

CHAPTER 3 THE RATIONALE UNDERLYING THE USE OF A QUESTIONNAIRE AS AN INSTRUMENT FOR THE MEASUREMENT OF ENVIRONMENTAL CONCERN

3.1 The measurement of environmental concern

Various researchers have developed scales with which to measure level of environmental concern and associated issues (Dunlap & Van Liere, 1978; Weigel & Weigel, 1978; Roper Organization, 1990; Stern, Dietz & Kalof, 1993; Steel, List & Schindler, 1994).

In order to assess public opinion regarding changes in ways of thinking proposed by the New Environmental Paradigm (NEP), Dunlap and Van Liere (1978) developed the NEP scale, a twelve statement, four-point Likert-type (1932) scale dealing with issues which distinguish the NEP from the Dominant Social Paradigm (DSP) such as "limits to growth" and "harmony with nature". Using two groups of Washington State participants, one consisting of 806 members of the general population, and the other consisting of 407 members of environmental organizations, these researchers found that, while members of the general population supported the NEP, members of environmental organizations showed an even stronger degree of acceptance - as could be expected. Dunlap and Van Liere (1978) concluded that, while the results at the time needed to be treated with caution, the proponents of the NEP seemed to have been successful in conveying their message.

The Environmental Concern Scale (ECS) was developed in the United States by Weigel and Weigel (1978). This measure consists of sixteen statements, seven positively stated and nine negatively stated, which deal with pollution and

conservation issues, each statement being scored on a five-point Likert-type (1932) scale. The ECS was shown to have sufficient internal consistency, stability and construct validity, and adequately predicts environmentally relevant behaviour. Weigel and Weigel (1978) add that, while there was substantial variation in socio-economic status, the participants were not representative of the different ethnic groups in the United States, therefore the reliability and validity of the ECS in relation to the American black (Afro-American), Mexican and Puerto Rican ethnic groups has not been verified.

Several researchers have developed scales, the values of which are indicative of the level of environmental concern by virtue of the latter's association with the actual construct being measured. For example, the Roper Organization (1990) developed a scale to measure people's attitudes regarding the role played by "people versus nature" in affecting environmental issues; Stern, Dietz and Kalof (1993) measured people's motivation to act in an environmentally responsible manner by assessing their "awareness of consequences" of actions which are potentially threatening to the environment; and Steel, List and Schindler (1994) developed the "forest values" scale which measures people's attitude towards the value of natural forest resources.

The statistical characteristics of environment concern scales which have been found in the literature, are described in Section 6.6.

When designing a research instrument such as mentioned above, it is important that (i) the researcher has defined the construct being researched; (ii) except in the case of an entirely new construct, there should be a high degree of commonality between the researcher's conceptualization of the construct under investigation and the conceptualization of the same construct held by other researchers; and (iii) the research instrument must focus on the object of this shared conceptualization, and not some other construct.

Environmental concern as an attitude has been described in Section 2.4. As one of the aims of this research is to measure environmental concern, it is appropriate to

consider how attitudes are measured and also what are the desirable characteristics as well as the problems associated with a measuring instrument.

Eagly and Chaiken (1993) describe two general methods of attitude scale construction. The **stimulus, then person scaling** approach involves the presentation of a stimulus, followed by the location of the participant on the attitude scale based on their response to the stimulus. With the **person scaling** approach, stimuli are evaluated beforehand as being favourable or unfavourable toward the attitude object, and participants are scored based on their degree of agreement or disagreement with the stimuli.

Important concepts which are relevant to the effectiveness of a measuring instrument, in general, are **reliability**, **validity** and the source of **variance**. The reliability, or internal consistency, of a measuring instrument refers to the degree to which the instrument yields consistently repeatable scores, irrespective of whether it is measuring what it is supposed to measure or not. When specifically testing for similar results, for example, in a test-retest evaluation with an interval of time between the tests, this perspective on reliability is commonly referred to as **stability**. Crano and Brewer (1973) state that it is possible for a measuring instrument to be both reliable and stable, and yet not measure what it is intended to measure. It is here that the validity of the measuring instrument provides the essential link between the construct being researched and its theoretical components.

