Social Indicators Research (2005) 74: 327–348
DOI 10.1007/s11205-004-4644-7


JOAR VITTERSØ, ROBERT BISWAS-DIENER and ED DIENER

# THE DIVERGENT MEANINGS OF LIFE SATISFACTION: ITEM RESPONSE MODELING OF THE SATISFACTION WITH LIFE SCALE IN GREENLAND AND NORWAY

(Accepted 7 October 2004)

ABSTRACT. Cultural differences in response to the Satisfaction With Life Scale (SWLS) items is investigated. Data were fit to a mixed Rasch model in order to identify latent classes of participants in a combined sample of Norwegians ($N = 461$) and Greenlanders ($N = 180$). Initial analyses showed no mean difference in life satisfaction between the two subsamples. After transforming the ordinal raw scores into interval scales while simultaneously controlling for response bias, different results appeared. First, approximately 80% of the participants in the Greenlandic subsample fit a latent class with a large degree of random responding to the SWLS. Second, relative to the Norwegians, more Greenlanders were using extreme categories in responding to the SWLS. After statistically controlling for this tendency, Norwegians were in general more satisfied with their lives than Greenlanders. Third, Greenlanders who belonged to one specific latent class were more satisfied than their Norwegian counterparts. A salient feature of this class was the relative unwillingness of respondents to change the circumstances of their lives if they were given such an opportunity. The above results are a reminder of the care that must be used in analyzing survey data across cultures. The analytical strategy applied in the article offers an improved approach to handling such data.

KEY WORDS: Greenland, item response modeling, Norway, satisfaction with life, subjective well-being

## INTRODUCTION

In the hermeneutic tradition of *verstehen* established by Dilthey, Weber, and others, meaning was the most important component of social science. To understand human behaviour we need to understand the subjective orientation of the actors, the argument went. Although it is well recognized that social and personality psychology depends on subjects' interpretations of the meaning of, say, survey

items (e.g., Campbell and Fiske, 1959; Bem and Allen, 1974), a common critique raised by qualitatively oriented researchers is the following. Within self-report research in general, and in cross-cultural versions of it in particular, too much weight is placed on the assumption that questionnaire items are interpreted in the same way by all respondents. Some researchers even attribute the slow progress of academic psychology to this assumption (Jackson, 1984; Smedslund, 1988; Scheff, 1997; Gigerenzer, 1998; van de Vijver and Leung, 2000).

Alan P. Fiske argues that it is virtually impossible to explore the variation of meaning across cultures without participant observation fieldwork (Fiske, 1995). Although we agree with the importance of investigating the diversity of meanings both between and within cultures, we disagree that such analyses are impossible without participant observation. To the contrary, this article suggests that survey data, when analysed by means of a mixed Rasch model, will provide a systematic determination of relational meaning of the items in a multi-item self-report scale. In the case of life satisfaction, we believe that this procedure generalizes to the broader issue of understanding how important elements of people's everyday lives are being diversely evaluated by different groups of people. The method may thus contribute to a more sophisticated understanding of the configuration of features making up the concept of life satisfaction.

### *Satisfaction with Life*

Life satisfaction is one of the major components of subjective well-being, and both concepts, with different levels of specificity, refer to the summation of evaluations regarding a person's life as a whole. Ruut Veenhoven provides a review of the development of this concept (Veenhoven, 1996), in which he argues that a distinction between objective and subjective aspects of life satisfaction was developed in the 1960s and a separation between overall and domain satisfaction took place in the 1970s. Another development in the 1970s was the distinction made between "cognitive" and "affective" components of the satisfaction with life concept. In current literature on subjective well-being, the concept of life satisfaction is sometimes used interchangeably with the term happiness, as the correlation

between the two is usually found to be very high (in the area of .80 or above, e.g., Eid, 1997; Schyns, 1998; Schwarz and Strack, 1999; Vittersø, 2003).

Although the field of life satisfaction and happiness research has been fortunate to have a number of self-report studies on life satisfaction with large, representative surveys, it is now time for well-being researchers to move beyond the level of simple cross-sectional and correlational studies (Diener and Lucas, 2000). One step in this direction is to further investigate different meanings the concept of life satisfaction has both between and within cultures.

