CHAPTER 7
OPERATIONALISATION OF VARIABLES (MEASUREMENT PROCEDURES)

The theoretical, methodological and empirical background to the present study, the research problem, and formulated objectives, all presented in the foregoing chapters, culminate at this point in operational decisions about and descriptions of how the constructs have been measured. The principles underpinning the arguments and choices about the design or selection of measurement instruments, were:

- Respondents' performance in terms of the main variables had to be assessed by applying more than one method or instrument in each case. Such a ~~multi-method~~ evaluation enables the cross-validation of measuring instruments.
- A ~~trade-off~~ had to be found between the ~~effort~~ (time, ease and cost) of applying the instruments, and the ~~quality~~ of data generated. The latter aspect has to do with both the completion of the measurement instruments in terms of providing honest and accurate answers, and the return rate of completed responses.
- The findings (analysis of data) had to be ~~comparable~~ and ensure optimal ~~continuity~~ with previous research findings, constructs, conceptualisations, operational aspects and statistical procedures.

The foregoing principles acted as parameters within which the operational elements had to stay. Consequently, the measurement instruments or fieldwork had to adhere to the following practical requirements:

- sex-role identity and relationship satisfaction each had to be evaluated by means of two instruments;
- measurement instruments had to be self-applied, and be suitable for distribution to individuals or small groups through informed and competent go-betweens;
- the instruments had to be in printed form, to be completed, as far as possible, only by selecting pre-coded responses; and
- frequently used and proven techniques had to be adhered to, provided that existing instruments used or adapted had to be substantiated by well-documented, acceptable psychometric properties.

More information about the instruments is appended at the end (Annexure 7.1 and onwards), and is cross-referenced within this chapter to avoid interrupting the flow of argument.

The structure and content of the present chapter are as follows:

- Section 7.1, comprising the general arguments and principles underlying the conceptualisation of the main (dependent and independent) and subsidiary variables; and
- Section 7.2, in which the detail about how the empirical measuring instruments have been selected or derived, is discussed.

7.1 General conceptualisation

Limited by practical considerations, the main (empirical) purpose of the present study is to establish the connection between sex-role identity and relationship satisfaction, and how these two variables, and the relationship between them, are moderated by other contributing factors. Relationship satisfaction is treated as the dependent variable, and sex-role identity as the independent variable. This arrangement suggests a loose causal relationship. However, the study is cognisant of the finer nuances applying to causality in this case. Of special concern is the direction of causal links, and the third-variable problem, i.e., the possibility that strong correlations between two variables may actually be dependent on more primary correlations between both these two variables and an unknown third variable. These issues have been taken into consideration during the analysis of the data (Chapter 9).

A more detailed exposition of the operational elements and the selection of the measuring instruments follow under 7.2, after the discussion of the more qualitative arguments in support of the conceptualisation and selection of dependent, independent and contributing variables.

7.1.1 Dependent variable: relationship satisfaction

The main objective in the present study is to learn more about the hypothesised impact of relevant personality determinants, such as sex-role identity, on relationship satisfaction. This broad, hypothesised link led to treating relationship satisfaction as the dependent variable in the theoretical model and corresponding analyses.

Much debate has occurred about how to operationalise this dependant variable in the context of conventional marriage. However, the initial measuring tools can easily be converted to assess close dyadic relationships irrespective of the gender orientation of the partners. In this way, one of the main research problems, outlined under 5.1.3, has also been addressed, namely that the instruments selected, should use language that is able to accommodate different kinds of relationship.

The possibility to widen the focus beyond marriage, is evident under 2.1.6, where the different approaches to operationalisation are introduced. Burr (1967) ordered the existing measurement practices into two approaches, which he called the generalist approach and the specific approach.

The former accepted that relationship (marital) satisfaction could be measured at an overarching or aggregate level. An identifiable, definable and measurable construct such as relationship satisfaction would thus exist. It could give a unique, overall impression of the success of intimate cohabitation within a single concept with symbolic meaning. Also, it did not rule out the possibility of using composite-criterion measures, such as the *DAS* with its total score and sub-scales.

In the second approach, it was assumed that only the measurement of specific aspects of relationship functioning was meaningful, and even possible. These aspects typically included satisfaction and success in aspects such as the joint management of financial affairs, and sexual fulfilment.

In deriving a composite-criterion scale, seemingly in pursuance of the first tradition, Spanier (1976) succeeded in addressing many of the difficulties by the way in which he proposed measuring dyadic satisfaction with his *DAS*. The first difficult element to blend into the final instrument, was the translation of the concept underlying the relevant research construct into an operationalised instrument. The second element on which a compromise had to be reached, was avoiding an oversimplification of the choice between so-called specific and generalist approaches. Spanier managed the balance in the form of a multi-criterion, dynamic adjustment approach. In this approach, process and development received proper emphasis, as is evident from the following statements by Spanier (1976, pp. 16-17):

> *Marital or dyadic adjustment may be viewed in two distinct ways – as a process, or as a qualitative evaluation of a state. … A "process" definition … is predicated not only on the existence of a continuum, but also on the belief (of)* movement *along the continuum. … The process consists of those events, circumstances and interactions which move a couple back and forth along this continuum. Definitionally, then, we could say that dyadic adjustment is a process of movement along a continuum which can be evaluated in terms of proximity to good or poor adjustment.*

For the purposes of the present study, the concept of relationship satisfaction is defined as the partners' simultaneous experience of their close dyadic relationship (of at least two to three years duration) in an overall sense as satisfactory, happy and successful. Such satisfaction is to be understood with special reference to partners' exercising sound communication skills, managing conflict constructively, experiencing emotional fulfilment and enjoyment, having genuinely positive subjective feelings, and pursuing an optimal number of common interests.

This definition does not exclude the possibility of gaining meaningful information about selected components of relationship satisfaction. The latter statement also includes the possibility of analysing individual satisfaction meaningfully, and relating it to the other research variables. Happiness, success and satisfaction are taken to describe a single semantic field. Understanding relationship satisfaction more as personal satisfaction and happiness, will reflect the intra-psychic, individual attributes (subjective experiences), without preventing aggregate dyadic configurations being studied. Success more manifestly reflects the dyadic patterns of congruence of experience between the partners in the dyad, leading them to continue living within a relationship perceived to be constructive and beneficial. The foregoing definition expands the preliminary one proposed in Section 2.1.6.

In conclusion, relationship satisfaction, as applied to close dyadic relations, serves as the dependent variable of the present study. One of the following configurations could apply to any dyad:

- both partners are satisfied and happy;
- both partners are dissatisfied and unhappy;
- any one partner is satisfied while the other is not.

Most of the "sub-variables" of relationship satisfaction suggested by proponents of the generalist approach (see again Section 2.1.6) are considered being covered well by the following array: dyadic consensus, dyadic satisfaction, dyadic cohesion, and affectional expression. These indeed formed the basis for the clusters of items proposed by Spanier (1976). The clusters are very briefly outlined

immediately below, whilst the instrument itself is only introduced under 7.2.1.1. Whether or not, and how, the dynamics of these four components of relationship satisfaction are determined or predicted by the kind / type of partners concerned, are gradually coming into sharper focus.

7.1.1.1 Dyadic consensus

Dyadic consensus relates to the conflict resolution skills of partners. As such, it addresses a fundamental component of interpersonal communication. Communication plays a highly salient role in identifying conflict, and in working out problem-solving strategies to the benefit of the couple as dyad. This should not result in a win-lose trade-off in favour of either partner, but rather in agreement on the resolution of problems.

7.1.1.2 Dyadic satisfaction

Dyadic satisfaction refers to the partners' overall, subjective experience of happiness about their relationship. It gauges the ability to accept their personal contribution to each other's experience of the relationship as positive. It also relates to accepting responsibility for creating such experiences for oneself and one's partner without uneven trade-offs.

7.1.1.3 Dyadic cohesion

Dyadic cohesion evaluates the direction of partners' energies towards mutually beneficial outcomes. The main focus is how, and how much, partners reach agreement on the definition and pursuit of common interests. This includes assessing how the balance is reached between individual interests and common ones, without compromising either dyadic success or individual fulfilment.