In general, the validity of a measuring instrument refers to the degree to which the instrument measures what it is supposed to measure, for example, the validity of an instrument used to measure the level of environmental concern relates to the degree to which scores on the instrument scale actually measure the level of environmental concern, as defined by the researcher. More specifically, the type of validity referred to here is known as **content validity**. Should the measuring instrument yield significantly higher scores for participants who are known to have higher levels of environmental concern than members of the general public, for example, active members of an organization involved in environmental conservation, then the

instrument can be said to have demonstrated good **construct validity** by the use of the **known-groups** method (Crano & Brewer, 1973).

In order to describe the concept of **variance**, it is convenient first to describe what is meant by the concept of the **mean**. The mean of a set of scores is the average value of the scores and provides a measure of central tendency. The variance of the scores refers to the degree of dispersion of the scores around the mean. In general, variance may be classed either as **systematic variance** or **error (unsystematic) variance**. Known or unknown factors which influence scores in a predictable way are the source of systematic variance, while the influence of random factors and events which occur "by chance", namely, factors which are unpredictable, are the source of error variance (Kerlinger, 1992). Total variance is made up of both systematic and error variance. The goal of the researcher is to maximize reliability and validity, and to minimize unsystematic variance. Often, however, the minimization of variance involves a "trade-off" type situation, as some techniques employed to minimize unsystematic variance in some areas may introduce new sources of unsystematic variance in other areas.

3.2 Problems associated with the use of a questionnaire

The following discussion deals with specific practical problems which may reduce reliability and validity by introducing or increasing variance in the scores on a questionnaire. Attention is given to problems which apply to the measurement of attitudes using this type of questionnaire and to techniques which may be employed to minimize them.

However, as mentioned above, minimizing variance often involves a "trade-off" or compromise regarding the nature and source of the variance. To arrive at an optimal situation, the benefit of reducing unwanted variance must be weighed against the negative effect of the possible introduction of variance from other sources. This decision is influenced by the nature and purpose of the measuring instrument.

3.2.1 The effect of scale length on reliability and validity

As described in Section 2.6, two dimensions of environmental concern, namely the *anthropocentric* and *ecocentric* dimensions, are proposed. The measuring instrument may therefore be seen to consist of two subscales. It is important that the items (statements) of a subscale measure the same construct, otherwise the reliability of the subscale will be low. Reliability may be increased by increasing the number of statements in the subscale which measure the same construct. However, what is important is for each subscale to have sufficient items so that acceptable levels of reliability and validity are achieved. The implication of having too many items is discussed in Section 3.2.4.

As mentioned in Section 2.5.3, there has been much debate concerning which factors provide an accurate measure of concern for the natural environment. While using only one factor per subscale to determine the level of environmental concern would increase the reliability of each subscale, it would also result in a reduction in external validity, that is, the scale would be less generalizable to other situations.

It is proposed that the statements used in the two subscales focus on two sub-classes of factors (called categories), namely the "air, land and water" (ALW) and "non-human life" (NHL) categories. It was reasoned that the use of the above two general categories within each dimension, or subscale, would cover a workable range of environmental issues, and at the same time would be specific enough to provide both a useful indication of the level of environmental concern of the participants as well as a platform for future research investigating the predictors of environmentally responsible behaviour. A suitable number of statements would be included in each category within subscale to enable amendments, for example, modifications to statements to reduce ambiguity (see Section 3.2.2); addition of new statements after the initial assessment for readability and the preliminary study; as well as deletion of statements after the evaluation by experts.

By specifying the "air, land and water" and "non-human life" categories, and by confining the statements of each subscale to these categories, it was reasoned that

a sufficient number of statements would be retained to enable a balance to be achieved between the length of the questionnaire and acceptable levels of reliability and validity.

3.2.2 Ambiguity

When using a questionnaire, ambiguity may take two forms. The first relates to the statement content, in the sense that different participants may interpret one or more statements differently. In this case the construct is not well-defined and undesirable variance is introduced.

By using statements which refer to specific incidents or characteristics of the construct under investigation (Cattell, 1974), or by purposefully excluding statements which are known to be ambiguous (Edwards, 1957), ambiguity relating to statement content may be reduced.