### *The Measurement of Meaning*

In social and personal psychology, meaning is often analysed by reducing sets of questions answered by samples of persons to a few generalizable quantities. For example, Osgood and his colleagues (Osgood et al., 1957) made one of the first attempts to measure meaning by asking subjects to rate words on 50 different bipolar scales, such as happy–sad or slow–fast. A quantitative analysis revealed that these adjectives could effectively be reduced to three factors (i.e., evaluative, potency, and activity). However, word meaning is extremely complex (e.g., Murphy, 2002), and Osgood's procedure could not be a complete representation of meaning, because it does not provide the information necessary to tell what the word means when it is used. The three factors are far too vague to represent the specific meanings of words. Osgood (1969) later restricted his proposal of a three-dimensional meaning space to account only for affective meaning.

One approach to meaning is based on response patterns given by respondents on a restricted universe of questions or propositions. This is provided when a particular Item Response Model (IRM) is applied to a set of items with a fixed response scale (for example that of a Likert format). The connection between IRT and the analysis of meaning is covered by the logic that a sample of items is representative for the concept they intend to indicate. A latent response pattern to these items may then be identified. If identified, such a response pattern can be taken as evidence that the items have a shared meaning in the population, or at least, the hypothesis of shared meaning cannot be rejected.

Shared meaning of items is a basic assumption in all measurement theory. Any true quantification in social sciences presupposes an underlying scale that is identical for the participants included in the study. If such a scale is not established, comparisons made between any two persons or groups of persons may be misleading for at least two reasons. First, the meaning of the concept being compared may not be the same for the two persons or groups of persons. Second, the scale units upon which the measurement rest may not be the same for the two persons or groups being compared. For example, a comparison of score level obtained in different groups is only meaningful when the scale on which the scores are expressed has the same origin (zero-point) and the same metric (scale units) across the different groups (Poortinga, 1989). Furthermore, scores may not be equivalent if items relate differently to the underlying content for the two individuals or groups. Item response theory is one way of transforming survey data with unknown zero-points and scale units into a comparison scale with equal zero-points and similar metric (e.g., Bond and Fox, 2001).

When combining item response theory and latent class analysis, the researcher is not restricted to the assumption that all items are understood unambiguously in the entire population. This means that the assumption of homogeneity in a population can be relaxed. For example, by reanalysing data from the Five-Factor Model of personality, Rost et al. (1997) discovered two latent classes in a 12-item extraversion scale. A subsample of extraverts easily agreed with the positive emotion facet of the extraversion construct (i.e., with items such as ''I am a cheerful person'', ''I am a joyful and humorous person'' and ''I am a happy-go-lucky optimist''), whereas another subsample of extraverts tended to disagree with these items. The meaning of positive emotions is thus differently related to extraversion for different groups of extraverts. Somewhat similarly, in a recent investigation of life events and subjective well-being, Lucas et al. (2003) found that the impact of marriage and widowhood was differently related to life satisfaction for different people. Although this analysis was not based on latent classes, the results clearly demonstrate how misleading a conclusion may be if the relations between relevant variables are treated homogenously for an entire population. Indeed, the idea that marriage will influence life satisfaction differently for people is commonsensical, and the lack of focus on these

matters may be attributed to methodological rather than theoretical restrictions. Thus, freeing up the classical test theoretical constraint that the entire population must be homogenous in terms of their reactions to either questionnaire items or life events may be an important step in further developing the concept of life satisfaction. In the current work, the mixed Rasch model will be utilized toward this goal.