7.1.1.4 Affectional expression

Affectional expression also has an important communicative side to it. However, the crucial ability here is to communicate emotional content successfully. Sexual satisfaction is included under this component, because of the way in which it operates as barometer of the degree to which mutual emotional needs are identified and communicated, without becoming a weapon or seat of pathology in a self-serving way to either partner.

7.1.2 Independent variable: sex-role identity

The main (empirical) objective of the present study is to establish the extent of the hypothesised impact of sex-role identity on relationship satisfaction. This determined the handling of sex-role identity as the independent variable in the theoretical model and corresponding analyses.

In a general sense, sex-role identity is a much less problematic construct than relationship satisfaction. This is already evident from the discussion under 2.1.4. There has also been a much more continuous and coherent development of the construct from the outset. One only needs to consider the route (see under 2.2.2) followed by Bem (1993) in expanding her ~~theory of gender schematicity~~ into the ~~enculturated lens theory~~. The latter relates to how gender-polarised social practices serve to transfer gender lenses from culture to the individual. Throughout the course of these developments, this polarised dichotomy found application by means of measuring the salience with which individuals perceived and adhered to the norms of masculinity and femininity attached to societal stereotypes. This situation also enabled the present researcher (see under 2.3.2) to generate, measure and study the various configurations of sex-role identity (as both traits and type).

For the purposes of operational measurement, sex-role identity is defined as in Section 2.1.4. It is possible to do so because of the almost classical nature that the concept of sex-role identity assumes at present.

Sex-role identity (type) is accordingly defined as the independent presence within any individual of high or low numbers of those psychological characteristics (traits) considered to represent the constructs of masculinity and femininity. The resulting four types of individual in terms of sex-role identity are: masculine, feminine, undifferentiated and androgynous. The nature of the primary psychological constructs or traits is determined as feminine or masculine through an analysis of the descriptive and prescriptive content of sex-role stereotypes.

Prinsloo (1990) amply recorded the latter analysis. Sex-role identity (mainly in the sense of "types") is treated as the main independent variable of this study.

Basow (1992, p. 13), in quoting the findings of a study (involving college students) reported in 1981 by Bem, provided a typical example of the distribution of sex-role identity type in a sample. Although any such distribution depends on the sample, measuring instrument, scoring procedure, and year of study, it would be informative to have a South African account too. According to Bem's findings, the following distribution is typical:

- undifferentiated: 18 % of females, and 27 % of males;
- masculine: 12 % of females, and 42 % of males;
- feminine: 39 % of females, and 12 % of males; and
- androgynous: 30 % of females, and 20 % of males.

The need for, and rationale underlying, evaluating both self-rated and ascribed sex-role identity in the present research, has to be taken up again, because previous research identified important gaps in this regard, and raised complex issues hitherto not sufficiently addressed (see under 2.1.4). The work of Lamke *et al.* (1994), based on the model proposed by Ickes, is particularly relevant here. The subtle difference at stake is that sex-role identity may help determine partners' experience of relationship satisfaction from two quite different angles. The one presupposes a more external, objective source, while the other is actually internal, subjective, and essentially projected. Therefore, an individual's level of relationship satisfaction may be influenced by:

- the ~~self-rated~~ sex-role identity type of the partner, as measured by the partner's personal responses to an appropriate instrument (in which the partner rates his / her own adherence to the relevant stereotypes); or by
- the ~~ascribed~~ or perceived sex-role identity type of the partner, as rated by the first individual (who, by completing the appropriate instrument, ascribes to his or her partner the latter's adherence to the relevant stereotypes).

It is obvious, therefore, that one partner's relationship satisfaction may be connected to the other partner's sex-role identity, either as ascribed by the first partner to the second (with the rating originating from within the first partner, almost as the result of an intra-psychic process), or as rated by the second partner himself/herself (with the rating originating in the second partner, resulting in an interpersonal process). This phenomenon deserves closer scrutiny. It is also important in view of indications that there is often a quite sizable agreement between the ascribed and self-rated sex-role identity scores of individuals.

Siavelis and Lamke (1992) found that the relationship satisfaction scores of individuals (male and female) correlated most closely with the perceived femininity of their partners. What is more, a person's own femininity significantly predicted the femininity which that person ascribed to his/her partner. It was hypothesised that this happened through mediation by the associated skills in self-disclosure. One caveat has to be sounded at this point. Both the studies in which Lamke participated, involved dating couples sampled mainly from undergraduate psychology student classes. This fact may render the generalisation of the findings problematic. The present study heeds the call of Lamke *et al.* (1994) for extending research to marital and same-sex friendships.

7.1.3 Contributing or nuisance variables

The theoretical framework reflected in Figure 3.1 dictates that all the variables other than the two main variables (sex-role identity and relationship satisfaction) have to be treated as contributing variables. In a certain sense, the various hypothesised relations between them and the main variables will be evaluated during statistical analyses that will treat them as additional independent variables in the models tested.

To avoid repetition, only the two broad clusters of contributing variables are defined in this section. They overlap largely with the intra-personal and extra-personal factors or variables introduced from the literature review in Sections 4.2 and 4.3 respectively. Figures 4.2 and 4.3 further clarify their role in terms of the theoretical framework. A general list of the various factors is provided after defining them. The operational aspects, such as the options that respondents had to select for each factor, are explicated further under 7.2.3 and in Annexure 7.5. With that, detailed explanations are given of how information on these elements will be gathered mainly by means of the biographical part of the measurement instruments.

The ~~intra-personal factors~~ are defined as those personality characteristics, values and intentions that serve to motivate individuals' behaviour from sources largely internal to that individual. The

components considered here include cognitive ability (intellectual ability); emotional functioning (stability / lability, and anxiety); attitudes (towards many of the relevant issues such as sex, and relationship type, being modern or traditional); personality traits (introversion, self-esteem, ego-maturity, independence, creativity, and general well-being); communication skills (styles and patterns, and self-disclosure, including of emotions); and perceptions (adhering to stereotypes).

The ~~extra-personal factors~~ are defined as those circumstances and physical or material conditions that serve as broader context determining individuals' behaviour from sources largely external to that individual. The components considered here typically include sex; age (which may represent life-stage development); health status (handicaps and other impairments); socio-economic status (material wealth, and cultural conditions); family stage (presence and number of children, and the duration of relationships); career and job factors (job status, sources of income, job satisfaction, and single- or double-career orientations); educational background (level and direction); and religious orientation.
Some additional ~~interactive factors~~, such as the age difference between partners, and the examples set by parents in terms of relationships (marital status, consistency, level of relationship satisfaction), are included as well. Particular attention is also given to the life-development perspective.

7.2 Measurement instruments

In this section, the tools applied for data gathering are discussed. The motivations behind selecting certain tools, and rejecting others, are both presented, because of the valuable insights gained through doing so.

7.2.1 Relationship satisfaction

It is evident from previous research that almost any assessment cocktail can be put together. For example, Lye and Biblarz (1993) chose to apply five separate measures of marital satisfaction. These were:

- a single item scored as yes/no;
- a single item scored on a 3-point scale;
- a single item scored on a 7-point scale;
- a five-item, 5-point scale assessing one area of marital happiness; and
- another five-item, 5-point scale assessing a different area of marital happiness.

It is evident that many reasons can determine the selection of particular techniques. The given example indicates a seemingly very economic battery of 13 easily applied and scored items. However, there also seems to be an insufficient theoretical base and a lack of common conceptualisation of the research construct.

In the case of the present research, the choice of technique for assessing relationship satisfaction rested on the following motivation:

- a well-accepted, classical instrument should serve as anchor measurement;

- the instrument should cover the central concept sufficiently to confirm its uniquely discernible character, whilst also representing important components or detailed facets;
- the instrument should be suitable, or easily adaptable, for measuring same-sex dyads;
- a second rating should be available to enable cross-validation of scores; and
- the combined battery of assessments should be feasible and economic in terms of yield of information against effort.