However, specific situations which are referred to may sometimes be closely allied to similar, perhaps transient, situations relevant in the participants' lives, or may (unbeknown to the researcher) specifically include, or exclude, a person from the set of participants likely to give a valid response. In general, the benefits of reducing irrelevant variance by increasing specificity may be offset by introducing new variance. The more a specifically described situation or characteristic affects, or relates to, an individual, the greater the likelihood is that the person's response will be based on personal, and perhaps short-term, situational factors, rather than on a personal opinion or belief.

The second form of ambiguity relates to participants' responses, in the sense that similar responses may not always have the same absolute meaning. For example, in terms of levels of agreement, the response "strongly agree" may have a different meaning to different participants. Also, a neutral response of "neither agree nor disagree" is different to, and may sometimes be confused with, a neutral response of "don't know, unsure".

The contribution of ambiguity to unwanted variance is most problematic when factual information or objective truths are being measured, but less of a problem when subjective interpretations of events, opinions and beliefs are being investigated (Ochse, 1983).

3.2.3 Faking

In relation to scientific research, faking refers to the distortion of responses obtained from participants by a researcher. It may occur with any method whereby information is acquired on a voluntary basis. It may result from reasons such as an unconscious effort on the part of a respondent to please the researcher, a desire to provide the most socially acceptable response, or a deliberate attempt to provide distorted information.

Fiske (1971) claims that deliberate faking tends to occur most often when the respondent (i) is aware of what is being measured, and why; (ii) is required to disclose his or her identity; and (iii) wants to present him- or herself in a particular way. Faking is likely to be most problematic when participants perceive that they may be affected in some way by the decisions or actions which may result from their responses. Anastasi (1988) expresses the opinion that most questionnaire statements have an answer which is sometimes recognizable as the socially desirable one. In this case, participants may "fake good" in order to create a favourable impression of themselves, especially when applying for admission to an institution of learning or for employment. Conversely, under different circumstances, for example, when an accused person is required, by a court of law, to undergo psychological assessment, he or she may be motivated to "fake bad" in order to create the impression that some type of psychological dysfunction exists.

For this study, it was decided to minimize the possibility of deliberate faking by requesting that participants remain anonymous, and also by not disclosing the precise purpose of the research. The subject of the research, and the content of the statements on the questionnaire, made it difficult not to disclose something about the nature of the research. (One cannot not ask environment-related questions when

researching environmental issues). However, neither the questionnaire itself, nor the instructions to the participants, contained any reference to "environmental concern" as such. For example, when providing information to the participants, the terminology "how people relate to the environment" was used instead of "how concerned people are about the environment". Distortion attributable to social desirability of responses, as a particular response set, is discussed further in Section 3.2.3.1.

3.2.3.1 *Response bias, response set and response style*

Response bias, response set and response style are terms which also do not have a common definition. All are associated with Cronbach's (1946, p476) original definition of response set, which was defined as "*any tendency causing a person to give different responses to items than he would when the same content is presented in a different form*". While most investigations into the effects of these constructs have been associated with personality research, their potential to be a source of unwanted variance in the scores on a questionnaire used when investigating other, non-personality type issues, cannot be ignored.

Various different distinctions have been used for differentiating between these concepts. Rorer (1965) suggests that "set" is associated with a desire on the part of a respondent to produce a certain image, and is therefore dependent on statement content, whereas "style" refers to the tendency to favour a certain response class independent of the item content. Rorer (1965) also does not distinguish between response bias and response set. Rundquist (1966) proposes that neither response set nor response style are independent of item content.

Anastasi (1988, p554) refers to the method of distinguishing response styles from response sets based on the former's association with "*broad and durable personality characteristics that were worth measuring in their own right*", but adds that this distinction is not universally accepted.

For the purpose of this research, response set refers to the tendency to select a particular response class based on statement content, with the purpose of promoting a particular image, response style refers to the tendency to respond in a certain way independent of statement content, and response bias is a general term referring to both response set and response style.

Despite differences in the definition of, as well as in opinions regarding the pervasiveness of, response set and response style (Rorer, 1965; Rundquist, 1966; Campbell, Siegman & Rees, 1967; Anastasi, 1988), it was considered prudent to evaluate, and attempt to minimize, where appropriate, the possible effects of response bias, in the process of developing a new questionnaire.

