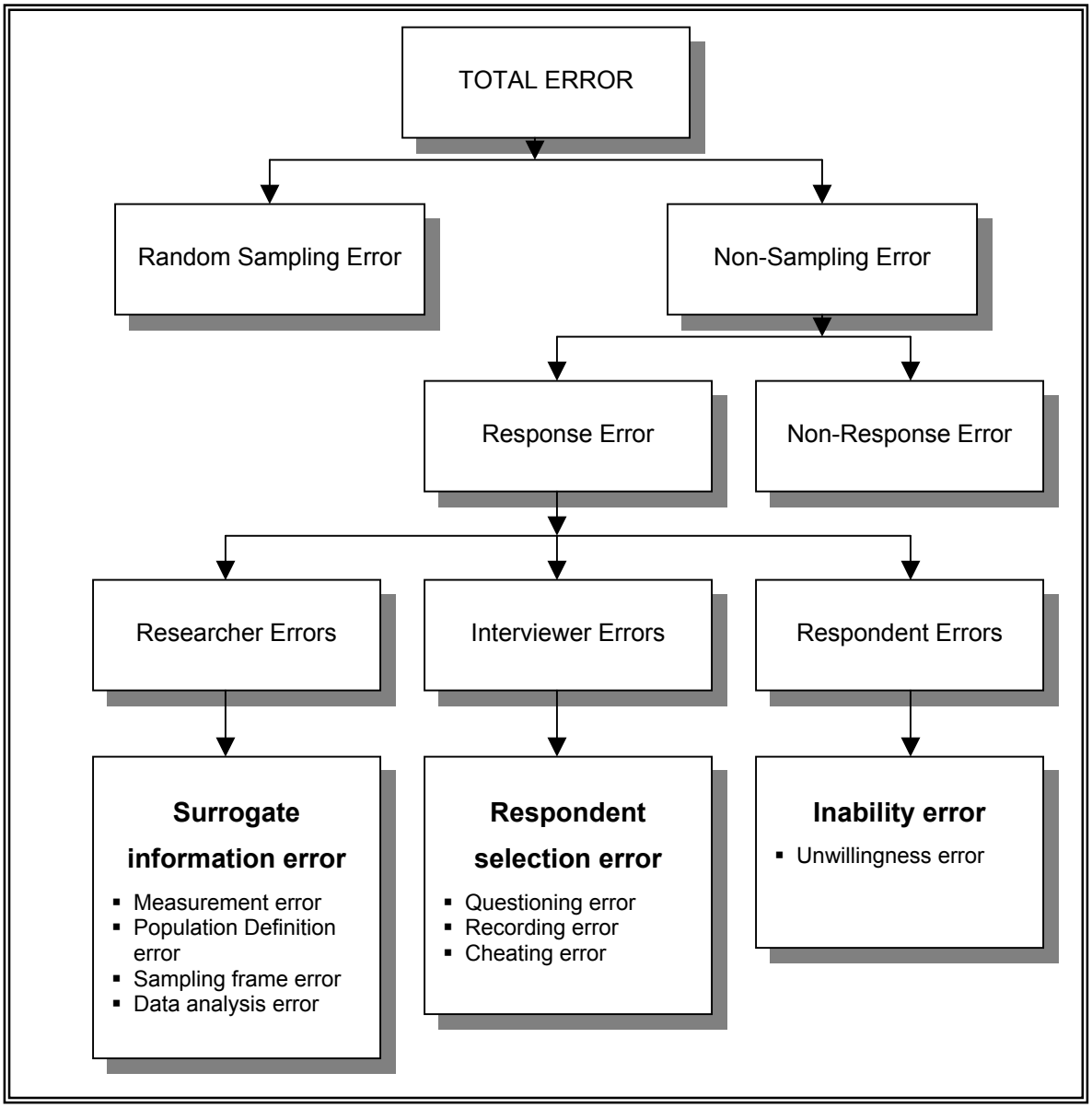
Malhotra (1996:102) describes a response error as the variation between the true value mean of the variable in the net sample and the observed mean value obtained in a research project. A response error is a non-sampling error arising from respondents who do respond but give inaccurate answers or whose answers are misrecorded or misanalysed. Researchers, interviewers or respondents can make response errors.