It soon became clear from the research literature that the *DAS* of Spanier (1976) was a worthy candidate. Not only has it been used very often, so much so that it has become the acknowledged standard, but its reliability and validity have been researched and documented abundantly. Should this impression hold during scrutiny, only a very concise confirmatory alternative (tool) would be required. In exploration of the techniques on offer (Table 3.1 in Column F lists almost 40), some of the most relevant information on the *DAS* is considered next.

7.2.1.1 Spanier's *Dyadic Adjustment Scale*

Spanier (1976) created a brief instrument of 32 items with quite impressive psychometric properties. It can easily be applied unchanged, or adapted only very slightly for use. The instrument is called the *Dyadic Adjustment Scale (DAS)*. The main benefit of this instrument is the close overlap between the circumstances for which it was designed and the dyadic focus of the current study. It also goes beyond traditional male-female marital relationships. Four different components (sub-scales) of dyadic adjustment can be assessed, namely dyadic satisfaction, dyadic cohesion, dyadic consensus and affectional expression. The performance of the sub-scales was verified through factor analysis. Spanier (1976, p.21) reported the following loadings:

13 items	for dyadic consensus	with loadings from 0,34	to 0,73;
4 items	for affectional expression	with loadings from -0,48	to -0,57;
10 items	for dyadic satisfaction	with loadings from 0,32	to 0,82;

and

5 items	for dyadic cohesion	with loadings from 0,50	to 0,71.

In only one case (Item 4), a competing loading of above 0,30 (0,35 to be exact) on a rival scale was evident. However, this is deemed insignificant in the light of the high loading of -0,57 achieved by the item in terms of its loading on its primary factor.

The conceptual rationale (or criteria) adhered to by the two developers of the *DAS* (Spanier and Cole, in 1974) was that (the definition of) adjustment had to satisfy the following five conditions:

- it should be distinguisable from other concepts;
- it should be operationalisable;
- it should account for all criteria considered to be important in the conceptualisation of adjustment;
- it should not be so abstract that it could not be clearly conceptualised, nor be so specific that it could not be applied to all marriages; and

- it should also apply to non-marital dyads.

(For Spanier's definition of dyadic adjustment, see Section 2.1.6.)

The technical detail about how the *DAS* was developed by taking into account a multitude of procedures can be found in Spanier (1976, pp. 17-20). A summary of these procedures shows that the researchers: identified an item pool (300 in total); had the items judged for content validity; pilot tested the experimental items (about 200 in number) through purposive sampling with married couples, divorcees, and non-marital dyads; analysed the items through frequency distributions and t-tests; and factor analysed the remaining/selected items (40). The result was a scale of 32 items. The test is short and easily self-administered, and can be adapted for use in interviews. Researchers can also use selected sub-scales without fear of compromising reliability or validity.

Sufficient indications were given by Spanier (1976) that conventionality and social desirability effects were minimised. In a way, the scale is sensitive to the issue of measuring individual *vis-à-vis* relationship functioning. Most of the items attempt to assess respondents' perception of the adjustment of the relationship (as a functional group). Reliability, and content, criterion-related and construct validity outcomes were also reported. They all were quite satisfactory. Three judges evaluated the items for content validity. Items were included when judges considered them to be: relevant measures of dyadic adjustment for contemporary relationships; consistent with nominal definitions of adjustment and its components as suggested by the developers; and carefully worded with appropriate fixed-choice responses. ~~Criterion-related validity~~ was established by predicting marital status (married or divorced) from the *DAS*-scores. For all sub-scales and the total scale sample means differed significantly (t-tests, $p < 0,001$). ~~Construct validity~~ was testified to by correlating the *DAS* with other instruments in existence. The correlations between the *DAS* and the Locke-Wallace *Marital Adjustment Scale* of 1959 were 0,86 for married and 0,88 for divorced respondents ($p < 0,001$). In addition, the outcome of the factor analysis also testified to construct validity. ~~Reliability~~ coefficients (Cronbach's alpha coefficient, a variant of the basic Kuder-Richardson formula) for the various (sub-)scales were: dyadic consensus - 0,90; dyadic satisfaction - 0,94; dyadic cohesion - 0,86; affectional expression - 0,73; and the total *DAS* - 0,96.

Many researchers who used the *DAS* afterwards confirmed its solidity as a psychometric technique for assessing satisfaction in dyadic relations. Some references are made to such reports.

Strickland (1981) used the instrument to good effect in his research on the effect of marital enrichment on marital satisfaction. He was impressed with the thorough content validation undertaken by Spanier. First, all items ever used in marital adjustment instruments since 1929 were collected. After eliminating duplicate items, the items were then submitted to judges to consider whether or not they could also assess dyadic adjustment in contemporary relationships, were consistent with Spanier's nominal definitions for adjustment (including its components satisfaction, cohesion and consensus), and were carefully worded and formatted in terms of fixed-choice responses. Strickland preferred the *DAS* above other instruments because the former controlled for acquiescence and direction-of-wording effects.

Antill (1983, p. 147) also used the technique, which he considered "psychometrically sound". For his sample, for both sexes, the alpha coefficient was above 0,75 for each of the sub-scales, except affectional expression. For the complete scale, the coefficients were 0,90 for males and 0,92 for females.

Martin (1985) used seven items from the *DAS* to measure marital adjustment. They were Items 8, 10, 11, 25, 27, 28, and 31. They were found to be as robust as the total scale.

Also Parmelee (1987, p. 434) submitted a good report about the *DAS*'s properties. The alpha coefficients for husbands and wives respectively were: for dyadic consensus - 0,78 and 0,84; for affectional expression - 0,65 and 0,67; for dyadic satisfaction - 0,77 and 0,82; for dyadic cohesion - 0,67 and 0,70; and for the total *DAS* scale - 0,86 and 0,90.

Li and Caldwell (1987) also attributed excellent psychometric properties to the instrument by citing routinely reported reliability coefficients in the 0,90s.

The Chinese version of the *DAS*, or *C-DAS*, developed in 1993 by Shek, Lam, Tsoi and Lam (Shek, 1995), also showed very acceptable psychometric properties. When Shek later, in another study by himself, involved 1 501 married adults aged between 30 and 60 years, he succeeded in replicating the factor structures originally documented by Spanier. The structures proved to be robust (similar and stable) across two random sub-samples and sex, supporting the universality of the concept of dyadic adjustment.

Piotrowski (1999) did a review of instruments used in the past to assess marital and family functioning. They found that the *DAS* was by far the most prominent tool used for assessing marital satisfaction. In addition, its use was on the increase. For the three eight-year periods from 1974 to 1981, 1982 to 1989, and 1990 to 1997, the studies in which the DAS was used, numbered 13, 147 and 159 respectively (the total was 319). Tools like the *Marital Adjustment Test*, *Marital Satisfaction Scale* and *Marital Satisfaction Inventory* did not even reach 20 % of the *DAS*'s turnover. Use of the former of the three was actually declining, with the others slightly in ascendancy, if not at a stable level.

Although busy validating the *Relationship Assessment Scale (RAS)*, Vaughn and Matyastik-Baier (1999) provided good criterion-related validity evidence for the *DAS* (correlation of 0,84 between the two tests).

Recently, Crane, Middleton and Bean (2000) used the *DAS* in their study of 486 married American individuals (through research at the Brigham Young and Auburn Universities), while evaluating various measures in terms of their ability to distinguish between distressed and non-distressed marital couples. They also produced conversion tables between the various instruments for individual and couple scores, and cut-off scores. The instruments at stake are the *Kansas Marital Satisfaction Scale (KMSS)*, the *DAS* and its revised version (*RDAS*), and the *Marital Adjustment Test (MAT)* and its revised

version (*RMAT*). The developers of the *RDAS* (Busby, Crane, Larson and Christiansen in 1995, as cited by Crane *et al.* in 2000) claimed brevity, acceptable construct validity and adequate internal consistency as its main strengths. The findings confirmed the sound psychometric qualities of the *DAS* and *RDAS*, and the equivalence of all the instruments in establishing distress levels. These findings strongly support using the *DAS* in the present study. (Note that the *RDAS* had not been released when the instrument selection was made for the present study, explaining why the original *DAS* was used.)