The **response set** which has received most attention with respect to its influence on the scores of a measuring instrument is **social desirability**. This refers to the tendency of a participant to choose responses which are considered likely to create the most favourable impression, that is, to *respond in a "socially desirable" way* (Crano & Brewer, 1973, p256). It appears to be more prevalent in the area of personality research, and has therefore received most attention from this perspective (Edwards, 1957; Anastasi, 1988). However, Fiske and Pearson (1970) express the opinion that, while social desirability response set exists, its effect is not as pervasive as originally thought.

Social desirability response set may be controlled in various ways, but each may introduce further difficulties. Subtle terms may be used to disguise statement content in an attempt to hide the true meaning from the participants. This approach may, however, result in loss of content validity due to the construct under investigation becoming less well-defined.

The technique of "forced-choice" requires participants to choose one of two items which are matched on social desirability. Unwanted variance may be introduced due to less-than-adequate item matching, and proper selection may become difficult for those participants who may feel that neither option is suitable (Cronbach, 1970).

A further technique to reduce the effects of social desirability response set is to use a social desirability subscale while developing a measuring instrument and to discard statements which load on social desirability, as well as the social desirability subscale itself, from the final form of the instrument (Neill & Jackson, 1970). A variation of this is to incorporate a social desirability subscale into the final form of a measuring instrument and to discard all completed questionnaires which indicate a level of social desirability above a pre-determined maximum (McClain, 1975).

A final method is also to use a social desirability subscale in the final form of the measuring instrument, and to partial out the effects of social desirability (Orlofsky, 1978). With this technique, statements which load on social desirability are not discarded, but variance due to social desirability is removed from interactions between components of the construct being researched.

However, values play an important part in the model of environmental concern used for this research, as described in Section 2.6. Due to the participants belonging to different cultures, and to the possibility that values differ fundamentally between these cultures (Edwards, 1957), it was reasoned that what might be socially desirable within one cultural group, may not necessarily be so in others, that is, different cultural groups may have different ideas about what constitutes socially desirable attitudes and behaviour with respect to the natural environment.

Further, it appears that there is no clear indicator of socially desirable behaviour or attitudes which is consistent across the different cultural groups in South Africa with respect to attitudes or behaviour which relate to the natural environment. On this basis, it was reasoned that any attempt to handle socially desirable responses may not be valid for all cultural groups, and therefore the decision was made not to attempt to control variance due to social desirability response set.

As was mentioned previously, the tendency to select a particular response class based on statement content is known as **response style**. Response styles take

various forms. Some of the more common ones which have received attention due to their influence mainly in the area of personality research are presented.

Acquiescence represents the tendency to agree with the statements presented on a measuring instrument (Crano & Brewer, 1973; Walsh & Betz, 1990; Cohen, Swerdlik & Phillips, 1996). Its inverse, **nonacquiescence** or **negativeness**, is the tendency to disagree. These may be counteracted by using four equally-sized classes of statements, defined as follows:

positively stated and positively scored;
positively stated and negatively scored;
negatively stated and positively scored;
and
negatively stated and negatively scored.

This technique may help to reduce, but not entirely eliminate, unwanted variance due to acquiescence or negativeness. However, participants may have difficulty with negatively-stated items, and the double negative, for example in Afrikaans, may confuse the understanding of the statements even further. On this basis, it was decided to relax the requirement of having an equal number of positively- and negatively-stated items on the questionnaire. While there were both positively- and negatively-stated items on the final form of the questionnaire, their inclusion was primarily based on other criteria, such as selection by expert evaluators and ease of understanding, rather than on rigid adherence to the requirement of having equal numbers of positively- and negatively-stated items. It was planned, however, to have approximately equal numbers of positively- and negatively-scored items on the questionnaire.

Evasiveness is the tendency to select the most neutral response, for example "neither agree nor disagree", or "uncertain". It may be controlled by not making this response category available, and therefore, usually, by requiring the participant to select from an even number of choices for each statement. For this research, it was

decided to control evasiveness by using four options for each statement and not to make a neutral response category available.

Extremity or **extremeness** is the tendency to select the most extreme responses available (Cohen, Swerdlik & Phillips, 1996). It may be controlled by reducing the number of options for each statement to two. However, it was reasoned that the benefits of doing this would be outweighed by the negative effects of the possible frustration of the participants due to their being forced to choose from a limited selection of options. This point is discussed further in Section 3.2.4.