5.8.1.4 Non-response errors

A non-response error is the variation between the true mean value of the variable in the original sample and the true mean value in the net sample. A non-response error occurs when some respondents included in the sample do not respond. Non-responses cause the obtained sample to be different in size or composition from the original sample (Malhotra, 1996:102).

Errors in research design are set out in Figure 5.2 (overleaf).

Figure 5.2: Errors in research design



Source: Malhotra (1996:100)

5.8.2 Dealing with non-responses

According to Sudman and Blair (1999:275), there has been a steady decline in sample co-operation in the past 25 years. They are of the opinion that there is a broad range of reasons for this phenomenon, and that most cannot be controlled by the researcher. The question arises whether careful probability design methods are valid and useful if co-operation rates continue to drop. Sudman and Blair (1999:275) believe that high-quality samples will continue to be possible, but only with greater effort and cost. New methods

will be needed but will only be justified if they can significantly improve the quality of the information obtained.

On the basis of the above tendency, Sudman and Blair (1999:275) suggest the following:

- Make more contact attempts to locate respondents.
- Make greater use of mixed modes to obtain co-operation.
- Provide higher compensation for interviewers.
- Ensure intensive efforts to obtain samples of previous non-respondents so that better post-survey adjustments of data are possible.

5.9 VARIABLES

Although the above literature takes an overall look at research questionnaire design, a more intense examination is required with regard to defining and understanding the functions of variables. Without knowledge of variables, one cannot conduct very effective research.

5.9.1 Defining variables

According to Morgan and Griego (1998:1), variables must be able to vary or have different values, or, a variable is any entity that can take on different values: 'The concept "variable" can be defined as a characteristic of the participants or situation of a given study that has different values in that study. In quantitative research, variables are defined operationally and are generally divided into independent variables (active or attribute), dependent variables, and extraneous variables.'

A variable is a quantity that varies over different instances. In mathematics, variables are often denoted by letters (such as x and y in the equation $y = x + 3$). In this example, x and y represent a family of pairs of values, which satisfy the equation ($[x,y] = [0,3], [1,4], [2,5]$).In statistics, x might be the weight of a particular test subject. Repeated weighing of the same test subject may yield different values. If one uses $y = x + 3$ in a statistical sense, x denotes the subject to chance variation (as in the example above), and whatever value x assumes, 3 is added to obtain y. Y is therefore also subject to the same chance variation, but is still related to x (Braverman, 1997:1).

Thus, a variable is 'any stimulus factor or behaviour that can change on some dimension and that can be observed, sometimes controlled, and measured. In scientific research variables that can be quantified with precision are preferred' (Walsh, 2002:1).

The *operational* definition denotes a variable in terms of the operations or techniques used to obtain or measure it. When quantitative researchers describe the variables in their study, they specify what they mean by demonstrating how they measured a variable. Demographic variables are usually defined by asking respondents to choose an appropriate category from a list, while abstract concepts need to be defined operationally by defining in detail how they were measured in a particular study (Morgan & Griego, 1998:1).

5.9.2 Independent variables

An independent variable in an experiment is 'the variable under the control of the scientist/investigator and which is believed to have the potential to alter or influence the dependent variable' (Walsh, 2002:1).

Independent variables can be further categorised into active independent variables and attribute independent variables.

5.9.2.1 Active independent variables

According to Morgan and Griego (1998:2), active independent variables are often called *manipulated independent variables*. They are often used to investigate the effect of a particular intervention. An example of this would be the effects of an innovative therapy as compared to those of a traditional treatment.