Highly acceptable psychometric properties were reported for a Turkish version of the *DAS* (Fişiloğlu & Demir, 2000). The same applied to a Chinese-American population (Lim & Ivey, 2000). In the latter case, the researchers were unable to replicate the four-factor pattern, leading them to conclude that only the global score was potentially useful. However, their small sample (n=63) renders the outcome of the factor analysis suspect. Both these findings increase confidence in the instrument as a valid and reliable tool across cultures.

It was decided to use the *DAS* for measuring relationship satisfaction in the present study. The instrument is included as Annexure 7.1. The following items load on the respective scales: dyadic consensus – Items 1-3, 5, 7-15; affectional expression – Items 4, 6, 29, 30; dyadic satisfaction – Items 16-23, 31, 32; and dyadic cohesion – Items 24-28. The only change made for the present study concerned Item 9, where "in-laws" suggest marriage. Reference to "own or partner's parents or family" was considered to address the problem.

7.2.1.2 Olivier's *Sexual Adaptation and Functioning Test*

The *Sexual Adaptation and Functioning Test (SAFT)* of Olivier (1985a; 1985b) was also considered. However, it is a fairly time-consuming procedure in terms of application and scoring. It is a projective technique intended for use in clinical situations to assist therapists in gathering information on clients' sexual needs and preferences. Such information aids therapists in designing sex therapy to address sexual dysfunctions or problems regarding sexual adaptation. The test can be applied with love-partners from both sexes in relationships with the same or the opposite sex. In the context of the present research, the instrument is suitable for collecting information on sexual preferences, inclination towards sexuality, meaning attached to sexuality and aspects thereof, the degree of fulfilment of sexual needs, the presence of sado-masochistic tendencies, and latent homosexual desires. The importance of certain contributing variables for relationship satisfaction could be explored in this manner. The fieldworker has to be trained properly to handle the technique.

A number of limitations motivated the decision not to make use of the technique. The most important reason is that the technique is too involved and time-consuming compared to the limited nature of the aspect of relationship satisfaction on which information would be gathered. What is more, the test has to be administered and scored by a psychologist. This prevents self-administration, and even using

assistants to apply and score it. Another limitation is that the technique is difficult to interpret, also in the sense that it does not render quantitative scores on interval or ratio scales. Thus, statistical comparisons are not possible.

Although it may have been advisable to apply the *SAFT*, or a sample of its items, to only a sub-sample of respondents to overcome some of the obstacles mentioned, the cost-benefit trade-off still seemed too high.

7.2.1.3 Direct questions

Another often-forgotten approach is to include one or more simple, single items directly requesting respondents to assess their relationship satisfaction. This also enables the easy cross-validation of measures. Scoring is normally done on a 3-, 4- or perhaps 5-point scale. The simplicity, economy and directness of the single-item strategy swung the decision in favour of its use.

More than 20 such items are listed in Table 3.1 (Column F). Some of them actually coincide with items already used in a related form in the *DAS*, and could be discarded. The following three items were preferred:

- Would you marry the same partner again if you had a chance? (Yes/No)
- I am happy when I'm with my partner. (Agree/Disagree?)
- During the previous year, did you consider your marriage to be in trouble? (Yes/No)

The set of three items seems to provide a good balance between past experience and future commitment, to cover positive and negative elements in the relationship, and to enable simple or scaled responses. As a result, the wording was slightly adapted to allow for the assessment of same-sex relationships, and to make provision for scaled options covering a wider score range. The final formulation of the three items is found in Annexure 7.2. Each item can be correlated individually with the *DAS* (total and scale scores). However, they can also form a short scale with a maximum score of nine to provide another opportunity for cross-validation.

7.2.1.4 Other relationship satisfaction instruments

Some measurement instruments deserved consideration because of their potential usefulness. After scrutiny with a view to possible use or adaptation, none of them were selected. A few notes are nevertheless warranted.

Venter and Zeelie (1992, p. 314) did substantive work on aspects of marital relationships in the South African context. They used the following instruments to assess marital relationships: *Enriching and Nurturing Relationship Issues, Communication and Happiness (ENRICH)*; *Personal Assessment of Intimacy in Relationships (PAIR)*; and *The Primary Communication Inventory (PCI)*.

❖ *Enriching and Nurturing Relationship Issues, Communication and Happiness (ENRICH)*

Olson, Fournier and Druckman designed this instrument in 1983 (Fowers, 1991). The *ENRICH* Inventory has 125 items in 14 scales (11 scales with 10 items each, and 3 scales containing 5 items each). The scales and what they measure are (those indicated with an "*" have fewer items):

- Idealistic Distortion*, measuring partners' tendency to answer questions in an unrealistically positive manner;
- Marital Satisfaction, using one item from each of the 10-item scales as a global measure of satisfaction;
- Personality Issues, examining the individual's satisfaction with his/her partner's behaviour;
- Communication, concerning the individual's feelings and attitudes toward communication in marriage (comfort levels in sharing and receiving emotional and cognitive information);
- Conflict Resolution, assessing the partner's perception of conflict and conflict resolution in the relationship;
- Financial Management, focusing on attitudes and concerns about how economic issues are managed;
- Leisure Activities, assessing preferences for spending free time;
- Sexual Relationship, examining the partner's feelings about the affectional and sexual relationship;
- Children and Parenting, assessing attitudes and feelings about having and raising children;
- Family and Friends, assessing feelings and concerns about relationships with relatives, in-laws and friends;
- Equalitarian Roles, focusing on individual's feelings and attitudes about various marital and family roles (occupation, household, sex, parent);
- Religious Orientation, examining the meaning of religious beliefs and practices within the marriage;
- Marital Cohesion*, describing how close the partners feel towards each other (including the balance between separation and togetherness); and
- Marital Change*, describing how a couple is able to balance stability and change in their marriage.

Reliability (alpha coefficient and test-retest) and validity (construct and discriminant) statistics were acceptable.

The decision not to use the instrument was based on three arguments. The instrument too specifically relates to marital situations. This implied substantive adaptation and restandardisation before use. Moreover, it is too long, requiring much effort in terms of application time for productive yield in comparison to the *DAS*, which is well-documented and short and covers all the important content of the *ENRICH* Inventory. Last, one cannot be sure that all 14 the scales cover unique, distinguishable factors with separate meaning.

❖ *Personal Assessment of Intimacy in Relationships (PAIR)*

Schaefer and Olson introduced this instrument in 1981. It tested couples' social, sexual, intellectual, recreational and emotional intimacy, among others. No copy or reports on its psychometric properties could be obtained. From the work of Venter and Zeelie (1992), however, it became clear that a few valid apprehensions existed. It covered a too limited aspect of relationships (intimacy), which was

nevertheless expanded into at least seven sub-scales. It is doubted whether they can all cover sufficiently unique constructs. The findings on the *PAIR* Inventory contradicted those of the *ENRICH* Inventory, suggesting a construct-validity problem. Also, some of the reliability coefficients were not acceptable (even lower than 0,60).

❖ *The Primary Communication Inventory (PCI)*

This inventory was designed by Locke, Sabagh and Thomas in 1956, and revised by Navran in 1967 (Venter & Zeelie, 1992). It was decided not to investigate its potential use in the present study further, after some immediate misgivings had become apparent. The scale covers only limited aspects, namely communication and empathy in relationships. It also suffered from low reliability coefficients (below 0,60) on some sub-scales.

❖ The *Marital Satisfaction Inventory (MSI)*

Snyder produced the *MSI* in 1981 (Juni & Grimm, 1994). The main reason why it has been investigated in more detail, is because it represents a multi-dimensional scale, in this sense contrasting with the uni-dimensional *DAS*. Its 280 true-or-false items evaluate nine scales, namely:

- Global Distress, estimating unhappiness with the marriage, and lack of commitment to the relationship;
- Affective Communication;
- Problem-solving Communication;
- Time Together, measuring quality and quantity of leisure time together;
- Disagreement about Finances;
- Sexual Dissatisfaction;
- Role Orientation, evaluating problems related to the adherence to marital and parental roles;
- Dissatisfaction with Children; and
- Conflict over Childrearing.