Omission is the tendency to avoid responding, while **cautiousness** is the terminology used when specifically referring to the tendency to avoid responding when in doubt (Cohen, Swerdlik & Phillips, 1996). These may be avoided by discarding all incomplete questionnaires, as was done during this research project. (Note that, as cautiousness is a subset of omission, by discarding all incomplete questionnaires, the researcher can effectively deal with omission and cautiousness without needing to distinguish between them).

Deviation or **deviance** is the tendency to select unusual or uncommon responses (Anastasi, 1988; Cohen, Swerdlik & Phillips, 1996). It may be controlled by adjusting the statement content in such a manner that none of the options can be considered "unusual". This assumes that the researcher has some idea of what "unusual" might be in relation to the construct under investigation. Also, depending on the nature of the construct under investigation, reducing variance due to deviation may result in a reduction in content validity. Apart from not using statements whose content may, based on an intuitive-type judgement, result in an "unusual" response, no attempt was made to control deviation in this research project.

A final response style to be mentioned is **gambling**, the tendency to guess when in doubt (Cohen, Swerdlik & Phillips, 1996). The "test-wise" participant (Miller-Jones, 1989) may guess to his or her advantage, whereas the non-testwise participant may be reluctant to do so, and may potentially end up with a lower score. This response

style is likely to be associated with personality traits (Cohen et al., 1996), and no attempt was made to control it in this research project.

3.2.4 Insufficient motivation to respond with attention

Attention to understanding the statements in a questionnaire, as well as to the selection of an appropriate response, can facilitate the minimization of unwanted variance in the scores on a questionnaire. Certain conditions may, however, discourage participants from devoting the necessary attention to their responses. It was previously stated that too short a subscale is likely to result in insufficient reliability. Conversely, a subscale, or questionnaire, which is too long, is likely to bore or tire the participants, or make them lose concentration, or lose interest in completing the questionnaire (Cronbach, 1970; Schriesheim & Denzi, 1980). This may result in reduced reliability and validity of the measuring instrument. A questionnaire therefore needs to be long enough to be sufficiently reliable, but not too long to result in a reduction in reliability.

Participants may also become frustrated if too few or too many alternative choices are available for each statement. The technique of reducing the number of options in order to overcome the problem of extreme responses may frustrate participants who feel limited by too few choices. While reliability increases with more options, too many options, on the other hand, may result in increased levels of unwanted variance due to different interpretations of the options by different participants, and also to the difficulty experienced by participants when they are required to make fine distinctions between the options.

Of specific relevance to South Africa is the language in which the questionnaire is presented. As was mentioned in Section 2.5.1, South Africa has eleven official languages. Added to this, a percentage of the population consists of immigrants whose home language may not belong to the above-mentioned group of eleven.

Participants who respond to a questionnaire presented in a language with which they are unfamiliar are likely to experience some degree of discomfort. The

possibility of introducing unwanted variance due to the misinterpretation of questionnaire content by these participants must therefore be considered. It was decided that the preliminary study should be conducted using only English language questionnaires. For the main study, questionnaires in five of South Africa's eleven official languages were used. It was hoped that the question identifying the "home language" of each respondent in Part A of the questionnaire would provide not only information regarding the relationship between home language and environmental concern, but also information regarding the percentage of participants whose home language was one of the five available for completing the questionnaire.

Furthermore, ideas for some of the statements on each subscale were obtained from existing associated English-language measuring instruments. Care was taken to avoid the use of idiomatic phrases which may be unique to their country of origin, or unique to a particular culture, and therefore inappropriate for general use in the South African context.

3.3 Problems relating specifically to the measurement of environmental concern

Research into the determinants and correlates of environmental concern has not always produced consistent results. One of the most common problems which has confronted researchers investigating environmental concern involves the question of how they and their participants understand the construct. These conflicting research results, together with an inordinately large number of influencing variables, have prompted Van Liere and Dunlap (1981) to express the opinion that researchers have not always conceptualized environmental concern in the same manner, and therefore also to question whether researchers are measuring the same, or different, underlying constructs, or whether they are perhaps measuring different components of an attitude.