In traditional experimental research, independent variables are variables that the investigator can manipulate as they seemingly cause a change in the resulting behaviour, attitude, or in the physiological measure of interest. An independent variable is considered to be manipulated or active when the researcher has the option to give one value to one group (experimental condition), and another value to another group (control condition).

However, Morgan and Griego (1998:2) note that often in applied research, one can have an active independent variable that is not manipulated by the researcher (for example,

where two comparative conditions use different stimuli, the researcher could compare the results without manipulating the variable).

5.9.2.2 Attribute independent variables

Morgan and Griego (1998:2) do not restrict the term 'independent variable' only to manipulated or active variables. They include any predictors, antecedents, or presumed causes or influences under investigation in the study. Attributes of the participants as well as active independent variables are included within this definition. A variable that cannot be given, yet is a major focus of the study, is called an attribute independent variable.

5.9.3 Dependent variables

Walsh (2002:1) defines a dependent variable as the response or behaviour in an experiment that is being studied in order to determine if it has been influenced by or altered by the independent variable. It is therefore the presumed outcome or criterion. Dependent variables are often test scores, ratings, readings from instruments, or measures of physical performance.

5.9.4 Extraneous variables

Extraneous variables are variables that are not of interest in a particular study, but that could influence the dependent variable. Environmental factors, for example, the temperature, the time of day, and the characteristics of the researcher are some possible extraneous variables that should be controlled (Morgan & Griego, 1998:5).

5.10 RESEARCH PRACTICES

Morgan and Griego (1998:5) identify five basic research practices and the criteria that distinguish them. They are represented in Table 5.2 and following is a brief discussion of them.

5.10.1 The randomised experimental method

In order for a research practice to be termed a *randomised experimental* method, two criteria must be met. The first is that the researcher must randomly assign participants to

groups and conditions (control and/or experimental). This criterion is what differentiates randomised (or true) experiments from quasi-experiments. The second criterion dictates that an independent variable must be active, as defined previously. In addition, the researcher is usually able to control the independent variable.

5.10.2 Quasi-experimental method

The quasi-experimental research method is similar to the randomised experimental method, but it fails to satisfy the condition of a random assignment of subjects to groups. Quasi-experimental methods have an active independent variable with a few values and also involve a comparison between, for example, an experimental and a control condition. Morgan and Griego (1998:5) warn researchers about the active independent variable: in the experimental method, the researcher usually has control over the independent variable in that one level can be randomly assigned to the experimental condition, and one level can be randomly assigned to the control condition. The strength of the quasi-experimental method is based on how much control the investigator actually has in manipulating the independent variable and deciding which group will receive which intervention. The strength of the design influences how confident the researcher can be about whether the independent variable was the cause of any change in the dependent variable.

5.10.3 Basic comparative method

The comparative research method differs from the two previous methods, as the researcher cannot randomly assign participants to groups and there is not an active independent variable. Like randomised experiments and quasi-experiments, comparative designs usually have a few levels or categories for the independent variable and make comparisons between the groups. Studies using the comparative method examine the presumed effects of attribute independent variables.

5.10.4 Basic associational method

This method is used where the independent variable is continuous or has several ordered categories, usually five or greater. Morgan and Griego (1998:7) cite the following example: a researcher is interested the relationship between giftedness and self-perceived confidence in children. Assume that the dependent variable is a self-confidence scale for children and the independent variable is giftedness. If giftedness has been divided into

high, average, and low groups, the research method would be deemed as comparative, as the logical approach would be to compare the groups. However, in the typical associational method, the independent variable is continuous or has at least five ordered levels or values. All participants would be in a single group with two continuous variables – giftedness and self-concept. A correlation coefficient could be performed to determine the strength of the relationship between the two variables.

5.10.5 Basic descriptive method

This method differs from the previous four methods in that only one variable is considered at a time, so that no relationships are made. Lack of comparisons or associations is what distinguishes this method. The basic descriptive method does not meet any of the other criteria such as random assignment of participants to groups.

