ANNA NIEBOER, SIEGWART LINDENBERG, ANNE BOOMSMA
and ALINDA C. VAN BRUGGEN

# DIMENSIONS OF WELL-BEING AND THEIR MEASUREMENT: THE SPF-IL SCALE[★]

ABSTRACT. What are the dimensions of well-being? That is, what universal goals need to be realized by individuals in order to enhance their well-being? Social production function (SPF) theory asserts that the universal goals affection, behavioral confirmation, status, comfort and stimulation are the relevant dimensions of subjective well-being. Realization of these substantive goals and the perspective on opportunities to realize these goals in the future contributes to the affective and cognitive component of well-being. The theoretical elaboration of this theory has been published elsewhere. This paper provides a measurement instrument for the dimensions of well-being. To measure levels of affection, behavioral confirmation, status, comfort and stimulation and empirically validate the dimensions of well-being, the SPF-IL scale was developed. This paper presents findings from a pilot study ($n = 145$), the main study ($n = 1094$), a test–retest examination ($n = 163$), and a validation study ($n = 725$). The measurement model was tested by means of structural equation modeling. Confirmatory factor analysis supported the dimensional structure of well-being indicating construct validity. The overall fit of the model was sufficient, in spite of the somewhat problematic measurement of status, and the test–retest study showed an acceptable level of stability. As for the content validity of the dimensions and their measurement, various sub-studies showed that the SPF-IL scale is a valid instrument, doing at least as well as popular measures of overall well-being but also specifying its dimensions.

KEY WORDS: measurement instrument, structural equation modeling, subjective well-being

## INTRODUCTION

Most researchers agree on the distinction between a cognitive and an affective component of well-being (Andrews and Whitley, 1976;

Lucas et al., 1996; Offer, 1996; Diener and Suh, 1997; Vitterso and Nilsen, 2001). Satisfaction with life is often used as a proxy for the cognitive evaluation of one's life (Diener et al., 1999; Eid and Diener, 2004). The affective component of well-being is mostly measured as a balance of positive and negative emotions (e.g., Bradburn, 1969; Watson et al., 1988). As such, these concepts only focus on very general dimensions of well-being. There has also been a discussion on the direct measurement of well-being. Kahneman (1999) for example is quite critical of direct questions on well-being because individuals are rather biased by their present state and what they remember at that moment. Instead, Kahneman suggests employing experience sampling in which the states the individual is in when answering direct questions on well-being are randomized. This method may indeed work (and has had already some success, see Csikszentmihalyi, 1992, 1994; Stone et al., 1999) but there is also another, more practicable possibility. One can try to identify the major dimensions of subjective well-being and ask questions that refer to activities and experiences related to these dimensions. In this way, one can avoid direct questions on well-being and approach well-being by the instruments individuals have at their disposal to reach it. There should be less memory bias for these questions. This approach has the added advantage that one has a much better understanding of the circumstances that affect well-being. For example, one would be able to compare groups or countries along the dimensions of well-being, one could trace over time whether loss in one dimension is compensated by increase in another, and for policy analysis one would have a good starting point to diagnose in what dimensions improvement would have the largest effect on overall well-being. These advantages depend on two things, a specification of the dimensions and a fairly secure measurement of the dimensions (see Land, 2001; Noll, 2002). In this article we try to do just that and develop a more specific and detailed conceptualization and measurement of the dimensions of subjective well-being. Assuming that subjective well-being reflects the extent to which people's goals (or needs) are obtained (or met), we base our dimensionalization of well-being on a theory that specifies people's universal highest order goals and their universal instruments for reaching the highest order goals. In fact, following the theory of social production functions (SPF, for short, see below), we concep-

tualize well-being as the degree to which people are able to realize these universal instruments.

In this paper, we will neither go into a comparative evaluation of SPF theory against other theories nor into a detailed theoretical elaboration. This has been done in a number of other publications (Lindenberg, 1996; Ormel et al., 1997, 1999; Van Bruggen 2001). Rather, we take SPF theory as given and proceed as follows. We put much effort into refining the conceptualization of the dimensions suggested by SPF theory and into creating secure measures for these dimensions. This included an extensive qualitative study (not reported here) as preparation of the operationalization, two pilot studies, a main study on a representative Dutch sample with confirmatory factor analyses, and a test–retest study. After reporting on these efforts, we present the validation studies, consisting of a confrontation of SPF-IL with traditional measures of well-being, of subgroup comparisons, and of a separate validation study 6 months after the main study.

## SPF THEORY

The theory we use is known as SPF theory (Lindenberg, 1986, 1990, 1996; Ormel et al., 1997, 1999). It has been applied in a variety of research projects on, for example, child care (Wielers, 1991), successful development and aging (Steverink, 2001; Ormel, 2002; Frieswijk, 2004; Schuurmans, 2004), problem behavior in schools (Van Liere, 1990; Kassenberg, 2002), division of labor in the household (Van der Lippe, 1993; Van der Lippe and Siegers, 1994), the level of community in Local Exchange Trading Systems (Hoeben, 2003), the consequences of life-events (Nieboer, 1997; Nieboer et al., 1998, 1999; Nieboer and Lindenberg, 2002), and the social impact on depressive symptoms (Meertens, 2004). SPF theory asserts that people try to improve their life situation by optimizing two universal goals (physical and social well-being) and five instrumental goals by which these universal goals are achieved (stimulation and comfort for physical well-being, and status, behavioral confirmation and affection for social well-being). In order to avoid confusion with other instrumental goals, we call these five goals "first-order" goals.[1] Levels of goals below this level are assumed to get progressively more

idiosyncratic to a particular culture, group or, finally, even a particular person.

It is a frequently observed reaction of readers to a list of goals that they would be able to come up with a completely different and yet equally plausible list of goals within 5 minutes of reflection. In our experience, this generally turns out not to be the case. The temptation comes primarily from the fact that in a goal hierarchy many things come to mind that people may want but they are either aspects of the goals stated (such as "respect"), or general means lower in the hierarchy and thus considered important but only in certain groups or cultures and only as instruments for reaching the first-order goals (such as money, power, autonomy, beauty, competence, control). To actually come up with a plausible alternative list for the first-order goals has proved difficult, though we do not claim that it is impossible. SPF theory, and especially the refined conceptualization used in this paper, has been very carefully crafted with an eye to many relevant discussions (Diener et al., 1999; Kahneman, 1999) and other goals theories that link well-being in some way to objective conditions, including, to name a few, Maslow's (1970) need hierarchy, the basic need theories of Doyal and Gough (1991), Max-Neef (1992), Fei et al. (1979); the multiple discrepancy theory by Michalos (1985); the resource theory by Schulz (1995); the life domain theory by Andrews and Withey (1976); the time use approaches by Dow and Juster (1985) and also Heady and Wearing (1989) (for a detailed discussion see Van Bruggen, 2001). The present paper also takes these discussions of goals a step further into measurement, validation and comparison.

In SPF theory, a theory of goals is integrated with a theory of behavior (which may be called a theory of "social rationality", see Lindenberg, 2001). The behavioral theory assumes that individuals take an active role in pursuing their goals intelligently (i.e., they do consider scarcities and differences in efficiency, they search for opportunities and they substitute), but that their rationality (beliefs and expectations, definition of the situation, and ways to come to a decision) is bounded and heavily influenced by social circumstances.

Subjective well-being is seen as an overall state of well-being that is determined by a person's ability to obtain the universal goals of physical and social well-being and their perspective on the future. In turn, social well-being is assumed to be realized through three forms

of social approval that have been widely acknowledged in the literature: affection (i.e., feeling loved); behavioral confirmation (i.e., belonging and doing things right); and status (i.e., distinction in valued aspects, such as skills, education, wealth). For the enhancement of physical well-being, it is assumed that people seek both comfort (i.e., reduction in noxious stimuli) and stimulation (both physically and mentally). All five first-order goals are assumed to have decreasing marginal value for the realization of well-being[2] (for descriptions of the first-order goals see Table I).

The structure is hierarchical[3] with, in descending order, subjective (or psychological, or overall) well-being on top, then physical and social well-being, then the five first-order goals below which we have a hierarchy of activities and endowments each of which is useful for realizing one or more goals higher in the hierarchy (see Ormel et al., 1999). Because people substitute ways to achieve well-being when the available means change, their level of well-being can be expected to be rather stable over time, and this is reported in the literature (McCrae and Costa, 1984). If, for example, our physical health deteriorates, intensifying our ties with loved ones buffers the negative effect on how we feel (Cohen and Wills, 1985; House et al., 1988). People are also sensitive to the efficiency and compatibility of means (activities and endowments) to achieve specific goals. This implies that the most important contributions to the well-being of an individual are his or her *multifunctional* means which contribute compatibly to all or most dimensions of well-being. For example, for most people in modern Western societies, an intimate partner is likely to be a multifunctional source for the achievement of all five first-order goals, providing comfort and stimulation, but also affection, behavioral confirmation and at least as much status as is compatible with affection. For many persons, work is also multifunctional. Lack or loss of multifunctional means is thus particularly damaging for subjective well-being.