Van Liere and Dunlap's (1981) analysis of the components of environmental concern provides some useful insights. These researchers distinguish between "substantive" and "theoretical" components, or measures, of environmental concern.

Substantive measures refer to environmental issues such as pollution, use of natural resources and conservation of wildlife. These issues have sometimes been researched as separate dimensions of environmental concern, and sometimes they have been combined into a single uni-dimensional measure of environmental concern. Theoretical components refer mainly to researchers' assumptions about how people (their participants) conceptualize environmental concern, and what strategies should be used to address environmental problems, for example, perceived seriousness of environmental problems, or support for stricter pollution control.

Van Liere and Dunlap (1981) propose that the degree to which each individual substantive or theoretical component has been included in a measure or not, contributes to variation in the measurement of environmental concern. Although one may speculate about the existence of a unique set of substantive and theoretical components which would provide an exact measure of environmental concern, Van Liere and Dunlap (1981) succeeded in emphasizing that varying conceptualizations of environmental concern make a difference with respect to how the construct is measured.

As was mentioned in Section 2.4.2, and discussed further above, Van Liere and Dunlap (1981) suggest the substantive-theoretical distinction as a basis for understanding how researchers have conceptualized environmental concern. However, the determination of this distinction is not a pre-requisite for the usefulness of a measuring instrument to be determined. The substantive versus theoretical properties of the instrument may be discussed, if appropriate, after the usefulness of the instrument has been demonstrated.

Also, the various environmental concern scales which have been developed may measure specific components of an attitude, such as beliefs about outcomes, others may measure both beliefs and evaluative components, while yet others may measure multiple environmental issues, suggesting a general environmental concern attitude scale (Tarrant & Cordell, 1997). In stressing the practical use of a measuring instrument, these researchers add that, if such an instrument is a good

predictor of environmentally responsible behaviour, then its usefulness is valid within an attitude-behaviour conceptualization, regardless of its theoretical orientation.

One may speculate about whether a unique set (or subset) of factors or dimensions is necessary to quantify environmental concern fully. The approach used in this research with regard to the dimensionality of environmental concern is described in Section 2.6.

3.4 The use and justification of a Likert-type scale

Several scaling techniques are available by which attitudes may be measured (Thurstone, 1931; Likert, 1932; Guttman, 1944). The Thurstone (1931) and Guttman (1944) scales appear to be more difficult to develop than Likert (1932) scales (Taylor, 1983), and tend to become unwieldy to administer when the number of participants becomes relatively large.

In contrast, Likert's (1932) **method of summated ratings** is a person scaling approach which is easy to develop and administer (Van Staden, 1983), and is therefore a widely-used instrument for the measurement of attitudes. This method requires the participants to rate a number of statements relating to the attitude object, with each statement receiving a score from, for example, 1 to 5, with 1 indicating strong disagreement and 5 indicating strong agreement. Scoring is reversed for negative (unfavourable) statements, so that a high score, for example 5, indicates strong disagreement. The sum of the scores for all the statements on the scale is an indicator of attitude strength. In the example used above, a neutrally scored statement is indicated by a response of 3. Should a researcher not want to make a neutral response available, then an even number of options, for example, from 1 to 4, is made available for the evaluation of each statement by the participants.

The Likert (1932) technique can easily combine items dealing with different issues into a single score, that is, it can easily accommodate statements compiled according to the "air, land and water" (ALW) and "non-human life" (NHL) categories

within each of the proposed anthropocentric (ANT) and ecocentric (ECO) dimensions of environmental concern. Furthermore, as it may be used with multiple classes of items, it is compatible with the cognitive, effective and behavioural categories of motivated response components of the attitudinal model described in Section 2.6.

Therefore it is proposed that a Likert-type (1932) self-report questionnaire is a suitable instrument to use for the measurement of environmental concern. When designing statements for the questionnaire, cognisance was taken of the contents of this chapter, as well as the "informal criteria for attitudinal statement construction" as stated in Appendix 5. The statement selection procedure is detailed in Section 5.5. It was planned to use this questionnaire in conjunction with a biographic/demographic-type questionnaire in order to investigate the relationship between environmental concern and its biographic/demographic influencing factors.