Morgan and Griego (1998:8) restrict basic descriptive methods to questions and studies that use only descriptive statistics, such as averages, percentages, histograms, and frequency distributions, and do not test null hypotheses with inferential statistics.

5.10.6 Complex research methods

Many research studies are more complex than implied by the previous approaches and almost all studies have more than one hypothesis or research question and may utilise more than one of the previous methods. Morgan and Griego (1998:8) believe that it is common to find a study with one active independent variable and one or more attribute independent variables. This type of study combines the randomised experimental method (if participants were randomly assigned to groups) and the comparative method. Most 'survey' type studies also have some descriptive questions; so it is common for published studies to use three or often even more of the methods.

Table 5.2 (overleaf) sets out a comparison of five basic research methods.

Table 5.2: Comparison of five basic quantitative research methods

Criteria	Randomised experimental	Quasi-experimental	Comparative	Associational	Descriptive
Random assignment of subjects to groups by researcher	√ Yes	X No	X No	X No (only one group)	X No groups
Independent variable is active	√ Yes	√ Yes	X No (attribute)	X No (attribute)	X No (independent variable)
Independent variable is controlled by the researcher[1]	Usually	Sometimes	X No	X No	X No
Independent variable has only a few levels/values[2]	√ Yes	√ Yes	√ Yes	X No[2] (many)	X No (independent variable)
Relationships between variables (comparison of groups or association of variables)	√ Yes (comparison)	√ Yes (comparison)	√ Yes (comparison)	√ Yes (association)	X No

[1]Although this is a desired quality of experimental and quasi-experimental designs, it is not sufficient to distinguish between the experimental and quasi-experimental methods.

[2]This distinction is made for heuristic/educational purposes and is only 'usually' true. In the association approach, the independent variable is assumed to be continuous, in other words, it has many values/levels. The approach is considered to be associational if the independent variable has five or more ordered categories. Except for this difference, the comparative and associational methods are the same.

Source: Morgan and Griego (1998:6)

5.11 RESEARCH HYPOTHESES

Research hypotheses (or questions) are classified by Morgan and Griego (1998:8) into three broad types: difference, associational, and descriptive hypotheses.

With *difference* hypotheses, groups or values of the independent variables are compared to their scores on the dependent variable. This type of hypothesis is typically used with the randomised experimental, quasi-experimental and comparative methods.

With *associational* hypotheses, independent variables are related or associated with dependent variables.

Descriptive hypotheses are not answered with inferential statistics as they merely describe or summarise data.

5.11.1 Basic difference versus associational research hypotheses

Morgan and Griego (1998:8) define hypotheses as *predictive statements about the relationships between variables.* Both difference and associational questions/hypotheses have as a general purpose the exploration of relationships between variables (see Table 5.3). Statisticians believe that all parametric inferential statistics are relational. This is consistent with the idea that the distinctions between the comparative and associational methods are arbitrary, but educationally useful. Difference and associational hypotheses (questions) differ in terms of their specific purpose and the kinds of statistics they use to answer the question.

Examples of *difference* or group comparison hypotheses include the following types of situations:

- The levels or values of the independent variable (for example, gender) are used to divide the participants into groups (male and female), which are then compared in order to note whether they differ in respect of their average scores on the dependent variables (for example, empathy).

- An example of a directional research hypothesis may be that women will score higher than men on empathy scores. The average empathy scores of the women will thus be significantly higher than the average scores for men (Morgan & Griego, 1998:10).

Examples of *associational* or relational hypotheses include the following:

- The scores on the independent variable (for example, self-esteem) are associated with or related to the dependent variable (for example, empathy). According to Morgan and Griego (1998:10), which variable is considered the independent variable

is often arbitrary, but most researchers conceptualise what they consider the predictor (independent) variable to be and what the outcome (dependent) variable is.

- An example of a directional research hypothesis is that there will be a positive association (relation) between self-esteem scores and empathy scores. Therefore those persons who are high on self-esteem will tend to have high empathy, those with low self-esteem will also tend to have low empathy, and those in the middle on the independent variable will tend to be in the middle on the dependent variable.