## METHOD

Thus far, operationalizations of affection, behavioral confirmation, status, comfort and stimulation in the various studies done have been rather *ad hoc*. To develop secure measurement instruments, we conducted a qualitative study (with focus groups) and a quantitative

TABLE I

Overview of the measurement instrument

| Goal | No. of items | Shorthand description of major aspects of the concept | Item numbers (see Appendix A) |
|---|---|---|---|
| Affection | 18 | The love one gets for who one is as a person, regardless of one's assets or actions | 1–18 |
| Behavioral confirmation | 12 | The feeling to have done "the right thing" in the eyes of relevant others | 19–30 |
| Status | 12 | Social approval given on the basis of the command over scarce resources relative to others (e.g., money and education) | 31–42 |
| Comfort | 8 | The absence of deleterious stimuli (i.e., physiological discomforts such as pain, thirst, hunger, or cold) | 43–50 |
| Stimulation | 8 | Activation which produces arousal, including mental and sensory stimulation and physical effort | 51–58 |

study. The results of the qualitative study (see Van Bruggen, 2001) enabled us to come to a more refined conceptualization of the different dimensions of well-being and to operationalize these central concepts of SPF theory. A pre-pilot study ($n = 326$) was conducted to adjust shortcomings of the questionnaire. Next, a pilot study among 145 respondents was used to select a comprehensive set of items, jointly forming the SPF-IL Scale that measures the dimensions of well-being.[4] In this paper we report on the pilot study ($n = 145$), the main study ($n = 1094$), a test–retest study ($n = 163$), and a validation study ($n = 725$).

## Sample Characteristics

For the *pre-pilot study* a random sample was drawn from Dutch postal addresses. By letter, eligible subjects were asked to cooperate when approached for a phone interview a few days later. The interviewer was instructed to interview the member of the household between 18 and 65 years old whose next birthday is closest to the date of interviewing to avoid only interviewing persons who first answered the phone. The response rate was about 60% ($n = 326$). This procedure resulted in a relatively large proportion of women (62%). For the *pilot study* we contacted a sub-sample of 213 of the pre-pilot respondents who agreed to be interviewed again at a later stage (95%). A total of 177 respondents were reached (36 respondents of the original sample were not reached, 12 of them due to incorrect phone numbers) and 145 of them (82%) cooperated a second time when contacted by the research team.

The sample of the *main study* consists of persons aged 18–65 years. Here we also randomly drew a sample from Dutch postal addresses and asked eligible subjects by letter to cooperate when approached for a phone interview a few days later. To avoid overrepresentation of women, the youngest male member of the household currently at home was interviewed. If not present, the youngest female was interviewed (Hess, 1994). The interviews lasted approximately 25 minutes. All interviews were carried out by well-trained social science students. Useful data are available for 1094 subjects, the response rate was 59% (53% female).

For the *validation study*, 1031 of 1094 respondents (94%) who agreed to be contacted for future research were approached 6 months

after the main study. We reached 928 of them and 725 (78%) agreed to cooperate a second time. The test–retest for the full battery of questions was conducted for 163 respondents of the validation study (51% female).

## DESCRIPTION OF THE INSTRUMENT

The SPF-IL is a multidimensional instrument to measure the first-order goals that enable people to realize well-being (for a complete list of items see Appendix A). The frame of reference in the questionnaire is 3 months.

*Affection*: For affection, questions are asked about the extent to which people feel liked, loved, trusted and accepted, understood, empathized with, know that their feelings are reciprocated, feel that others are willing to help without expecting something in return, feel that their well-being is intertwined with others, and feel that others like to be close and hug them (Van Bruggen, 2001). Eighteen items were selected to assess nine different aspects of affection. Each aspect included a positive and a negative item. Examples of items are: 'Would you say it is difficult for people to put themselves in your shoes?', 'Do people really like you?', 'Do people act like they care very little about how you feel?'. The 18 items were coded on a 4-point scale with the categories 'never', 'sometimes', 'often', or 'always' (range: 0–3). The scores of the items were recoded in order to have higher scores reflect higher levels of affection. The overall scale score was transformed to a range of 0–100 (dividing it by the maximum scale score and multiplying it by 100).

*Behavioral confirmation*: The level of behavioral confirmation was conceptualized by six aspects, viz. feeling that you: do good things, do things well, are a good person, are useful, are part of a functional group, and contribute to a common goal. Examples of items to measure these aspects are: 'Do people think your contribution is inadequate?', 'Do people think that you make the right choices?', 'Do you feel useful to others?'. The instrument consists of 12 items; the measure ranges from 0 to 100.

*Status*: The level of status is conceptualized by six aspects: the feeling of being treated with respect, being independent, self-realization, achievement as compared to others, influence, and reputation.

Examples of items are: 'Do people think that you do better than others?', 'Do people belittle your achievements?', 'Do people find you an influential person?'. The instrument consists of 12 items and the measure ranges from 0 to 100.

*Comfort*: The level of comfort is conceptualized as the absence of feelings of discomfort, such as pain or stress. Respondents were asked, for example: 'How often do you have pain?', 'In the past few months did you feel: ... fit', '... perfectly healthy'. The instrument consists of eight items. The measure ranges from 0 to 100.

*Stimulation*: The level of stimulation is conceptualized by mental and physical activation. Respondents were asked, for example, 'Do you find your life boring?', 'Are your activities challenging to you?', 'Do you really enjoy your activities?'. The instrument consists of eight items and the measure ranges from 0 to 100.

*Perspective on the future*: People's perspective on their future realization of well-being is measured by five items. First, a general question was asked 'At times, do you have a negative outlook on your own future? Please answer with yes or no.'; 'Do you feel that your financial situation keeps on improving?'; 'Do you feel that you continue to develop certain skills?'; 'Do you feel that your relationships with friends, family and acquaintances is getting worse, stays the same or is improving?'; 'In the next years, do you expect your health to improve, to stay the same or to deteriorate?'. The first three items only have the answering categories yes and no. The last item has been recoded to reflect perceived stability of one's health as opposed to anticipated positive or negative changes. A sum score of the items was calculated. The measure ranges from 0 to 6.

The overall measure of well-being is defined by a Cobb–Douglas function of the first order goals (see Lindenberg, 1996). The reason for this is that the first order goals are assumed to be "needs" up to a certain point after which they turn into "wants" in the sense that having little of one can be compensated by getting more of another of the five first-order goals. For example, with very little comfort, people's well-being would be very low as well, no matter how high the score on the other four goals is. For this reason, the measure must allow a small score on any of the five first-order goals to drag down the entire score for well-being. A Cobb–Douglas function is well-suited for expressing these relations, and having no particular theory (yet) to distinguish the weights of the five first-order goals, we give

each the same weight. In order to do this, the standardized latent constructs are transformed to a minimum of 0 (mean value 1) to yield the function

$$\text{Well-being}_{\text{SPF-IL}} = (\text{affection}^{0.2})*$$
$$(\text{behavioral confirmation}^{0.2})*(\text{status}^{0.2})*(\text{comfort}^{0.2})*$$
$$(\text{stimulation}^{0.2})$$

Further developments of this indicator of well-being might adjust the exponents in applications to particular groups or cultures. For example, for the US, status might be weighted relatively more heavily; for adolescents stimulation might get a higher weight. In principle, the exponents can be considered open parameters that can be estimated for particular populations.

### Other Measures

As part of the validation of the dimensions and their measurement, we also included traditionally popular measures of overall, *subjective well-being*. Even though direct measures have been criticized as subject to a memory bias, they have been used in much research on well-being in the past and proven very useful. The measure presented here does not directly ask for subjective well-being or life satisfaction but for the degree to which individuals believe they have realized various aspects which SPF theory identifies as making up subjective well-being. The advance consists of avoiding direct questions on well-being and concentrating on lower level but still universal aspects that supposedly jointly create a feeling of well-being. Still, there should be a considerable correlation of our new measure with these more traditional measures and for this reason we use them for validation.

There is a broad consensus that the affective and cognitive components of well-being are complements of each other. Overall subjective well-being is measured with respect to life satisfaction (supposedly measuring the cognitive component), and positive and negative affect (supposedly measuring the affective component). Cantril's Ladder (1965) is widely used to assess *satisfaction with life* and reflects a general, cognitive evaluation of a person's overall well-being. 'On a scale of 1 to 10, how satisfied are you with your life as-a-whole now?' A 10-item version of the PANAS (Watson et al., 1988)

was used to assess positive and negative affect. *Positive affect* consists of five items: During the past 3 months, how often did you feel … excited, enthusiastic, alert, inspired, and determined. Due to low scalability of the item on feeling excited, it was removed from the analyses. *Negative affect* consists of five items: sad, upset, afraid, nervous, and scared. For some of the analyses, SPF-IL was compared to an overall score of subjective well-being by adding the standardized scores of satisfaction with life, positive and negative affect.

Within SPF, the cognitive component consists mainly of the future perspective and the first-order goals that play their most obvious part for well-being when the individual is deprived of a satisfactory level: comfort (for example, having no pain, no hunger) and affection (not being lonely, not lacking friendliness around you). The affective component (positive and negative affects) consists mainly of first-order goals that provide a positive "kick" with improvement and anxious feelings with threatened loss: stimulation, status, and behavioral confirmation. Because SPF-IL combines the cognitive and the affective components, it should be compared to the combined measure of Cantril's Ladder and PANAS.