### *The Mixed Rasch Model*

The Mixed Rasch Model belongs to the family of Item Response Models (e.g., Embretson and Reise, 2000; Bond and Fox, 2001). Briefly, it extends the polytomous Rasch model's assumption that item responses depend only on the trait of a person and the difficulty of the item, to also include a mixture of latent classes (a more technical description of the Mixed Rasch Model is provided by Rost and Langeheine, 1997). The Polytomous Rasch Model (which sometimes is referred to as the partial credit model) defines the probability of passing each threshold of an ordinal scale. The probability estimates, or threshold parameters, define the intersection points of each of two adjacent values on the ordinal scale, where one of the response categories has highest probability. These probabilities are often visually expressed by Item Characteristic Curves. The model only holds if the threshold parameters are ordered, that is, if the threshold difficulties increase with increasing scores on the ordinal scale. The Mixed Rasch Model extends the Polytomous Rasch Model by allowing differences in the threshold parameters in two or more subpopulations of individuals. In contrast to ordinary latent class models (e.g., Lazarsfeld, 1950), the model allows differences among respondents within a class by means of the ability (or trait) parameter.

The trait parameter is corrected for the effect of a possible response set on the original scale. The trait parameter is scaled at the interval level. The Mixed Rasch Model also estimates the location of each item on the latent trait under scrutiny, that is, the item location parameter (which is called item difficulty parameter for dichotomously scored variables). Finally, the Mixed Rasch Model provides goodness-of-fit estimates for the entire model, for each of the items, and for each participant.

Compared with other approaches applied in comparisons of well-being across nations, the Mixed Rasch Model has some impressive advantages. First, the dispute over of psychological measurement addressed, for example, by Michell (1990, 1999) is partly overcome due to the transformation of ordinal raw scores into interval scales with an additive structure. Second, the assumption that a sample from a particular culture is homogenous, that is, that all its members assign the same meaning to the items of an inventory is not made. Third, since the mixed Rasch model provides goodness-of-fit estimates for the degree to which the scoring pattern of each participant deviates from the model's prediction, aberrant respondents may be excluded from further analyses. By excluding aberrant individuals, the influence of measurement errors on the final results is reduced, and the probability of class membership is increased.

### *Main Purposes of the Study*

The current study will address the issue of cross-cultural comparisons of self-reported well-being by discovering whether members of different cultures ascribe the same meaning to a set of survey items. To this end, the Satisfaction With Life Scale (SWLS, Pavot and Diener, 1993) will be fitted to a Mixed Rasch Model. Examining group differences in response patterns for the SLWS enables us to analyse the possible heterogeneous meaning of items both within and between cultures. Furthermore, we will analyse cultural differences in response biases, and after controlling them, compare life satisfaction across cultures on an interval scale rather than the original raw score scale, which is only assessed at the ordinal level of measurement.

## METHOD

### *Participants*

We analysed a mixed sample of Norwegians ($N = 461$) and Greenlanders ($N = 180$). Data for the Norwegian subsample were collected in 2000 and are approximately representative of the adult population in Northern Norway, with a mean age of 47.8 years, ranging from 19 to 88, with a standard deviation of 15.8 years. The sample had 45.2% females and is further described in Vittersø and Nilsen (2002). The

Greenland data were collected in 2002 as a convenience sample among the Inughuit of the Northwestern Greenland. It has a mean age of 39.3 years (Range = 18–76, SD = 12.7) with 58.3% females. It is further described in Biswas-Diener et al. (2003).

### *Measures*

For both subsamples, the SWLS (Pavot and Diener, 1993) was administered along with several other personality and emotional inventories in self-reported questionnaires. The SWLS contains the following items: (1) *In most ways my life is close to the ideal*; (2) *The conditions of my life are excellent*; (3) *I am satisfied with my life*; (4) *So far I have gotten the things I want in life*; and (5) *If I could live my life over, I would change almost nothing*. The items are answered on a 7-point Likert-scale according to the labels Strongly disagree (1); Disagree (2); Slightly disagree (3); Neither agree nor disagree (4); Slightly agree (5); Agree (6); and Strongly agree (7). The Cronbach alphas were 0.87 for the Norwegian and 0.58 for the Greenlandic subsamples, respectively. After recoding the scale into a five point scale (see below), the metric and labels should read: 0 = Strongly disagree; 1 = disagree, 2 = Neither agree nor disagree; 3 = Agree; 4 = Strongly agree.