Table: 5.3: Representation of how purpose, approach and type of research hypothesis correspond to the type of statistics used

GENERAL PURPOSE	Explore Relationships Between Variables		Description (only)
SPECIFIC METHOD	**Randomised experimental., Quasi-experimental, and Comp ative**	**Associational**	**Descriptive**
SPECIFIC PURPOSE	Compare Groups	Find Associations, Relate Variables, Make Predictions	Summarise Data
TYPE OF HYPOTHESIS (QUESTION)	**DIFFERENCE**	**ASSOCIATIONAL**	**DESCRIPTIVE**
GENERAL TYPE OF STATISTICS	Difference Inferential Statistics (e.g. *t* test)	Associational Inferential Statistics (e.g. correlation)	Descriptive Statistics (e.g. histograms, means, percentages)

Source: Morgan and Griego (1998:9)

5.12 TYPES OF RESEARCH QUESTION

Morgan and Griego (1998:10) distinguish between six types of research question, divided into basic (univariate) and complex (multivariate) research questions, which both include

descriptive, difference and *associational* hypotheses. Table 5.4 represents these types of research question, as well as examples of the types of statistics that are used with them.

Morgan and Griego (1998:10) note that some complex descriptive statistics (for example, cross-tabulation tables) could be tested for significance with inferential statistics – if they were so tested they would no longer be considered descriptive. Most qualitative/constructivist researchers ask complex descriptive questions, as they often consider more than one variable/concept at a time but do not use inferential or hypothesis-testing statistics. Furthermore, complex descriptive statistics are used to check reliability and reduce the number of variables (for example, factor analysis).

5.12.1 Difference versus associational inferential statistics

Inferential statistics can be divided into two types, corresponding to difference and associational hypotheses/questions. Difference inferential statistics are used for the experimental, quasi-experimental and comparative approaches, which test for differences between groups (for example, using analysis of variance). Associational inferential statistics test for associations or relationships between variables and use correlation or multiple regression analysis (Morgan & Griego, 1998:11).

Table: 5.4: Types of research question

Type of Research Question (Number of Variables)	Statistics
Basic Descriptive Questions – 1 variable.	Mean, standard deviation, frequency distribution
Complex Descriptive Questions – 2 or more variables, but no use of inferential statistics.	Box plots, cross-tabulation tables, factor analysis, measures of reliability
Basic Difference Questions – 1 independent and 1 dependent variable. Independent variable usually has a few values (ordered or not).	*t* test, one-way ANOVA
Complex Difference Questions – 3 or more variables. Usually 2 or a few independent variables and 1 or more dependent variables considered together.	Factorial ANOVA, MANOVA
Basic Associational Questions – 1 independent variable and 1 dependent variable. Usually at least 5 ordered values for both variables. Often they are continuous.	Correlation
Complex Associational Questions – 2 or more independent variables and 1 or more dependent variables. Usually 5+ ordered values for all variables but some or all can be dichotomous variables.	Multiple regression

Source: Morgan and Griego (1998:11)

5.13 VALIDITY

Validity can be described as the extent to which one is measuring what one is supposed to measure (Christensen, 1994:201). Rulers, thermometers, measures of weight and other instruments all have demonstrated validity. Validity tends to become more of a problem when applied to psychosocial aspects, where the instruments used may have to have their validity established. Eachus (1999:1) is of the opinion, however, that problems of validity are not restricted to the social and behavioural sciences, but are also prevalent in other sciences.

Various types of validity are of interest to researchers. These are set out below.

5.13.1 Construct validity

The construct can be described as the phenomenon being studied or measured. What matters is whether the construct, as described, is a valid conceptualisation of the phenomenon (Eachus, 1999:1).

5.13.2 Face validity

Face validity refers to the *perception* of the phenomenon. It refers to the requirement that the instrument actually measures this phenomenon and that it must be able to measure it to such an extent that a deduction can be arrived at (Rosenthal & Rosnow, 1991:124). Face validity is therefore concerned with the extent to which the contents of a test or procedure look as though they are measuring what they are supposed to measure.