*Context-specific means*: To validate the SPF-IL we also studied the contribution of specific means to the realization of affection, behavioral confirmation, status, comfort and stimulation. Certain activities, endowments and situations should contribute to the first-order goals on face validity. Means included in the questionnaire concerning personal relations in the private sphere are: a partner; equal contribution of both partners; children; number of friends; difficult friendships; size of the network as compared to others; and success as compared to others in surroundings. Job-related means are: difficult work relations; clarity of tasks; and supervising others. For leisure time, we asked about the extent to which time is spent with others; the number of activities; and the number of vacations. In the context of impersonal relations, we asked about how easy it is to make contact; the willingness to help others; and whether or not people are impressed by your things. A particularly interesting step in the validation procedure is to compare the contribution these means make to SPF-IL with the contribution they make to well-being measured in the traditional way. Again, if these contributions are roughly similar, SPF-IL is to be preferred because it is linked to specified dimensions, whereas the traditional measures are not.

*Background variables:* included were age, gender and the number of chronic conditions (e.g., asthma, heart disease or diabetes).

## ANALYSES

For the construction of the scales for the main study, we first analyzed the pilot data. The main selection criteria to remove items were: (1) overlap of items within and between goals and (2) lack of contribution to the underlying dimension of well-being. In the end, we selected 58 items from a pool of 108. We included a positive and a negative item for each aspect of affection, behavioral confirmation and status (42 items) and 16 items to measure the level of comfort and stimulation.

The analyses consisted of three parts. First, analyzing the pilot data resulted in a selection of 58 items to reflect the level of affection, behavioral confirmation, status, comfort and stimulation. Descriptives at the item and scale level (i.e., mean, SD, item correlations and Cronbach's $\alpha$) are used to compare the results of the pilot and main study. Second, the measurement model was tested by means of structural equation modeling. Confirmatory factor analyses were conducted to test whether or not items load on the intended dimension. Overlap between items and dimensions could be traced via the modification indices. Potential model improvement by covariation of measurement errors was often indicative of overlap (i.e., correlation of error terms). To test the measurement model, the normal-theory weighted least squares $\chi^2$ statistic, the standardized root mean square residual (SRMR), the root mean square error of approximation (RMSEA) and the Incremental Fit Index (IFI) were calculated using cut-off criteria proposed by Hu and Bentler (1999). The analyses were based on an imputed data matrix using the expectation-maximization algorithm of LISREL 8.50 (Jöreskog and Sörbom, 1996; Du Toit and Du Toit, 2001). The imputed data matrix was used to avoid losing cases as a result of item non-response. We did not censor items with a kurtosis larger than 3 (items 5, 18, 21, 30, 40, 42, 44, 51, 57) because it did not really influence the covariance matrix. For the confirmatory factor analyses of the full model (58 items) we had to use maximum-likelihood estimates because the number of cases was too small to use a more robust estimation method accounting for non-normal distri-

bution of variables (required number of cases $58*59/2 = 1711$). However, for the 15-item version (see below) we were able to use robust ML-estimates, given an estimated asymptotic covariance matrix, which resulted in a still better model fit than the non-robust ML-approach (Table X).

## DESCRIPTIVE RESULTS

First, in Table II we report descriptive analyses of the scales for affection, behavioral confirmation, status, comfort and stimulation.

The descriptive results of the pilot study, the main study and the test–retest are similar. Cronbach's α for the five scales range from 0.68 to 0.88 for the pilot study and from 0.60 to 0.86 for the main study. For affection, behavioral confirmation, comfort and stimulation these estimates exceed the conventional standard of 0.70 (Nunnally, 1978), which points to one-dimensionality of these scales. In the case of status, Cronbach's α is a low 0.60 for the main study results. Moreover, internal-consistency reliability estimates for status differ between the two independent samples (the pilot and the main study), and is lowest for the retest study.

As for the missing values, only 1% of the respondents had a missing value on one or more of the items for the scale on comfort and stimulation. Thirteen percent of the respondents had a missing value on at least one of the items of both the affection and behavioral confirmation scale. In total 18% of the respondents had a missing value on one or more of the status items suggesting more difficulties in answering status related questions.

Correlations among the observed item and scale scores for affection, behavioral confirmation, status, comfort and stimulation are available at http://www.bmg.eur.nl/personal/nieboer. Item discrimination among scales was very good for comfort, stimulation, behavioral confirmation (except for item 25), and affection (except for items 3 and 5). For status, items only discriminated well when separate analyses were conducted for the positive and negative dimension (except for items 31 and 35). The scales for affection, behavioral confirmation, status, comfort and stimulation are not independent of each other (see Table Va), however. This may reflect the fact that individuals seek multifunctional means. For the

TABLE II

Descriptives of the scales for affection, behavioral confirmation, status, comfort and stimulation

| | # items | Pilot study (n = 145) | | | Main study (n = 1094) | | | Retest study (n = 163) | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Mean | SD | α | Mean | SD | α | Mean | SD | α |
| Affection | 18 | 72.3 | 10.7 | 0.84 | 73.8 | 10.3 | 0.79 | 72.0 | 10.4 | 0.81 |
| Behavioral confirmation | 12 | 76.6 | 10.1 | 0.76 | 77.9 | 9.2 | 0.71 | 75.9 | 9.1 | 0.71 |
| Status | 12 | 69.2 | 10.1 | 0.68 | 69.1 | 9.6 | 0.60 | 67.9 | 8.8 | 0.54 |
| Comfort | 8 | 68.8 | 18.5 | 0.88 | 70.1 | 19.1 | 0.86 | 71.5 | 16.6 | 0.85 |
| Stimulation | 8 | 76.7 | 12.4 | 0.76 | 77.2 | 14.2 | 0.79 | 76.8 | 13.4 | 0.80 |

Note: α means Cronbach's α.

measurement model such intercorrelations between scales are only problematic if they introduce (too) much error. In the next paragraph a test of the measurement model is discussed.

## TESTING THE MEASUREMENT MODEL

In order to test our measurement model we included all 58 items. When discussing a short-version of the instrument, we will suggest items that may be removed resulting in a better model fit. The results of the structural equation models are reported for the pilot and main study data in Tables IIIa and b, respectively.

Four criteria were used to test the model. First, the conventional overall test of goodness-of-fit assesses the discrepancy between the model implied and the sample covariance matrix by means of a normal-theory weighted least squares $\chi^2$ test. A plausible model has low, preferably non-significant $\chi^2$ values. Second, the RMSEA reflects the estimation error divided by the degrees of freedom as a penalty function. If it is smaller than 0.06 it points to small differences between the estimated and observed model. Third, we also used the SRMR, a scale invariant index for global fit that ranges between 0 and 1; lower scores reflect a better fit with a 0.08 cut-off value (Hu and Bentler, 1999). Fourth, we used the IFI. Here the independence model (i.e., observed variables are unrelated) is compared with the estimated model. Preferably, the IFI should be larger than 0.95. In the case of an IFI-score lower than 0.90, the specified relations between variables should not be considered as being supported by the data.

It is most important to look at the results of the main study (Table IIIb), because the item-selection process based on the pilot data is partly a result of capitalization on chance. The significant normal-theory weighted least squares $\chi^2$ statistic is not surprising given its sensitivity to sample size. The SRMR statistic however is below 0.08 for affection, behavioral confirmation, status, comfort and stimulation separately, as well as for the overall model. The overall model fit seems therefore sufficient. The RMSEA is higher than 0.06 for status and comfort (but is still well within the 90% confidence range). There clearly are differences between the estimated and observed model, but they are not very large. When we look at the IFI-index, one may notice that it is quite good for most goals but below 0.90 for status.

TABLE III

Factor loadings, normal-theory weighted least squares $\chi^2$ statistic, SRMR, RMSEA and IFI index for affection, behavioral confirmation, status, comfort and stimulation

| | $\chi^2$ | df | $p$-Value | SRMR | RMSEA | 90% CI | IFI |
|---|---|---|---|---|---|---|---|
| (a) *Pilot study:* ($n = 145$) | | | | | | | |
| Affection | 179.27 | 134 | 0.01 | 0.065 | 0.047 | 0.025–0.065 | 0.91 |
| Behavioral confirmation | 73.92 | 53 | 0.03 | 0.068 | 0.053 | 0.017–0.079 | 0.93 |
| Status | 83.52 | 53 | 0.05 | 0.091 | 0.067 | 0.040–0.091 | 0.89 |
| Comfort | 26.93 | 19 | 0.11 | 0.024 | 0.048 | 0.000–0.092 | 0.99 |
| Stimulation | 27.45 | 19 | 0.10 | 0.022 | 0.061 | 0.000–0.100 | 0.99 |
| Overall model | 2441.55 | 1550 | 0.00 | 0.084 | 0.053 | 0.048–0.058 | 0.80 |
| (b) *Main study* ($n = 1094$) | | | | | | | |
| Affection | 452.13 | 134 | 0.00 | 0.040 | 0.047 | 0.042–0.051 | 0.91 |
| Behavioral confirmation | 180.78 | 53 | 0.00 | 0.037 | 0.047 | 0.040–0.055 | 0.92 |
| Status | 350.69 | 53 | 0.00 | 0.069 | 0.072 | 0.065–0.079 | 0.81 |
| Comfort | 108.75 | 19 | 0.00 | 0.029 | 0.066 | 0.054–0.078 | 0.97 |
| Stimulation | 60.47 | 19 | 0.00 | 0.030 | 0.045 | 0.032–0.058 | 0.98 |
| Overall model | 4087.88 | 1550 | 0.00 | 0.052 | 0.039 | 0.037–0.040 | 0.85 |

*Note:* Positive and negative items are modeled as separate but correlated latent constructs.