The test manual gave the values of Cronbach's alpha coefficients of internal consistency as ranging from 0,80 to 0,97 (mean of 0,88). Test-retest reliability coefficients ranged from 0,84 to 0,94 (mean of 0,89). In Juni and Grimm's (1994) study, the scale scores represented two sets of issues (factors?), namely Child Issues and a general measure of Dyadic Satisfaction. They can be added to give a composite measure of marital satisfaction. Alpha coefficients for these three clusters respectively were 0,71, 0,77 and 0,84, for men, and 0,71, 0,73 and 0,81, for women (supporting their internal consistency).

The following reservations tipped the scales against the instrument's inclusion in the present research. It is far too long, and thus not sufficiently cost-effective. This also refers to the yield of quality data, which are at risk when respondents get frustrated or bored because of repetition in the items. The instrument clearly also evaluates some specific marital relationships and issues. It may also focus too much on children's issues, while many couples nowadays may prefer not to have any.

In essence, its non-coverage of relationship satisfaction as a developmental process of adjustment is suspect, because it contradicts the fluid nature of relationships. It thus fails to satisfy the accepted standards set by the *DAS* of retaining the best from the generalist and specialist approaches. The recent reliability coefficients could have been higher, besides the fact that only internal consistency statistics were given, instead of also indications of test-retest reliability and validity.

❖ The *Relationship Assessment Measure for Same-Sex Couples (RAM-SSC)*

Burgoyne (2001) adapted the *Waring Intimacy Questionnaire*, intended for heterosexual, married couples, for same-sex couple evaluations. Waring and Reddon developed this original instrument in the early 1980s. The scales, each measured by 10 true-false items, relate to:

- Conflict Resolution (ease of resolving differences of opinion);
- Affection (degree of expressing feelings of emotional closeness);
- Cohesion (feelings of commitment to or belonging in the relationship);
- Sexuality (degree of communication and fulfilment of sexual needs);
- Identity (level of self-confidence and self-esteem);
- Compatibility (ability to work and recreate together comfortably);
- Autonomy (relationship to outside parties such as parents, children and friends);
- Expressiveness (degree of sharing thoughts, beliefs, attitudes and feelings; self-disclosure); and
- Social desirability (favourability / positivity response bias).

Although the instrument could distinguish between clinical and non-clinical respondents, the samples were quite small, and no reliability and validity evidence could be provided.

In summary, the choice was made for the *DAS* and a few supplementary single-question items. The choice is supported by a few main reasons, which include:

- the *DAS* extensively analysed and replaced the existing instruments before it;
- the battery is very cost-effective, using a limited number of items with great ease of application and scoring to yield quality data; and
- the *DAS* has exhibited highly sound psychometric properties over time in various studies.

The latter fact was recognised in a recent study by Calahan (1997), who compared the *DAS* to the *Kansas Marital Satisfaction Scale*. Calahan preferred the latter instrument, with the *Quality Marriage Index*, for his research for their brevity and simplicity of administration.

The following few main reasons exist against the use of the other instruments that have been introduced:

- they often relate to relationship aspects that are too narrow or limited in scope;
- they are not cost-effective, owing to their length or difficulty in scoring;
- their psychometric properties are either not well-documented, or the coefficients reported are low; and
- their items are often limited to marital situations, and would require extensive adaptation before being ready for use in same-sex situations.

7.2.2 Sex-role identity

About 15 measurement instruments have been used in previous research to evaluate sex-role identity (see Column E of Table 3.1). In about half of these studies, the *Bem Sex Role Inventory (BSRI)* was administered. As with the *DAS*, this figure reveals the wide acceptance and use of the *BSRI*.

Almost the same requirements as those set for relationship satisfaction (7.2.1) were set for the measuring instrument of sex-role identity. Such measurement technique had to comprise:

- a well-accepted, classical instrument, that could also serve as anchor measurement;
- an instrument sufficiently covering the core concepts in line with a well-established theoretical position;
- an instrument suitable for measuring sex-role identity among males and females equally well;
- the availability of a second rating tool to enable cross-validation of scores; and
- a cost-effective combined battery of assessments in terms of yield of quality information for effort.

The literature review made it very clear that the *BSRI* is the instrument for measuring sex-role identity. It is commonly acknowledged as the standard in the field. Documentation on its reliability and validity suggests, as more fully shown below, that a compact additional technique is all that is needed for cross-validation purposes.

7.2.2.1 *Bem Sex Role Inventory (BSRI)*

Arguments are submitted here for sex-role identity (traits and type) to be measured by means of the well-documented and accepted *Bem Sex Role Inventory (BSRI*) (Bem, 1975; 1978; 1979; 1981a, 1981b, 1985).

Essentially, the *BSRI* assesses individual differences in gender schematicity. Sex-typed persons (those with masculine or feminine sex-role identity types) are considered gender schematic, while androgynous and undifferentiated persons are not. The median-split scoring method was used to establish this. After about 1987, however, quite a few studies failed to replicate previously reported findings (based on the *BSRI*), or did not support hypotheses derived from schema theory (Deaux & Kite, 1993). In addition to measurement inconsistencies, the criteria used to define gender schematicity were questioned. In the late 1970s and early 1980s, Markus proposed that her ~~self-schema theory~~ be extended to gender-related domains, implying that masculinity and femininity should be separate self-schemas. Accordingly, a person would be masculine schematic, feminine schematic, neither, or both. The failure of researchers to distinguish between the simple use of sex as a biological category, and more subtle inferences about psychological gender, may have caused the confusion underlying some of the criticisms. At the time, further research was called for.

Bem (1993), in answer to many criticisms regarding the use of the concept of androgyny, even femininity and masculinity, and specifically the application of the *BSRI*, argued that the technique was still valuable. Respondents are unaware of the categories (masculinity, femininity, and fillers) represented by the items, and even more so of the fact that the technique actually attempts (not to classify individuals into sex types, but) to uncover the salience with which gender schemata operate

during completion of the task. Many features of the *BSRI* still directly challenge some of the most problematic assumptions made in studying the nature of maleness and femaleness, -- assumptions given specific scientific legitimacy by the work of Terman and Miles many years before (as early as 1936).

First, Terman and Miles reified mental masculinity-femininity as a deep-seated, enduring aspect of the human personality, and tested it likewise in individuals. The way in which the *BSRI*'s masculinity and femininity items were derived and operated, was one whereby informed judges from a culture rated the items in terms of their desirability or stereotypicality, putting it into the discourse of culture, rather than the personality of the individual.

Second, masculinity and femininity were polarised as mutually exclusive (or bipolar) ends of a single dimension. The *BSRI* scoring strategy assumed that masculinity and femininity represented two fully independent scales of culturally defined entities. The "either ... or" thinking was replaced with a "both ... and" one.

Third, direct links were theorised between sex-role identity, homosexuality and psychological normality. Instead of negatively contrasting sex-typed persons with their abnormal inverts, the *BSRI* positively held the contrast group to have a very acceptable range of androgynous traits.

And so, after at least 35 years, during which the Terman-Miles assumptions at the basis of the measurement of masculinity-femininity flourished, the *BSRI* reflected the changes that had taken place.

A number of informative studies of the *BSRI* followed. Blanchard-Fields, Suhrer-Roussel and Hertzog's (1994) work is reported at some length, because of the radical nature of their investigation. They used the data collected from 671 participants in their psychometric analyses. These included LISREL (confirmatory factor) analyses to evaluate the item factor structure of the *BSRI*, the second-order factor model for the item factors, and structural equation models to estimate age and gender differences in the first- and second-order factors.

The researchers questioned almost all previous findings indicating the multi-dimensionality of the gender-identity construct measured by the *BSRI* and emphasising the existence of anything from two to seven first-order factors. Such previous findings were criticised for: almost exclusively using exploratory factor analyses, using principal components to extract factors, and using *Varimax* rotation to an orthogonal solution. However, this approach was considered questionable because of the high likelihood that the *BSRI*-measured attributes were correlated and therefore not statistically orthogonal. This would result in erroneous inferences regarding the identification of factors.