5.13.3 Content validity

Content validity is the extent to which the content of the test or procedure adequately represents all that is required for validity (Eachus, 1999:1). Content validity means that the questionnaire items represent the kind of material that they are supposed to present. This is usually a basic consideration in the construction phase of any questionnaire.

5.13.4 Criterion validity

This is a measure of validity that is established by the use of a criterion measure; in other words, a test's validity is demonstrated against a known criterion (Eachus, 1999:2). Another form of criterion validity is identified as concurrent validity.

5.13.5 Discriminant validity

Discriminant validity is similar to criterion validity in that it demonstrates the ability of a scale or test to discriminate between different groups.

5.13.6 Predictive validity

Predictive validity is used to make a prediction of future behaviour or occurrences in terms of a determined criterion, based on the grounds of psychological test results (Rosenthal & Rosnow, 1991:124).

5.13.7 Internal and external validity

Internal validity is concerned with ruling out plausible rival hypotheses that may jeopardise statements about whether *x* causes *y.* External validity refers to the generalisability of a causal relationship to circumstances beyond those studied or observed (Rosenthal & Rosnow, 1991:124).

5.14 RELIABILITY

Reliability is concerned with the extent to which a test instrument – whether it is concerned with measuring physical, biological or psychosocial phenomena – is able to produce the same data when the phenomenon is or the phenomena are measured at different times (Eachus, 1999:2).

Reliability may be characterised as either internal or external. External reliability is the easiest to comprehend, as it simply implies the extent to which data measured at one time are consistent with data from the same variable measured at another time.

Internal reliability is more correctly a measure of internal consistency. When analysing reliability in terms of internal consistency, there are several ways of examining the data. To test the reliability of standardised tests, item analysis in the form of Cronbach's alpha coefficient is often used. The alpha coefficient is computed by correlating all the scores on individual items with the overall score on the test. Tests with reliability (those with high internal consistency) will achieve an alpha coefficient of 0.75 or greater on a scale of 0 to 1 (Eachus, 1999:2).

Rosenthal and Rosnow (1991:125) believe that methods of testing reliability can take two forms:

- the measurement instrument is completed by respondents at a specific time, and then the constancy of the responses is measured; or
- the measurement instruments are completed by different respondents at different times, and then the respondents' answers are measured over a determined time frame.