As reported earlier, we assumed a relation between the positive and negative dimension of our status-measure, but here we did not find it. Without status, the IFI-index for the overall model is sufficient (0.90).

The completely standardized solution of the confirmatory factor model is presented in Table IV. A few items load rather low on the underlying dimension of affection and status. These confirmatory factor analyses point out that there is a weak relationship between these indicators and the respective latent constructs. The model fit improves considerably if we remove all items with a factor loading lower than 0.40 ($\chi_{1035} = 1854.07$, $p = 0$; SRMR: 0.045; RMSEA: 0.032; IFI: 0.91).

The associations between the latent constructs of the different goals are higher than between the observed total scale values (see Table Va) because they account for measurement error. The inter-correlations in Table Vb reveal high associations between the observed and latent scores with the exception of the negative dimension of status, pointing to low construct validity. The relatively low association between the observed and latent score of the negative dimension of status is probably due to large measurement errors. This is yet another indication that the measurement of status, especially via negative items, is problematic. The root of this problem may lie in the interpretation of "others" and "people" when items are negatively phrased. For example, items such as "Could you have accomplished more according to others?" or "Do people look down on you?" may create mixed reactions for lower status persons, since thinking of different sets of "people" may lead to different answers.

## TEST–RETEST STUDY

Structural equation modeling is ideal to analyze the variability in the dimensions of well-being because systematic instability is separated from measurement error by distinguishing observed and latent variables. Eid and Diener (2004) showed higher correlations for life satisfaction and mood without measurement error ($r = 0.74$ vs. 0.65). We carried out a test–retest after 1 week on an earlier version of the instrument ($n = 59$) and a test–rest after 6 months on the final instrument ($n = 163$). What should we expect? There is no firm ground

TABLE IV

Main study: Confirmatory factor analysis for the positive and negative items of affection, behavioral confirmation, status, comfort and stimulation

| | Affection | | Behavioral confirmation | | Status | | Comfort | | Stimulation | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Positive | Negative | Positive | Negative | Positive | Negative | Positive | Negative | Positive | Negative |
| 1st item | 0.39 | 0.37 | 0.47 | 0.47 | 0.39 | 0.59 | 0.76 | 0.61 | 0.70 | 0.69 |
| 2nd item | 0.38 | 0.48 | 0.50 | 0.54 | 0.35 | 0.23 | 0.55 | 0.69 | 0.71 | 0.59 |
| 3rd item | 0.61 | 0.34 | 0.38 | 0.52 | 0.54 | 0.27 | 0.79 | 0.74 | 0.72 | 0.72 |
| 4th item | 0.66 | 0.29 | 0.39 | 0.48 | 0.51 | 0.55 | 0.84 | 0.63 | 0.48 | 0.46 |
| 5th item | 0.54 | 0.56 | 0.50 | 0.45 | 0.62 | 0.39 | | | | |
| 6th item | 0.53 | 0.33 | 0.58 | 0.50 | 0.62 | 0.43 | | | | |
| 7th item | 0.43 | 0.52 | | | | | | | | |
| 8th item | 0.60 | 0.25 | | | | | | | | |
| 9th item | 0.68 | 0.61 | | | | | | | | |

TABLE V

Intercorrelations between the observed total scores and the latent scores of (a) and observed and latent scores of the positive and negative dimensions (b) of affection, behavioral confirmation, status, comfort and stimulation ($n = 1094$)

*Panel a*

| Observed | Latent Affection | Behavioral confirmation | Status | Comfort | Stimulation |
|---|---|---|---|---|---|
| Affection | [0.97] | 0.71 | 0.54 | 0.32 | 0.46 |
| Behavioral con-firmation | 0.55 | [0.98] | 0.67 | 0.25 | 0.52 |
| Status | 0.33 | 0.43 | [0.93] | 0.18 | 0.51 |
| Comfort | 0.28 | 0.21 | 0.14 | [0.99] | 0.43 |
| Stimulation | 0.37 | 0.43 | 0.42 | 0.37 | [0.99] |

*Panel b*

| | Affection | | Behavioral confirmation | | Status | | Comfort | | Stimulation | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Positive | Negative | Positive | Negative | Positive | Negative | Positive | Negative | Positive | Negative |
| | 0.977 | 0.916 | 0.951 | 0.964 | 0.977 | 0.890 | 0.982 | 0.979 | 0.974 | 0.977 |

*Panel a: Note:* All correlations $p < 0.001$. On the diagonal [between brackets] the correlations between the observed and latent scores are reported. Above the diagonal intercorrelations between the latent scale scores and below the diagonal between the observed total scores.
*Panel b: Note:* All correlations $p < 0.001$.

on which to expect certain degrees of stability or change. However, a few expectations can be generated. First of all, we expect that the measures scale as well at Time 2 as they did at Time 1 and that the means do not change significantly. Of course, we also expect the retest after 1 week to show more stability than the retest after 6 months. Second, we expect the highest stability in the dimensions that mainly represent the cognitive aspects (viz. affection and comfort), and less stability in the dimensions that mainly represent positive and negative affective aspects (viz. stimulation, status, and behavioral confirmation). Table II shows the descriptives (the first test–retest study is not included in this table, but there were no significant differences in the mean levels of the dimensions between Times 1 and 2). Indeed the scales show a stable mean and Cronbach's $\alpha$ when we compare main and retest studies, except for the 6 month test–retest of behavioral confirmation and status. A paired $t$-test for Times 1 and 2 did reveal a significant reduction in behavioral confirmation ($t$-value 2.8; $p < 0.01$) and status ($t$-value 3.2; $p < 0.01$). The Cronbach's $\alpha$ are all very acceptable, except, as we have observed above, the $\alpha$ for status which is quite low for both Time 1 and Time 2. As can be seen from Table VI, the intercorrelations among the five dimensions are acceptably stable over time, again with correlations involving status being less stable (see for comparison Badia et al., 1996; Eid and Diener, 2004).

Regarding the intercorrelations between the same dimensions over time, we found for the 1 week retest the following correlation coefficients for Time 1 and Time 2 (all significant with $p < 0.001$): affection 0.68; behavioral confirmation 0.65; status 0.80; comfort 0.80; stimulation 0.75). For the 6 months interval retest study, we found correlation coefficients for Time 1 and Time 2 (all significant with $p < 0.001$): affection 0.72; behavioral confirmation 0.64; status 0.58; comfort 0.62; stimulation 0.58). As expected, the stability declines over time and it declines most for status and stimulation and, surprisingly, also for comfort. The explanation for this may be that comfort may have a larger affective component than we thought and this is corroborated by the regression analysis (see Table VIIb). Behavioral confirmation changed less than expected. All in all, the test–retest studies provide no particular cause for alarm concerning the stability. This conclusion will be reinforced by the special follow-up validation study to be presented below. There, life events do have

TABLE VI

Main study and retest study, product–moment correlation matrix affection, behavioral confirmation, status, comfort and stimulation ($n = 163$)

| Main study | Retest Study | | | | |
|---|---|---|---|---|---|
| | Affection | Behavioral confirmation | Status | Comfort | Stimulation |
| Affection | [0.72] | 0.87 | 0.42 | 0.23 | 0.46 |
| Behavioral confirmation | 0.75 | [0.64] | 0.46 | 0.39 | 0.56 |
| Status | 0.53 | 0.62 | [0.58] | 0.20 | 0.36 |
| Comfort | 0.33 | 0.31 | 0.17 | [0.62] | 0.47 |
| Stimulation | 0.47 | 0.57 | 0.51 | 0.38 | [0.58] |

*Note*: For all correlations $p < 0.05$. On the diagonal [between brackets] the associations between the retest and main study data are reported. Above the diagonal correlations for the retest and below for the main study data.

the effect on the dimensions of well-being one would expect them to have.

## VALIDATION STUDIES

*First-order Goals and the Perspective on the Future*

As indicated above, we included some traditional measures of the cognitive and the affective aspects of well-being in order to help validate the new measure. Since SPF-IL combines cognitive and affective components, it should be compared to a combination of life satisfaction and PANAS scales. The intercorrelation between the traditional well-being measure and overall well-being as measured by SPF-IL is 0.63 ($p < 0.001$). As can be seen from Table VIIa, the first-order goals and the perspective on the future do indeed explain a good deal (42%) of the combined traditional measure of subjective well-being. As expected, the first-order goals – except for the relative low contribution of the status dimension – have a much higher impact than the future perspective.

We also tested whether the affective side of subjective well-being can indeed be conceptualized in terms of stimulation, status and behavioral confirmation, and the cognitive side primarily in terms of comfort, affection and future perspective. Table VIIb shows the results of this test. As expected, stimulation, status and behavioral confirmation are crucial for positive and negative affect (life's kicks or fears); and affection, comfort and future perspective are important to the cognitive side of subjective well-being (measured by life satisfaction). We also see that comfort and future perspective play an important role for (negative) affect and stimulation for life satisfaction. Seemingly, the pure distinction between cognitive and affective components does not hold. Individuals also worry about comfort (e.g., pain) and the future, and this shows up as negative affect. Individuals also seem to have a general impression of how interesting their lives are, which shows up as a contribution of stimulation to life satisfaction.