Some previous studies availing themselves of confirmatory factor analyses were cited. The study of Marsh and Myers, reported on in 1986, posed that femininity and masculinity were higher-order constructs, each defined by a number of traits. Femininity was proposed to include the following

characteristics: emotional, dependent, nurturant, traditional, empathetic (sensitive to others), verbally expressive, and romantic. Masculinity would cover: aggressive/dominant, self-sufficient, competitive/assertive, physical/athletic, rational, goal directed, and tough. They found masculinity and femininity to be positively and significantly correlated (0,58), but not as polar opposites. However, the Windle and Sinnott study of 1985 failed to detect the two-factor structure, as well as correlation between the two factors.

In response to the latter finding, the study of Blanchard-Fields *et al*. (1994) set out to achieve several objectives. It planned to identify an appropriate confirmatory factor analysis model with a relatively large number of first-order factors. Then it would test whether these could be modelled with a single second-order bipolar masculinity-femininity factor, or in terms of two second-order masculinity and femininity factors (as originally conceptualised by Bem). The factorial invariance across gender groups would be assessed next. The relationship between gender and the constructs of (scores on) masculinity and femininity would also be evaluated. In general, the conceptualisation of the second-order factor(s) would have to receive attention.

The 20 filter items as well as four additional (traditionally feminine) adjectives not correlating with other *BSRI* items were omitted in the analyses. Models were first searched in an exploratory half-sample (n = 336), after which they were cross-validated in the validation half-sample (n = 335). Re-estimation of the final model was done afterwards on the total sample, as were gender comparisons of the factor structure and the run of structural equation models.

The three researchers found that males rated themselves higher on masculinity and females on femininity. The correlation between masculinity and femininity for males was -0,49 and for females -0,42 (aggregate -0,87).

A ten-factor model (first-order items) achieved the best fit. Three of the *BSRI* items were modelled as single indicator factors (“masculine”, “feminine” and “analytic”). The other seven factors exclusively modelled *BSRI* masculinity (four factors) and femininity (three factors) items. They were: “decisive”, “self-sufficient”, “athletic”, and “dominant”, as against “interpersonal affect”, “shy”, and “compassionate”. (Cf. Table II in Blanchard-Fields *et al*., 1994, pp. 436-437, for details.) The standardised factor loadings were substantively similar for males and females. Factor correlations were also similar in pattern for males and females. High correlations were found between “decisive” and “dominant”, and “interpersonal affect” and “compassionate”. Thus, gender equivalence in factor loadings could be accepted, although not to the degree of the existence of equal factor covariance matrices between the groups. Analyses to localise the differences showed that they were associated mostly with the items (adjectives) “masculinity” and “femininity”. Thus, males and females differed in the magnitude of correlations between the masculinity and femininity items, and the other *BSRI* items. Some less striking differences were found between males and females regarding the correlations between “dominant” and “decisive”, “dominant” and “athletic”, and “compassionate” and “self-sufficient”. In summary, the same factors are present in both males and females, and these factors have equivalent

relationships to *BSRI* items. However, they differ somewhat in their interrelationships, particularly with masculinity and femininity self-ratings.

The researchers in the end preferred a second-order factor model specifying two factors (masculinity and femininity), with “dominant” also loading on femininity (negatively correlated), and “shy” only on masculinity (negatively correlated). Although alternative specifications were possible, the fit of the revised two-factor model was considered sufficiently good for use in structural regression analyses, even though some possible additional relationships between *BSRI* item factors were not fully accounted for. The relationships of age and gender (considered together) to *BSRI* factors were small.

All in all, the revisit of the *BSRI*, reported above, seemed to strongly support the *BSRI* as a valid and appropriate assessment tool for two multi-faceted factors, namely masculinity and femininity. These two factors, as psychological constructs (traits) for sex-role identity (type), were also sufficiently discernible from the notion of biological sex (as reflected in gender identity within the items "masculine" and "feminine", according to the Blanchard-Fields research team). However, Blanchard-Fields *et al*. (1994) confused some matters. They argued that Bem and the *BSRI* subscribed to masculinity and femininity as the bipolar opposites of a single continuum/dimension. This is a notion that Bem rejected very early in her work! A valid criticism is that some items previously considered to be associated with femininity as a construct, may also be conceptualised as negatively weighted masculinity items (for example, “shy”), or the inverse.

Harris (1994) conducted investigations into the validity of the *BSRI* as an indicator of current cultural definitions of masculinity/femininity. He also wanted to establish whether such cultural definitions, as indicated by the *BSRI*, as well as the social desirability of the items, differed between the various American cultural groups. The questionnaire tasks to be completed were changed accordingly. Factor analysis largely supported the original construction of the *BSRI*. Harris (1994, p. 263) concluded that the *BSRI* "appears to still be a valid indicator of ‘American’ cultural definitions of masculinity and femininity today". Slightly different notions of masculinity and femininity were detected between the various sub-groups, though. Male subjects (African-American, Hispanic and Anglo-American) always rated masculinity in males significantly higher, compared to their female counterparts. However, African-American male subjects were the only sub-group that rated femininity in females significantly higher compared to their female counterparts, where only the other two sub-groups rated femininity highest. The African-American male subjects’ ratings of both masculinity in males and femininity in females were always the lowest, relative to the two other sub-groups, while their female counterparts’ ratings of masculinity in males and femininity in females, were always the highest. In short, this pattern entails that African-American subjects considered more of the traits (masculine and feminine) common to both sexes. No Hispanic respondents outperformed their Anglo-American counterparts in any rating, invalidating the historically held assumption about extreme gender stereotyping in that culture (cf. the notions of *macho* and *machismo*). In addition, the African-American sub-group was the only one where both male and female subjects did not rate any of the masculinity items as significantly more desirable for men than for women, or the femininity items as

more desirable for women than for men (with the exception of “sensitive to the needs of others” and “tender” in the latter case).

Williams and D'Alessandro (1994) studied the median-split technique of scoring the *BSRI*. Their main concern about this scoring method was that a continuous variable was artificially dichotomised. This could lead to a reduction in available information, or the misclassification of individuals close to the median. This, in turn, may be the cause of reduction in measurement precision, underestimates of the magnitude of bi-variate relations, the lowering of the statistical power for detecting true effects, and lowered reliability. Some continuous-scoring procedures were compared with the median-split one. However, in the end the "examination of the relationship between each of these measures and a series of external criteria ... revealed no differences among the scoring methods" (pp. 477-478). This outcome provides some further confidence when using the *BSRI* with its conventional scoring technique in the present study.

Van Strien (1994) demonstrated that the *BSRI*, as a sex-role inventory, measured the degree of gender-schematic processing well. Preceding theoretical developments did not marginalise the instrument. (Also see Section 2.2.1.)

As part of more recent developments, Chung (1995) investigated the construct validity of the *BSRI* by studying 60 heterosexual and 63 **gay** men (from a large town in the US Midwest). The two groups were matched for age, socio-economic status, race, student status and educational level. Internal consistency (corrected item-total correlations and α-coefficients), discriminant validity and confirmatory factor analyses were employed. Findings were that:

- the *BSRI* was equally valid for heterosexual and gay men;
- the new data essentially replicated the psychometric data reported in the 1981 *BSRI* Manual; and
- the short form was more appropriate than the long form, because of the problematic femininity items in the long form.

Campbell, Gillaspy and Thompson (1997) also explored the construct validity of the *BSRI*, noting that some of the recent findings had been controversial. Confirmatory factor analysis methods were used on the data obtained from 791 respondents. They were undergraduate and graduate students enrolled at a large American university (in Texas). After testing several models through LISREL, also minding the findings from many other bi-variate correlation procedures and chi-square statistics, the researchers declared that:

- no fit to a single bipolar factor, as assumed in the pre-Constantinople era (before 1973), could be made;
- two essentially orthogonal factors (masculinity and femininity) were revealed by the fit; and
- the short form showed itself to be superior to the long form.