5.15 SENSITIVITY

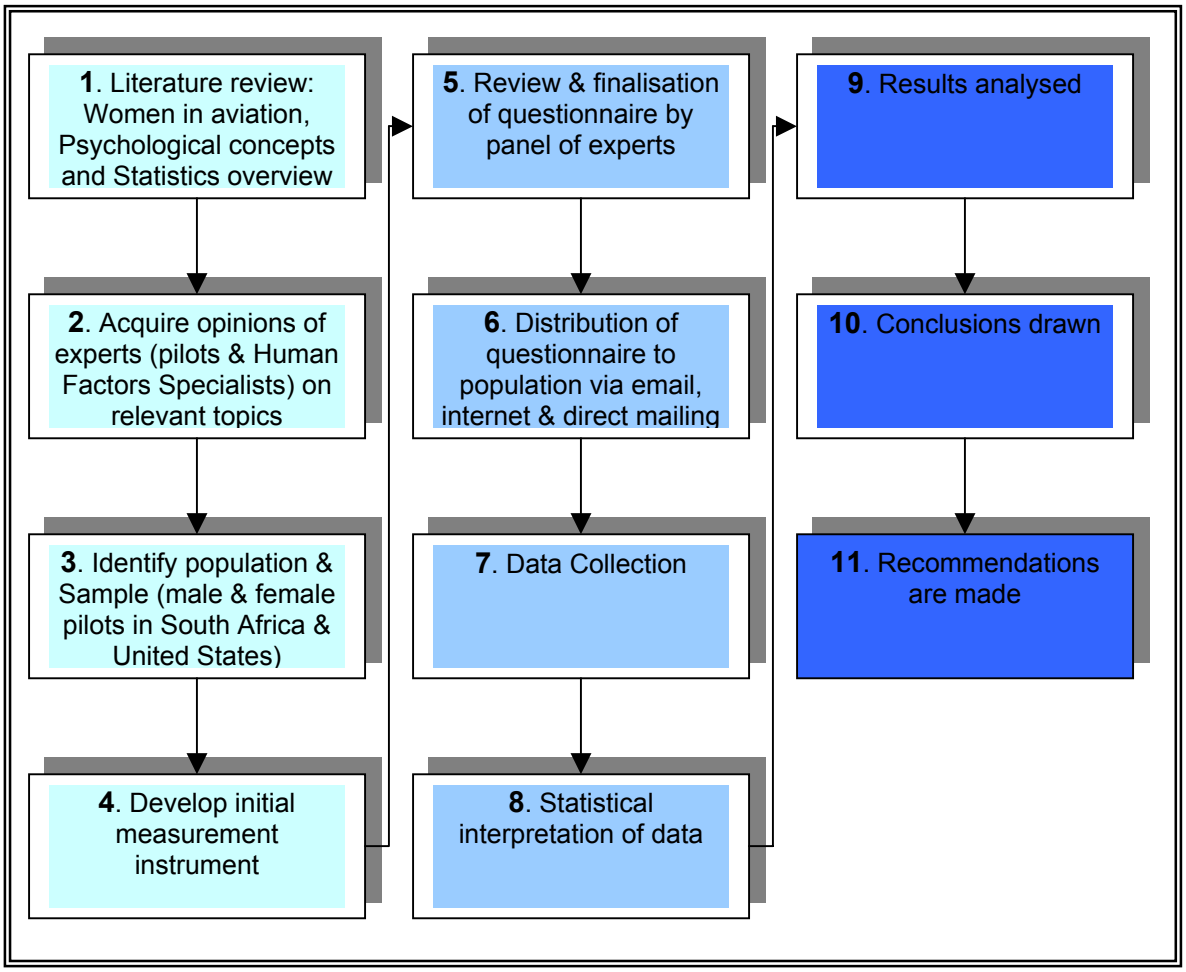
The sensitivity of a measurement instrument refers to the ability of the instrument to discriminate (Eachus, 1999:3). For example, a ruler marked in millimetres has the ability to discriminate (in times of size) to a greater degree than a ruler marked in inches. However, it does not necessarily follow that (for example) a satisfaction scale that measures satisfaction on a twenty-point scale is more sensitive than one with a ten-point scale. This would only be the case if the validity and reliability of the twenty-point scale had been assessed as satisfactory.

5.16 INTEGRATED CONCLUSION

It was noted in previous chapters that although a fair amount of research has been conducted in the fields of CRM and Human Factors in Aviation, very few studies have focused on gender differences/similarities and how these affect cockpit and aviation management. Indeed, a stigma is still widely attached to the role of women as pilots and in combat aviation roles.

The scope of research of this study is summarised in Figure 5.3.

Figure 5.3: Scope of research – summarised



In accordance with the literature cited in Section 5.2.1, the author of this research identified and defined the information this study intends to obtain. This was detailed in Section 1.3 as the research goals. From this information, it became clear that the research process for this study consisted of three distinct phases.

Firstly, a comprehensive literature review was conducted in order to gain a sound understanding of the factors that have influenced women in the field of aviation. The literature study also formed the basis of this investigation and generated important theoretical constructs.

Secondly, an instrument was developed in order to measure the perceptions of male and female pilots regarding gender-based issues in aviation. The questionnaire was evaluated and refined by a panel of experts before being disseminated amongst pilots.

Thirdly, the information obtained from the completed responses were analysed and interpreted, and the findings are discussed in the following chapters.