*Context-specific Means*

As an additional way to validate the instrument, we take a further look at some context-specific means. Again, well-confirmed and

TABLE VII

Multiple regression analyses of subjective well-being (a) and positive affect, negative affect and satisfaction with life (b) on affection, behavioral confirmation, status and the perspective on the future

| Well-being | $R^2$ change | F change | $\beta$ | |
|---|---|---|---|---|
| | | | Step 1 | Step 2 |
| *Panel a* | | | | |
| *Step 1* | 0.41 | 148.9 | | |
| Affection | | | 0.15 | 0.15 |
| Behavioral confirmation | | | 0.10 | 0.10 |
| Status | | | 0.06 | 0.06 |
| Comfort | | | 0.24 | 0.21 |
| Stimulation | | | 0.30 | 0.28 |
| *Step 2* | 0.01 | 12.8 | | |
| Perspective on the future | | | | 0.09 |
| Adjusted $R^2$ for equation | 0.42 | | | |

| Independent | Dependent | | |
|---|---|---|---|
| | Positive affect $\beta$ | Negative affect $\beta$ | Satisfaction with life $\beta$ |
| *Panel b* | | | |
| Affection | −0.01 | 0.05 | 0.26* |
| Behavioral confirmation | 0.06 | 0.13* | 0.02 |
| Status | 0.17* | 0.02 | −0.06 |
| Comfort | 0.00 | 0.29* | 0.16* |
| Stimulation | 0.27* | 0.11* | 0.23* |
| Perspective on the future | 0.03 | 0.08* | 0.09* |
| Adjusted $R^2$ for equation | 0.19 | 0.25 | 0.27 |

Panel a: *Note*: For all $\beta$ coefficients $p < 0.05$; Significance F change $< 0.001$.
Panel b: *Note*: * $p < 0.01$.

common sense expectations are tested against our data, this time expectations on lower level means for the realization of subjective well-being.

Multifunctional means, such as personal relationships and work should have a considerable influence on subjective well-being. As one can see from the correlational and multiple regression analyses in Table VIII, personal relations of people are indeed their most important source of well-being. The largest proportion of variance (13%) is explained by having a partner, friends, a relatively larger network and being more successful as compared to people in your surroundings. Work related means explain 8% of variance in well-being, and people's leisure activities 7%. Impersonal interactions for example on the street or in a bus only explain 1% of variance.

Even though the context-specific means are often multifunctional, this does not mean that they contribute equally to all first-order goals. In accordance with our expectations, friendships, for example, contribute primarily to the level of affection, whereas supervising others at work is important for one's status-level. Also, as expected, there is a stronger association between the means and specific goals than with overall subjective well-being (SPF-IL and traditional measures). Moreover, in accordance with our earlier findings, the positive and negative dimensions of status are distinctly different. Some means, such as knowing more people or supervising others at work, contribute primarily to the positive dimension whereas not having a partner and not having children contribute to negative status.[5]

We also see from Table VIII that the contribution of all these means to well-being measured in the traditional way is similar to their contribution to SPF-IL. Given this similarity, not much seems to be lost by using SPF-IL instead of the popular traditional measures, but much is gained by the fact that SPF-IL has specified and empirically corroborated dimensions.

*Group Differences*

For an additional validation of the measures, we tested some well confirmed (and common sense) expectations for each of the five first-order goals against these data. Such a face-value validation is quite elementary but it is still worth reporting because it gives

TABLE VIII

Main Study: Intercorrelations and multiple regression analyses of context-specific means and well-being (n = 1094)

| | Affection | Behavioral confirmation | Status | pos/neg | Comfort | Stimulation | Well-being (SPF-IL)[5] | β, adjusted R² | Well-being (traditional) | β, adjusted R² |
|---|---|---|---|---|---|---|---|---|---|---|
| **Personal relations** | | | | | | | | | | |
| Partner | 0.16 | 0.14 | 0.09 | 0.01/0.11 | 0.09 | 0.20 | 0.17 | 0.12 | 0.20 | 0.16 |
| Equal contribution partners[a] | 0.16 | 0.12 | 0.07 | 0.03/0.07 | 0.07 | 0.13 | 0.14 | – | 0.19 | – |
| Children | 0.09 | 0.14 | 0.10 | 0.00/0.14 | 0.08 | 0.13 | 0.13 | 0.06 | 0.13 | 0.05 |
| # of friends | 0.19 | 0.13 | 0.07 | 0.06/0.04 | 0.07 | 0.12 | 0.15 | 0.08 | 0.14 | 0.06 |
| Difficult friendships | −0.19 | −0.15 | −0.07 | 0.04/−0.14 | −0.09 | −0.15 | −0.16 | −0.12 | −0.16 | −0.12 |
| Know more people | 0.19 | 0.16 | 0.16 | 0.17/0.05 | 0.10 | 0.17 | 0.21 | 0.15 | 0.19 | 0.15 |
| Less successful | −0.15 | −0.15 | −0.18 | −0.08/−0.18 | −0.16 | −0.20 | −0.23 | −0.20 | −0.18 | −0.15 |
| | | | | | | | | $R^2 = 0.13$ | | $R^2 = 0.11$ |
| **Paid work** | | | | | | | | | | |
| Having paid work | 0.00 | 0.00 | 0.07 | 0.11/−0.02 | 0.10 | 0.08 | 0.08 | – | 0.11 | – |
| Troublesome relations[b] | −0.17 | −0.16 | −0.09 | 0.05/−0.17 | −0.17 | −0.12 | −0.19 | −0.18 | −0.17 | −0.16 |
| Tasks clear[b] | 0.13 | 0.11 | 0.00 | −0.04/0.04 | 0.15 | 0.15 | 0.15 | 0.14 | 0.15 | 0.14 |
| Supervise others[b] | 0.04 | 0.06 | 0.17 | 0.25/−0.02 | 0.10 | 0.20 | 0.16 | 0.16 | 0.18 | 0.19 |
| | | | | | | | | $R^2 = 0.08$ | | $R^2 = 0.08$ |

TABLE VIII

Continued

| | Affection | Behavioral confirmation | Status | | Comfort | Stimulation | Well-being (SPF-IL)[5] | β, adjusted R² | Well-being (traditional) | β, adjusted R² |
|---|---|---|---|---|---|---|---|---|---|---|
| **Leisure time** | | | | | | | | | | |
| Spending time with others | 0.14 | 0.12 | 0.09 | 0.08/0.04 | 0.16 | 0.14 | 0.18 | 0.13 | 0.15 | 0.11 |
| # of activities | 0.16 | 0.10 | 0.14 | 0.13/0.07 | 0.23 | 0.17 | 0.23 | 0.19 | 0.15 | 0.11 |
| # of vacations | 0.07 | 0.03 | 0.14 | 0.13/0.07 | 0.10 | 0.12 | 0.13 | 0.06 | 0.15 | 0.11 |
| | | | | | | | | $R^2 = 0.07$ | | $R^2 = 0.05$ |
| **Impersonal interactions** | | | | | | | | | | |
| Easy to make contact | 0.08 | 0.09 | 0.10 | 0.11/0.03 | −0.02 | 0.11 | 0.09 | 0.06 | 0.11 | 0.09 |
| Willingness to help others | 0.08 | 0.12 | 0.07 | 0.08/0.03 | −0.03 | 0.07 | 0.08 | 0.06 | 0.05 | 0.03 |
| People impressed by your things | 0.00 | 0.04 | 0.13 | 0.23/−0.05 | −0.02 | 0.03 | 0.05 | 0.04 | 0.04 | 0.04 |
| | | | | | | | | $R^2 = 0.01$ | | $R^2 = 0.01$ |

*Note:* [a] $n = 907$; [b] $n = 770$; for all correlations higher than 0.06, $p < 0.05$.

examples of how the first-order goals might be linked to lower order means for the realization of well-being. We used the transformed latent scale scores (minimum 0; mean 1). In Table IXa the mean level of *affection* is reported for people with or without a spouse. As expected, people with a partner have a significantly higher level of affection (almost 1/2 SD).

Also, as expected, the presence of children brings responsibilities and therefore opportunities to realize *behavioral confirmation*, as shown in Table IXb. In Table IXc significant differences between people with paid work vs. housewives are shown for *positive status*, confirming this expectation. For *negative status* we did not find this effect, challenging the content validity of this measure.

The realization of well-being deteriorates as a result of chronic conditions (Stewart et al., 1989; Verbrugge et al., 1994). Subjects with more chronic conditions are therefore expected to have a lower level of *comfort*. Table IXd reports indeed large differences in the level of comfort for people with one or more chronic conditions. In Table IXe, lower levels of *stimulation* are reported for people with fewer roles. The highest level of stimulation is found among people who have a partner, children and paid work.

Finally, in accordance with our expectations we found age and gender differences in the realization of the different first-order goals. The level of affection was lower for men than for women (0.98 vs. 1.02; $t$-test 3.4, $p < 0.001$), whereas men scored better with regard to positive status (1.06 vs. 0.95; $t$-test 5.1, $p < 0.001$), comfort (1.03 vs. 0.97, $t$-test 4.1, $p < 0.001$), and stimulation (1.01 vs. 0.99, $t$-test 2.5, $p < 0.05$). Age was associated with the level of behavioral confirmation ($r = 0.13$, $p < 0.001$), positive and negative status ($r = 0.06$, $p < 0.001$; $r = 0.15$, $p < 0.001$), and stimulation ($r = 0.10$, $p < 0.001$).