The replication study of Wilcox and Francis (1997) produced less optimistic findings. Their research participants were 236 adolescent females from North-east England, aged mainly 16 and 17. Only three each of the masculinity and femininity items were deemed (by half of the sample in each case) to

still satisfy the criteria for inclusion in the scale. However, the respondents were only required to rate the social desirability of the items for the respective sexes. This fact, the age of the respondents, and unique features of the English sample, could explain the findings, and would not necessarily imply an outright rejection of the instrument for measuring sex-role identity.

Based on their replication study, Holt and Ellis (1998) confirmed that all but two feminine adjectives still served as good indicators of gender-role perceptions. In fact, even the two "deviant" adjectives performed in the predicted direction, and significance was only marginal. This finding lends support to the validity of the instrument. The study was undertaken because of concern that role changes over the previous decade had rendered the *BSRI* suspect.

Katsurada and Sugihara (1999) evaluated the appropriateness of the *BSRI* (translated into Japanese through sound parallel and back-translation processes) for the Japanese culture. Their conclusion, that it was a valid instrument overall, lends support to arguments that the tool can be transported to quite different cultural contexts. Only three feminine items were cause for slight concern. The researchers involved 289 college students. However, Sugihara and Katsurada (1999), reporting on research involving 265 college students, found that there were no statistically significant differences between the Japanese male and female students with regard to both the masculinity and femininity scales. Also, students from both sexes scored slightly higher on femininity. Confirmatory factor analysis did not provide support for Bem's gender-role model either. The two researchers proposed that the unique nature of the Japanese culture, placing a high premium on internal strength for both sexes, as well as the limitations of using a small sample of students, could have contributed to the findings.

Auster and Ohm (2000) did more work on the *BSRI*. Their study was prompted by factors such as women's increased participation in the workforce and their higher educational achievement profile, as well as men's role changes with regard to occupying previously female-dominated positions and taking on more domestic chores. The impact of such factors on the rating of gender traits through the *BSRI* by American respondents was studied by replicating what Bem had done in 1972. Students (n=133, weighted to 100) were the participants. They were from a small private liberal arts college in a mid-Atlantic state. The findings revealed that almost all of the traits (two exceptions on femininity) still qualified as descriptions of masculinity or femininity in American society. However, with regard to desirability, large shifts took place. For masculinity, a minority (eight) of the items was still more desirable for the target gender (males) compared to the reference group (females). Femininity traits' desirability largely stayed the same. Overall (irrespective of the sex for which desirability was rated), the desirability levels of the traits increased slightly. The relative rankings of the stereotypes were remarkably similar to what they had been before. Last, male and female respondents' ratings of ideal traits for themselves, compared to prevalent societal stereotypes, seemed to start drifting apart. The continued value (validity) of the *BSRI* for evaluating societal stereotypes (and schemata) was essentially confirmed.

Hoffman and Borders (2001) pointed out that the form (long and short) and scoring method (median split and hybrid) made a difference to the classification of respondents. (See again Section 3.4, where it is indicated how using a student sample could have influenced their findings.)

In summary, the reasons for selecting the *BSRI* for the present study, are that it:
- is based on sound and recently updated theory;
- still is the classic instrument for measuring sex-role identity;
- exhibits good psychometric properties; and
- is cost-effective.

None of the alternatives to the *BSRI* has the track record or maintained frequency of reliable use.

Only slight adaptations have been required to change the *BSRI* into a self-administered technique for ascribing sex-role identity characteristics (traits) to one's partner, namely:
- changing the instructions to indicate who has to be rated (self or partner); and
- modifying the 7-point scale descriptions from true of "me" to true of "my partner", where relevant.

The rest of the procedures remained the same.

It was decided to use the Short Form of the *BSRI* (see Annexure 7.3) as the main measurement instrument for the independent variable of the present study, namely sex-role identity (type and traits). The decision was based on this form's greater cost-effectiveness, as it is only half the length of the original form. In addition, the very few suspect items reported from time to time, could be omitted in this way. This would enhance reliability and validity too. Antill (1983) documented some of the psychometric properties of the Short Form of the *BSRI*. Alpha coefficients of 0,85 and 0,86 respectively were reported for the masculinity and femininity scales.

7.2.2.2 *Adjective Checklist (ACL)* procedure adapted by Prinsloo

It was decided to use previous experience as the basis for dealing with the second measure of sex-role identity. Prinsloo (1990) reported in great length on the outcome of his exploratory research on sex-role stereotypes. During the main activity of comparing South African sex-role stereotypes with those in 25 other countries, much energy also went into studying sex-role identity measurements. Many techniques were analysed critically (Prinsloo, 1990, Chapter 4.1). Subsequently, personal efforts at measuring sex-role stereotypes followed the format of the 300-item *Adjective Checklist (ACL)* developed by Gough and Heilbrun in 1965. However, the adjectives that were selected for Prinsloo's research, were taken from known lists of sex-role stereotypes worldwide. Through analyses of the stereotypes, a number of adjectives could be derived for use in a Guttman-type adjective scale for the measurement of sex-role identity. Those adjectives would then, on the basis of being confirmed as stereotypes by the South African sample, validly serve as the items of a sex-role identity instrument. By selecting those items that describe themselves, future respondents' sex-role identity type could be determined.

Two aspects of the foregoing analysis of the ratings of the sex-role stereotypes were important for selecting the items for the sex-role identity measurement instrument of the present study. They were

the weight with which stereotypes had been endorsed, and the social desirability of each stereotype. Based on these, a Guttman-type scale could be created, whereby endorsement of the masculine and feminine stereotypes could be counted as a measure of respondents' sex-role identity. The adjectives were eventually selected in the way described below.

From the 171 adjectives tested, those that had M%-scores of above 66 (indicating endorsement of a masculine stereotype) or below 34 (endorsement of a feminine stereotype), were selected. Of these, only those items were retained that achieved S%-scores of above 50, indicating firm endorsement of the item as a stereotype by a majority of the sample. If the meaning of an item had not been clear to more than 95 % of the sample, the item concerned was also dropped at this stage.

The next major decision covered the effects of social desirability. The main criterion was to retain equal numbers of socially desirable and undesirable items for both masculinity and femininity. This was done to be in a position to monitor the effect of social desirability on participants' endorsement of the stereotypes. In addition, each masculinity and femininity item retained, had to be equally desirable (or undesirable) for males and for females. Its so-called m+%- and f+%-scores, therefore, had to differ only minimally, irrespective of being above or below 50. This measure was taken to minimise the effect that the different social desirability levels of masculinity or femininity stereotypes may have on respondents' selection of the latter items. [The complete list of 171 original items and their statistics can be found as Table 6.16 in Prinsloo (1990, pp. 159-162).]

In summary, the main arguments in favour of the decision to use this adapted *ACL* procedure, were that:

- the sex-role stereotypes on which it is based, were researched recently within the South African context, as also referred to in Chapter 3 (Prinsloo, 1990);
- the results of a thorough analysis of the effects of social desirability and cultural bias were available, implying that items that were suspect, could be avoided;
- its application and scoring is brief and easy, as that of most *ACL* procedures; and
- it is widely acknowledged as a useful technique for gathering data.

The scores calculated from the data gathered through this adapted *ACL* procedure can be used to validate those obtained through the *BSRI*. The instrument is reported in Annexure 7.4.

Having considered the arguments above, the *BSRI* and the adapted *ACL* technique were considered most suitable for measuring self-rated and ascribed sex-role identity as the independent variable of the present study.

7.2.2.3 Other sex-role identity instruments

Very few other sex-role identity instruments were identified during the course of the research. Brief references are made to a few that appeared promising, but in the end could not compare favourably with the assessment battery selected. They deserve mention because of the potentially useful perspectives they bring.