*Changes in Important Means*

The validation studies also involved a follow-up of the main study incorporating a 15-item version of the instrument.[6] It enabled us to look at changes in the level of affection, behavioral confirmation, status, comfort and stimulation[7] as a result of important changes in life-circumstances. For example, changes in one's work-situation, finding a partner, etc.

TABLE IX

Level of affection for people with or without a spouse, (b) Level of behavioral confirmation for people with or without children, (c) Level of status and work situation, (d) Level of comfort and chronic conditions, (e) Level of stimulation and number of roles (partner, parent and work role)

| | n | Mean | SD |
|---|---|---|---|
| *Panel a* | | | |
| Partner | | | |
| No | 187 | 0.93 | 0.23 |
| Yes | 907 | 1.01 | 0.19 |
| Total | 1094 | 1.00 | 0.20 |
| *Panel b* | | | |
| Children | | | |
| No | 319 | 0.95 | 0.25 |
| Yes | 775 | 1.02 | 0.25 |
| Total | 1094 | 1.00 | 0.25 |

| | | Status positive | | Status negative | |
|---|---|---|---|---|---|
| | n | Mean | SD | Mean | SD |
| *Panel c* | | | | | |
| (Early) retirement | 57 | 1.07 | 0.37 | 1.04 | 0.15 |
| Paid work | 748 | 1.03 | 0.36 | 1.00 | 0.22 |
| Student | 61 | 1.01 | 0.36 | 0.94 | 0.21 |

| | N | | | | |
|---|---|---|---|---|---|
| Disability | 44 | 0.97 | 0.41 | 0.97 | 0.29 |
| Unemployed | 43 | 0.95 | 0.40 | 0.99 | 0.27 |
| Housewife/man[a] | 127 | 0.82 | 0.37 | 1.04 | 0.21 |
| Total | 1080 | 1.00 | 0.37 | 1.00 | 0.22 |
| **Number of chronic conditions** | | | | | |
| *Panel d* | | | | | |
| None | 821 | 1.07 | 0.21 | | |
| 1 chronic condition | 202 | 0.82 | 0.27 | | |
| 2 chronic conditions | 71 | 0.68 | 0.31 | | |
| Total | 1094 | 1.00 | 0.26 | | |
| **Number of roles** | | | | | |
| *Panel e* | | | | | |
| None | 39 | 0.89 | 0.26 | | |
| 1 role | 169 | 0.95 | 0.21 | | |
| 2 roles | 396 | 0.99 | 0.19 | | |
| 3 roles | 490 | 1.04 | 0.16 | | |
| Total | 1094 | 1.00 | 0.19 | | |

*Panel a: Note:* The standardized latent scores have been transformed (minimum = 0; mean = 1). *t*-test 5.2, *p* < 0.001.

*Panel b: Note: t*-test 4.6, *p* < 0.001.

*Panel c: Note:* One-way ANOVA for work-situation, positive status ($F_{\text{work situation}}$ = 7.5; *p* < 0.01) and negative status ($F_{\text{work situation}}$ = 2.6; *p* < 0.05).[a] All but one of the housewives is male. However, the difference in the positive status-level can only partly be contributed to a gender effect (men 1.06 vs. women 0.95, *t*-value 5.1, *p* < 0.001).

*Panel d: Note:* One-way ANOVA for number of chronic conditions, comfort ($F_{\text{chronic conditions}}$ = 167.2; *p* < 0.001).

*Panel e: Note:* One-way ANOVA for number of roles, stimulation ($F_{\text{roles}}$ = 16.4; *p* < 0.001).

There was a significant increase in stimulation for respondents who experienced an increase in roles as a result of finding a job (from 5.2 to 5.7, $t$-test 2.0; $p < 0.05$). We also found an increase in comfort for people who expected their health to improve over time (from 4.2 to 4.8, $t$-test 3.0, $p < 0.01$). However, we did not find a further decrease for respondents who expected a deterioration, although their level of comfort is still significantly lower than for respondents who expected their health to remain stable (4.8 vs. 4.6, $t$-test 3.7, $p < 0.001$).

Eleven respondents experienced the ending of a relationship which resulted in a significant loss of affection in terms of feeling loved as measured by item 17 ($t$-value 2.4; $p < 0.05$). The new relationship of 10 respondents did not (immediately) result in an increase in affection; they already had a higher level of affection to begin with (7.1 vs. 6.0 at Time 1; $p < 0.05$). Having a new-born child ($n = 17$) did not raise the level of behavioral confirmation ($t$-value 0.7; $p =$ ns), although cross-sectionally we did find a higher level of behavioral confirmation for people with children (see Table IXb).

Fifteen respondents lost their job between Time 1 and Time 2, whereas others found a job ($n = 18$). There was a significant improvement in status for the newly employed. Those who found a job had an increase from 1.8 to 2.9 ($t$-value 3.6; $p < 0.01$). Although losing one's job did make people feel less influential (reducing the average score on item 39 from 2.0 to 1.5; $p < 0.05$), it did not lower the overall level of status significantly (2.5 vs. 2.3, $t$-value 0.4; $p =$ ns). This may be the result of the short time elapsed between the measurements and the presence of buffers. When people have other means for the realization of status, the effect of losing one's job is buffered, at least on the short run (Nieboer and Lindenberg, 2002).

## SHORTER VERSION OF THE INSTRUMENT

To reduce respondent's burden we developed two short-versions of the SPF-IL. A 15-item version with three positive items per goal and a 24-item version with three positive and three negative items for affection, behavioral confirmation and status and three items for

TABLE X

Comparing the normal-theory weighted least squares $\chi^2$ statistic, **SRMR**, **RMSEA** and **IFI** index for affection, behavioral confirmation, status, comfort and stimulation for the main study and validation study data

| | n | $\chi^2$ | df | p-Value | SRMR | RMSEA | 90% CI | IFI |
|---|---|---|---|---|---|---|---|---|
| **15-item version** | | | | | | | | |
| Main study data | 1094 | 168.24 | 80 | 0.00 | 0.032 | 0.043 | 0.025–0.038 | 0.97 |
| Asymptotic covariance | 1094 | 1600.91 | 80 | 0.00 | 0.050 | 0.030 | 0.024–0.037 | 0.98 |
| Validation study data | 725 | 181.17 | 80 | 0.00 | 0.040 | 0.042 | 0.034–0.050 | 0.96 |
| Asymptotic covariance | 725 | 165.09 | 80 | 0.00 | 0.083 | 0.038 | 0.030–0.047 | 0.97 |
| **24-item version** | | | | | | | | |
| Main study data | 1094 | 433.81 | 224 | 0.00 | 0.036 | 0.029 | 0.025–0.033 | 0.95 |
| Retest data | 163 | 317.76 | 224 | 0.00 | 0.051 | 0.050 | 0.037–0.063 | 0.93 |

*Note*: The 15-item version consists of items 7, 15, 17, 19, 27, 28, 37, 39, 41, 47, 49, 50, 52, 56, and 58. The 24-item version of 4, 7, 8, 9, 15, 18, 19, 20, 21, 22, 28, 29, 33, 37, 38, 39, 41, 42, 43, 44, 49, 51, 54, and 56.

comfort and for stimulation (see Table X). Both of the short versions of the measurement instrument were developed on the basis of the pilot data. In order to test the content validity of these scales a number of comparative analyses were conducted.

In Table X data are presented for the 15-item version comparing the results for the main study and the validation study. The results for the 15-item version are most favorable with a high IFI score, and low SRMR and RMSEA scores. The cross-validation index in which the estimated model based on the main study data is compared with the observed model as measured in the validation study was 19.88. The 15-item version of SPF-IL (see starred items in Appendix A) is thus a reasonable alternative for the full SPF-IL battery. We indicate this version by SPF-IL(s).

For the 24-item version we compared the model for the data from the main study and the test–retest sample of the validation study. This version of the instrument includes both positive and negative items. As a consequence, the IFI-index that compares the independent model with the estimated model is somewhat lower. The comparison of the very good overall fit of the 15-item version with the 24-item version shows that leaving out negatively worded items improves the fit, mainly because of the problem of measuring status.

## DISCUSSION

The aim of this study was to develop a conceptualization and measurement of the dimensions of subjective well-being. SPF theory specifies a hierarchy of human goals with the universal, first-order goals affection, behavioral confirmation and status for social well-being, and comfort and stimulation for physical well-being. An extensive qualitative study for the elaborate conceptualization of these goals helped us operationalize the dimensions of well-being and develop a quantitative measurement instrument. The resulting 58-item survey assesses the level of affection, behavioral confirmation, status, comfort and stimulation.

The results are very encouraging. With respect to the dimensional structure of the SPF-IL, a confirmatory factor analysis revealed that by and large the measurement model reflected the

underlying theory. The overall fit of the structural equation model is adequate with a low SRMR statistic (i.e., small differences between the estimated and observed model) and an acceptable RMSEA score. The only problem encountered was the poor construct validity of the negative status dimension. Without this status measure, the overall fit is very good. A revised version of the instrument may therefore have to be limited to the positive status dimension. As for the retest study 6 months after the main study, we found scales and individual levels of well-being to be reasonably stable and intercorrelations among each dimension over time to be acceptable.

In order to validate the instrument, we conducted a number of sub-studies. First of all, we compared the SPF-IL with popular traditional measures of well-being and we found that SPF-IL does as well as these measures, but has the advantage that it offers empirically corroborated dimensions. Second, we tested a number of expectations about the specific means for the realization of affection, behavioral confirmation, status, comfort and stimulation. As expected, personal relations are the most important contributor to well-being, followed by work, leisure time activities and impersonal interactions, in that order. Third, we tested and confirmed expectations on the difference in SPF-IL for specific groups (for example, people with and without work, people with many and few roles, people with and without partner, people with and without chronic conditions). Fourth, we conducted a separate validation study 6 months after the main study, in order to trace the effect of life events on the dimensions of well-being. Here too, the findings were supportive of the five dimensions and their measurement. Overall, the validation studies supported the instrument and the underlying dimensions of well-being.

Two short-form versions were developed to reduce respondent's burden. The 15-item version only included positive items, whereas the 24-item version consisted of negative items as well. The fact that the fit of the observed models was better (for the SPL-IL(s)) or not substantially lower than those for long-form measures (for the 24-item version) is encouraging. Tradeoffs between short- and long version measures in detecting changes in well-being over time do need further evaluation, but at this point, the 15-item version SPF-IL(s) can already be seen as a reasonable alternative to the full

58-item battery. It was used to measure changes over time in goals and means (in the validation study 6 months after the main study) and has since been used in a number of other studies.

There are some limitations associated with the study. First, we observed a relatively high rate of missing values for items in the status scale (18%). This suggests difficulties in answering status-related questions. Instructions should make clear that even when people feel awkward about items referring to status differences, it is essential they try to answer every question. As a result of cultural differences we expect this to be more of a problem in the relatively collectivist Dutch society than in more competitive individualistic societies such as the US (see for example Lucas et al., 2000). Second, the sample consisted of respondents between 18 and 65 years of age. An adjusted version of the instrument is currently used in the TRAILS panel study of adolescents, and the 15-item version is successfully applied in groups of frail elderly (Frieswijk, 2004, Schuurmans, 2004). Another limitation concerns the need for further research into issues associated with different forms of administration (for instance, self-completion or interviewer administration). Issues of social desirability – especially with regard to status – may be reduced by using self-completion methods (De Leeuw et al., 1996).

## APPENDIX A

The full 58-item SPF-IL battery (the 15-item version of SPF-IL(s) is indicated by starred items). All questions were originally stated in Dutch.

Meaning of the labels:

Questions 1 through 18 and 23 through 58:
    never.................................................... (1)
    sometimes............................................ (2)
    often .................................................... (3)
    always.................................................. (4)

Questions 19 through 22:
    never.................................................... (1)
    sometimes............................................ (2)

often .................................................... (3)
always.................................................. (4)
n.a. .................................................... (5)

I will ask you a number of questions about how you feel. These questions refer to the past 3 months. For your answer, will you please choose between NEVER, SOMETIMES, OFTEN or ALWAYS? If you HARDLY EVER have that feeling you can answer NEVER. If you ALMOST ALWAYS feel that way, answer ALWAYS. Use whichever answer is CLOSEST to the way you feel, NEVER, SOMETIMES, OFTEN or ALWAYS.
(Questions 1–18: Affection)

1.  Can you be completely at ease and be yourself with others? Have you NEVER, SOMETIMES, OFTEN or ALWAYS felt that way in the past 3 months?
2.  Would you say it is difficult for people to put themselves in your shoes?
3.  Do people really like you?
4.  Do people act like they care very little about how you feel?
5.  Do you feel that people can't stand you?
6.  Do people keep their guard up when they are around you?
7.* Do people pay attention to you?
8.  Do people show indifference to your thoughts?
9.  Do people sympathize with you?
10.  Do you miss being cuddled/hugged?
11.  Do people care about how you feel?
12.  Are your feelings towards others reciprocated?
13.  Do you feel your relations with others are out of balance?
14.  Do people cuddle/hug you?
15.* Do people help you if you have a problem?
16.  Do people expect something in return when they help you?
17.* Do you feel that people really love you?
18.  Do you feel that people don't care enough about you?

(Questions 19–30: Behavioral confirmation)

19.* There are situations in which we deal with groups of people, for example at home, at work or during our leisure time. Do others appreciate your role in the group?

Interviewer: n.a. DO NOT offer, only score if the respondent cannot give another answer

20.   Do people think your contribution is inadequate?
21.   Do people think that you function poorly in the group?
22.   Do others think that you achieve something together?
23.   Do people think that you strive for the wrong things?
24.   Do people think that you make the right choices?
25.   Do people have difficulty with the way you look at life?
26.   Do people believe that you are competent in whatever you do?
27.*  Do people find you reliable?
28.*  Do you feel useful to others?
29.   Do people get disappointed by what you achieve in the end?
30.   Do others feel that you are a burden to them?

(Questions 31–42: Status)

31.   Do people take you seriously?
32.   Do people look at you as an independent person?
33.   Do people look down on you?
34.   Do people think that you have strongly developed yourself?
35.   Could you have accomplished more according to others?
36.   Do people see you as being dependent?
37.*  Do people think you do better than others?
38.   Do people belittle your achievements?
39.*  Do people find you an influential person?
40.   Do people think you have no say in things?
41.*  Are you known for the things you have accomplished?
42.   Do people think poorly of you?

(Questions 43–50: Comfort)

43.   Now a few questions about how you feel physically. How often do you have pain? (i.e., physical pain)
44.   In the past few months have you felt:
     . . . ill?
45.   . . . fit?
46.   . . . physically poor?
47.*  . . . relaxed?
48.   . . . restless?

49.* … perfectly healthy?
50.* … physically comfortable?

(Questions 51–58: Stimulation)

51.  The next questions are also about the past 3 months. Do you find your life boring?
52.* Are your activities challenging to you?
53.  Do you feel too passive?
54.  Do you do things you find interesting?
55.  Is your life monotonous?
56.* Do you really enjoy your activities?
57.  Are you indifferent to your activities?
58.* How often are you fully concentrated when doing something?

## NOTES

[1] We speak of "instrumental goals" in the sense that they are both goals and resources. In a goal hierarchy, every level represents goals when seen from below and resources when seen from above. Thus, people are assumed to go after achieving these intermediary goals; yet, once achieved, the goals turn into resources for the realization of higher order goals.

[2] The decreasing marginal value is not the same for all five first-order goals. For example, status is not just a first-order goal for the achievement of social well-being but it can also be used as a means to realize all sorts of other higher-level or lower-level goals (such as stimulation, or a good job). Thus, the marginal value of status is likely to diminish less than the marginal value of an instrumental goal that cannot be used as a multipurpose instrument. We have empirical confirmation for this (see Nieboer and Lindenberg, 2002).

[3] Strictly speaking, the structure is a semi-lattice because lower order means (such as money) may help realize various higher-order goals.

[4] SPF-IL stands for "Social Production Function Instrument for the Level of well-being."

[5] In calculating the Cobb–Douglas function the positive and negative status dimension were included separately.

[6] See next section a description of this version.

[7] Here, we computed sum scores per 3-item subscale and checked the contribution of the separate items.

## REFERENCES

Andrews, F. M. and S. B. Withey: 1976, Social Indicators of Well-Being: American's Perceptions of Life Quality (Plenum Press, New York).

Badia, X., F. Gutierrez, I. Wiklund and J. Alonso: 1996, 'Validity and reliability of the Spanish version of the Psychological General Well-being Index', Quality of Life Research 5, pp. 101–108.

Bradburn, N. M: 1969, The Structure of Psychological Well-being (Aldine, Chicago).

Cantril, H.: 1965, The Pattern of Human Concern (Rutgers University Press, New Brunswick, NJ).

Cohen S. and T. A. Wills: 1985, 'Stress, social support, and the buffering hypothesis', Psychological Bulletin 98(2), pp. 310–357.

Csikszentmihalyi, M.: 1992, Flow: The Psychology of Happiness (Rider, London).

Csikszentmihalyi, M.: 1994, Flow: The Psychology of Optimal Experience (HarperCollins, New York).

De Leeuw, E. D., G. J. Mellenbergh and J. J. Hox: 1996, 'The influence of data collection method on structural models', Sociological Methods and Research 24, pp. 443–472.

Diener, E. and E. Suh: 1997, 'Measuring quality of life: Economic, social, and subjective indicators', Social Indicators Research 40, pp. 189–216.

Diener, E., E. M. Suh, R. E. Lucas and H. L. Smith: 1999, 'Subjective well-being: Three decades of progress', Psychological Bulletin 125, pp. 276–302.

Dow, G. K. and F. T. Juster: 1985, 'Goods, time and well-being: the joint dependency problem', in F. T. Juster and F. P. Stafford (eds.), Time, Goods and Well-being (Survey research center, University of Michigan, Ann Arbor), pp. 397–414.

Doyal, L. and I. Gough: 1991, A Theory of Human Need (MacMillan, Houndmills).

Du Toit, M. and S. Du Toit: 2001, Interactive LISREL: User's guide (Scientific Software International, Chicago).