❖ The *Sex Role Trait Inventory (SRTI)*

The *SRTI* was developed by Street and Meek in 1980 (Street *et al.*, 1995). The researchers considered the *BSRI* inappropriate for assessing respondents other than the self, especially in their investigation of the stability of sex-role perceptions over time. The instrument was designed to evaluate respondents' ratings of five different "objects", being the ideal man and women, most men and women, and the self. The test stimuli are 33 masculine and 33 feminine traits, which have to be rated on a 5-point Likert scale. (For the "ideal-person" object, both men and women preferred an androgynous ideal woman. However, while women also preferred an androgynous ideal man, men preferred a masculine sex-typed ideal man. In as far as stereotypes were concerned, both women's and men's perceptions of most men and women in society were sex-typed. Men mostly rated themselves as androgynous, while women rated themselves as feminine.)

❖ The *Personal Attributes Questionnaire (PAQ)*

The *PAQ*, developed in 1974 by Spence, Helmreich and Stapp, was used by Todt-Stockman (2000) in another investigation of the consistency of sex-role stereotypes over time. She found that women appeared to shift towards instrumentality, away from expressiveness, from 1973 to 1997. Twenge (1997) did a large-scale meta-analysis of 63 studies using the *BSRI*, and 40 using the *PAQ*. Only American college students were involved in these studies. Similar to Todt-Stockman's finding, the masculinity scores of women on both instruments increased steadily over time. However, no time-effects were observed with regard to femininity. Incidentally, these changes also signalled an increase in androgyny scores among females. On the *BSRI* only, men's masculinity scores also showed a small increase over time, while a decrease in sex-differences (smaller gap between the masculinity scores of men and women) over time also became evident. Possible explanations for the shifts in female masculinity were the increasing numbers of mothers working outside the home, their daughters' greater career aspirations, and/or girls' increasing participation in sport.

7.2.3 Contributing or nuisance variables

Many questions had to be formulated to gather the information required for testing the influence of other contributing variables. These questions had to be designed in such a way that the hypothesised, theory-based connections between the contributing variables and the two main variables of the study, could be evaluated. These items not only cover what is conventionally called biographical data, but also information on intra- and extra-personal factors (see Sections 4.2 and 4.3).

The main criterion applied in formulating these questions, was their cost-effectiveness. Optimal information had to be gained through as few as possible relevant items. Their response options also had to be as simple as possible, in an easily applied and scored format that would require very little writing by participants.

A good example of the kind of information usually gathered to support theory development, can be gleaned from Strickland (1981). Strickland cited the research by Spanier and Lewis in 1980 on the implications of ~~Social Exchange (Interaction) Theory (SIT)~~ for research on marital quality. They tested the hypothesis that marital quality is enhanced under certain circumstances that are anticipated by theory. The following circumstances were considered relevant: certain minimum social and personal resources available for adequate marital role functioning; acceptable satisfaction with their life style among spouses; and enough rewards from spousal interaction. In terms of SIT, the interaction between two people is a function of the pay-offs each gets out of the relationship. Such pay-offs include aspects such as positive regard for the spouse, emotional gratification, effective communication, and sexual enjoyment. The measuring instruments, as a consequence, had to be designed in such a way as to enable sub-group comparisons between the marital quality scores of couples with low, medium and high scores for the hypothesised constructs. Inter-partner comparisons are also relevant here.

The circumstances and conditions listed below were considered sufficiently important for collecting information on. Annexure 7.5 can be consulted for the specific format and formulation of the various items, and for the relevant sections of the research instruments, covering such demographic and other contextual information.

❖ General information to enable structuring the data set for sub-group comparisons:

- Type of sub-sample: same-sex dyads (gay/lesbian); and heterosexual dyads, with the latter divided into happy and unhappy couples. (The inclusion of perturbed couples heeds Langis *et al*.'s (1991) call (see 2.4), and serves as an effort to increase the range of relationship satisfaction scores (see 5.2.5). Same-sex couples were not treated in the same way, because their range of relationship satisfaction scores are assumed to be wide already given the odds against which they have to function, and because they do not report readily for couple therapy given the high level of stigmatisation still assumed to exist.)
- Couple number: a unique, serially allocated number for each respondent.
- Partner number: separate codes to identify each partner in a dyad.
- Couple relationship satisfaction category: both satisfied; only husbands satisfied; only wives satisfied; one only of the males or females in a same-sex couple satisfied; and neither partner satisfied.
- Couple sex-role identity type configuration: identical; only femininity (trait) identical; only masculinity (trait) identical; and not identical (qualified further by at least one partner being androgynous; feminine; masculine; or undifferentiated).

❖ Intra-personal factors (also cf. Section 4.2):

- Academic achievement: reflected by participants' highest academic qualification. (This item could serve as an imperfect proxy for cognitive ability. However, it is intended more to assist exploration of the influence of different levels of academic standing between partners on the dyad's

relationship satisfaction. Practical considerations also obviated any effort to obtain well-founded intelligence or mental alertness scores.)

- Emotional/affective functioning: reflected by items for rating levels of moodiness and depression.
- Feelings or attitude towards sex: reflecting respondents' enjoyment of sex.
- Relationship type preference or attitude: egalitarian (modern); comradeship (modern); and traditional.
- Personality characteristics: reflecting participants' self-ratings on extravertion – introvertion; emotional stability – lability; and dependence – independence.
- Skill at solving conflict through communication: reflected by an item for rating the presence / lack of the skill.
- General life-satisfaction or personal happiness: reflected by an item for rating level of happiness.
- Job-satisfaction: reflected by an item for rating level of satisfaction.

❖ ~~Extra-personal factors~~ (also cf. Section 4.3):

- Sex: male or female.
- Age (development or life-stage): in intervals of 10 years.
- Age difference between partners: number of years that either male or female (or one same-sex partner) is older.
- Health status: reflected by an item for rating the level of illness or health.
- Socio-economic status: derived from personal and couple gross income before deductions.
- Family stage: categorised according to age of eldest child (cf. Section 4.3.5 for Categories I to VIII).
- Number of children: living in the household.
- Career / job patterns: both; either; or neither professionals (including double-earner, single earner, no-earner patterns (cf. Section 4.3.6).

❖ ~~Other demographic information:~~

- Marital status: married; cohabiting (first time; or second; or more).
- Duration of relationship: in completed, full years.
- Type and consistency of relationship of parents: heterosexual marriage; heterosexual cohabitation; homosexual dyad; divorced; and single (moderated by various sequences; number of changes; and duration of longest arrangement).
- Level of relationship satisfaction of parents: reflected by an item for rating the level of happiness.
- Mother tongue: home language.
- Living area: rural/farm; village/town; city/urban.
- Ethnic group: black; brown; Asian; white.
- Religious orientation and commitment: reflected by items for rating the fact and type of religious persuasion.

7.2.4 Background information and covering letter

Special attention was given to establish positive rapport with the research participants in an effort to increase the return rate of instruments and the quality of the responses. A concise, to the point covering letter accompanied the survey material. This document intended to:

- secure the participation (motivation) of the respondents;
- ensure them of the anonymity of their contributions;
- provide them with the necessary background to the research; and
- set clear instructions for dealing with the research materials and responses.

The letter also stated that only couples who had been in close personal relationships for at least three years were eligible to participate. This requirement was considered necessary to enhance the quality of responses on the personality dimensions and behavioural aspects in relationships. In order to recruit more participants, this requirement was lowered to two years in some cases.

Participants, who wanted to learn more about themselves or their relationships, or just receive feedback in a general sense, were provided the opportunity to leave contact details. It is not uncommon that projects, such as the present one, create expectations for further information, deeper insight, or even assistance. For those situations, and to follow responsible ethical practices, the required channels were made available.

The background information and covering letter is reflected in Annexure 7.6.

7.3 Conclusion

In conclusion of this chapter, it is submitted that the tools required for gathering data in support of a scientific contribution, in response to the objectives of the present research, have been selected, or modified where necessary, to satisfactory standards. The literature surveyed has been helpful in mapping the theoretical, methodological and empirical dimensions of the field of study, making it possible to define the research problem, and develop or select the appropriate tools to gather information that, hopefully, will lead to answers to the research questions and objectives.