Eid, M. and E. Diener: 2004, 'Global judgments of subjective well-being: Situational variability and long-term stability', Social Indicators Research, 65, pp. 245–277.

Fei, J., G. Ranis and F. Stewart: 1979, A Macro-Economic Framework for Basic Needs. Backgroundpaper for World bank (World Bank, Washington, DC).

Frieswijk, N.: 2004, Frail, But Happy. The importance of Self-management Ability and Social Comparison for the Subjective Well-being of Elderly Persons. Dissertation, Department of Psychology (University of Groningen, Groningen, NL).

Heady, B. and A. Wearing: 1989, 'Personality, life events and subjective well-being: Toward a dynamic equilibrium model', Journal of Personality and Social Psychology 57(4), pp. 731–739.

Hess, T.: 1994, Random individual samples within households or how inadequate methodologies lead to a waste of time and money. Paper presented at the 47th ESOMAR Marketing Research Congress.

Hoeben, C.: 2003, LETS' be a Community: Community in Local Exchange Trading Systems (Thela thesis, Amsterdam).

House, J. S., Umberson, D. and K. R. Landis: 1988, 'Structures and processes of social support', Annual Review of Sociology 14, pp. 293–318.

Hu, L. and P. M. Bentler: 1999, 'Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives', Structural Equation Modeling 6, pp. 1–55.

Jöreskog, K. and D. Sörbom: 1996, LISREL 8: User's Reference Guide (Scientific Software International, Chicago).

Kahneman, D.: 1999, 'Objective happiness', in D. Kahneman, E. Diener and N. Schwarz (eds.), Well-Being: Foundations of Hedonic Psychology (Russell Sage Foundation Press, New York), pp. 3–25.

Kassenberg, A.: 2002, Wat scholieren bindt: Sociale gemeenschap in scholen [Community in Schools] (Thela Thesis, Amsterdam).

Land, K. C.: 2001, 'Models and indicators', Social Forces 80, pp. 381–410.

Lindenberg, S.: 1986, 'The paradox of privatization in consumption', in A. Diekmann and P. Mitter (eds.), Paradoxical Effects of Social Behavior. Essays in Honor of Anatol Rapoport (Physica-Verlag, Heidelberg/Wien).

Lindenberg, S.: 1990, 'Homo socio-oeconomicus: the emergence of a general model of man in the social sciences', Journal of Institutional and Theoretical Economics 146, pp. 727–748.

Lindenberg, S.: 1996, 'Continuities in the theory of social production functions', in S. Lindenberg and H. B. G. Ganzeboom (eds.), Verklarende sociologie: opstellen voor Reinhard Wippler (Thesis Publishers, Amsterdam), pp. 169–184.

Lindenberg, S.: 2001, 'Social rationality versus rational egoism', in J. Turner (ed.), Handbook of Sociological Theory (Plenum, New York), pp. 635–668.

Lucas, R. E., E. Diener, and E. Suh: 1996, 'Discriminant validity of well-being measures', Journal of Personality and Social Psychology 71, pp. 616–628.

Lucas, R. E., E. Diener, A. Grob, E. S. Suh and L. Shao: 2000, 'Cross-cultural evidence for the fundamental features of extraversion', Journal of Personality and Social Psychology 79, pp. 452–468.

Maslow, A. H.: 1970, Motivation and Personality, 2nd edn. [1954], (Harper and Row, New York).

Max-Neef, M.: 1992, 'Development and human needs', in P. Ekins and M. Max-Neef (eds.), Real Life Economics: Understanding Wealth Creation (Routledge, London, New York ), pp. 197–213.

McCrae, R. R. and P. T. Costa: 1984, Emerging Lives, Enduring Dispositions: Personality in Adulthood (Little Brown, Boston).

Meertens, V.: 2004, Depressive Symptoms in the General Population: a Multifactorial Socal Approach. ICS-dissertation, University of Nijmegen (Thela thesis, Amsterdam).

Michalos, A. C.: 1985, 'Multiple discrepancies theory (MDT)', Social Indicators Research 16, pp. 347–413.

Nieboer, A. P.: 1997, Life-events and Well-being: A Prospective Study on Changes in Well-being of Elderly People due to a Serious Illness Event or Death of the Spouse (Thesis Publishers, Amsterdam).

Nieboer, A. and S. Lindenberg: 2002, 'Substitution, buffers and subjective well-being: a hierarchical approach', in E. Gullone and R. A. Cummins (eds.), The Universality of Subjective Well-being Indicators: A Multidisciplinary and Multinational Perspective. Social Indicators Book Series 16 (Kluwer Academic Publishers, Dordrecht), pp. 175–189.

Nieboer, A. P., R. Schulz, K. A. Matthews, M. F. Scheier, J. Ormel and S. Lindenberg: 1998, 'Spousal caregivers' activity restriction and depression: a model for changes over time', Social Science and Medicine 47, pp. 1361–1371.

Nieboer, A. P., S. M. Lindenberg and J. Ormel: 1999, 'Conjugal bereavement and well-being of elderly men and women: a preliminary study', Omega 38, pp. 113–141.

Noll, H.: 2002, 'Towards a European system of social indicators: theoretical framework and system architecture', Social Indicators Research 58, pp. 47–87.

Nunnally, J. C.: 1978, Psychometric Theory (McGraw-Hill, New York).

Offer, A.: 1996, In Pursuit of the Quality of Life (Oxford University Press, Oxford).

Ormel, J.: 2002, 'Social Production Function (SPF) Theory as a heuristic for understanding developmental trajectories and outcomes', in L. Pulkkinen and A. Caspi (eds.), Paths to Successful Development: Personality in the Life Course (Cambridge University Press, Cambridge), pp. 353–379.

Ormel, J., S. M. Lindenberg, N. Steverink and M. VonKorff: 1997, 'Quality of life and social production functions: A framework for understanding health effects', Social Science and Medicine 45, pp. 1051–1063.

Ormel, J., S. Lindenberg, N. Steverink and L. M. Verbrugge: 1999, 'Subjective well-being and social production functions', Social Indicators Research 46, pp. 61–90.

Schulz, W.: 1995, 'Multiple-discrepancies theory versus resource theory', Social Indicators Research 34, pp. 153–169.

Schuurmans, H.: 2004, Promoting Well-being in Frail Elderly People: Theory and Intervention (Thela Thesis, Amsterdam).

Steverink, N.: 2001, 'When and why frail elderly people give up independent living: the Netherlands as an example', Ageing and Society 21, pp. 45–69.

Stewart, A. L., S. Greenfield, R. D. Hays, K. Wells, W. H. Rogers, S. D. Berry, E. A. McGlynn and J. E. Ware Jr.: 1989, 'Functional status and well-being of patients with chronic conditions. Results from the Medical Outcomes Study', Journal of the American Medical Association 262, pp. 907–913.

Stone, A. A., S. S. Shiffman and M. W. De Vries: 1999, 'Ecological Momentary Assessment', in D. Kahneman, E. Diener and N. Schwarz (eds.), Well-Being: Foundations of Hedonic Psychology (Russell Sage Foundation Press, New York), pp. 26–39.

van Bruggen, A. C.: 2001, Individual Production of Social Well-being. An Exploratory Study. Dissertation at the ICS, University of Groningen (electronically available at www.ub.rug.nl/eldoc/dis/ppsw/a.c.van.bruggen/).

van Liere, C.: 1990, Lastige leerlingen. Een empirisch onderzoek naar sociale oorzaken van probleemgedrag op basisscholen [Problematic Pupils. An Empirical Study on Social Determinants of Problem Behavior in Schools] (Thesis Publishers, Amsterdam).

van der Lippe, T.: 1993, Arbeidsverdeling tussen mannen en vrouwen [Division of labor between men and women] Ph.D. thesis University of Utrecht (Thesis Publishers Amsterdam).

van der Lippe, T. and J. J. Siegers: 1994, 'Division of household and paid labour between partners: effects of relative wage rates and social norms', Kyklos 47, pp. 109–136.

Verbrugge, L. M., J. M. Reoma and A. L. Gruber-Baldini: 1994, 'Short-term dynamics of disability and well-being', Journal of health and social behavior 35, pp. 97–117.

Vitterso, J. and F. Nilsen: 2001, 'The conceptual and relational structure of subjective well-being, neuroticism, and extraversion: Once again, neuroticism is the important predictor of happiness', Social Indicators Research 57, pp. 89–118.

Watson, D., L. A. Clark and A. Tellegen: 1988, 'Development and validation of brief measures of positive and negative affect: the PANAS Scales', Journal of Personality and Social Psychology 54, pp. 1063–1070.

Wielers, R.: 1991, Selectie en allocatie op de arbeidsmarkt [Selection and allocation at the labor market] (Thesis Publishers, Amsterdam).

*Department of Health Policy and Management*          Anna Nieboer
*University of Rotterdam*
*P.O. Box 1738*
*3000 DR Rotterdam*
*The Netherlands*
*E-mail: nieboer@bmg.eur.nl*


*Department of Sociology and*                      Siegwart
*Interuniversity Center for Social Science*        Lindenberg
*University of Groningen*                  Alinda C. Van Bruggen
*The Netherlands*


*Department of Statistics and*                Anne Boomsma
*Measurement Theory*
*University of Groningen*
*The Netherlands*