### *Procedure*

Using the Expectation-Maximization (EM) procedure in SPSS 10.0 for Windows, 37 missing data values were replaced with their imputed estimates. To reduce the cells of the multidimensional contingency table, we rescaled the seven-point SWLS ordinal scale into five response categories by collapsing the two lowest of the original categories into one category, and similarly collapsing the two highest categories (see for example Rost et al., 1997; Eid and Diener, 2001 for analogous strategies). The SWLS was then fitted to a Mixed Rasch Model using the WINMIRA software (von Davier, 2000). The number of latent classes were increased until an acceptable goodness-of-fit was reached. When the *p*-values of both the Pearson chi-square and the Cressie Read reached the level of 0.05 or higher, we considered the model fit to have been achieved (von Davier, 2000). However, a widespread problem with chi-square

parameters for survey data fitted to item response models, is that the number of cells becomes substantially larger than the number of observed answering patterns, bootstrapping is recommended as a remedy (Langeheine et al., 1996). For our analysis, the bootstrapped *p*-values were used to evaluate goodness-of-fit. For the current data, appropriate fit coefficients were reached with four latent classes. However, among the 641 participants in the original data file, 73 showed severe deviations in their person-fit parameter (i.e., digressions larger than two standard deviation from the expected fit). These persons were dropped, yielding a revised sample of 568 persons. There were no national biases in the proportion of aberrant persons in the two countries; 18 (11%) and 55 (11%) participants were removed from the Greenlandic and Norwegian subsamples. The WINMIRA program also estimates the probability of class membership. Based on these probabilities, a participant's class membership can, with some degree of certainty, be identified and be grouped according to his or her class membership, to be described later. Four classes were generated and the mean class membership probabilities in our sample were 92%; 76%; 84% and 71% for classes 1–4 respectively.

## RESULTS

### *National Differences*

Table I shows mean scores for the rescaled 5-point SWLS, and the mean of the rescaled person parameters estimated by the mixed Rasch model for the Greenlandic and the Norwegian subsamples. Presented in the Table are also the person parameter means for each of the four groups. Clearly, splitting the sample into four latent classes as well as using person parameters rather than raw scores make a significant difference. Whereas no group difference in life satisfaction was found for the raw scores, the Norwegian subsample shows significantly higher life satisfaction compared with the Greenlanders when the numbers were transformed into an interval scale ($M^{\mathrm{Nor}} = 1.17$, $\mathrm{SD}^{\mathrm{Nor}} = 1.95$ vs. $M^{\mathrm{Green}} = 0.64$, $\mathrm{SD}^{\mathrm{Green}} = 1.20$, $p < 0.001$). Similarly, Norwegians belonging to the first latent class scored higher on the transformed scale than did Greenlanders in this class (M($M_{\mathrm{C1}}^{\mathrm{Nor}} = 0.44$, $\mathrm{SD}_{\mathrm{C1}}^{\mathrm{Nor}} = 0.46$ vs. $M_{\mathrm{C1}}^{\mathrm{Green}} = 0.32$,

$SD_{C1}^{Green} = 0.38$, $p < 0.05$). For members of the third class, on the other hand, Greenlanders showed higher life satisfaction as measured both with the ordinal and the interval scale. (For the transformed scale: $M_{C3}^{Nor} = -0.15$, $SD_{C3}^{Nor} = 1.63$ vs. $M_{C3}^{Green} = 1.26$, $SD\text{-}L_{C1}^{Nor} = 1.29$, $p < 0.01$). All the four latent classes had different group means on the interval scale ($F$ [3, 564] = 218.9, $p < 0.001$).

*Latent class Differences*

The mean item location across the four classes were 0.17; −0.37; −1.87; −0.47; and 2.54 for items SWLS1, SWLS2, SWLS3, SWLS4, and SWLS5, respectively. The third item has the lowest item location (−1.87), which means that this item is the easiest of the five to agree with. Technically, the item location represents the mean of all threshold parameters as defined above, and further explained below. The analogy of item location to the original scoring procedure of the SWLS would be to say that this item has the highest mean. Item 5 of the SWLS is the hardest to agree with (analogous to the lowest mean on the ordinal scoring). Figure 1 shows the locations for each item in the four latent classes.

*Class 1*. With regard to item difficulty, all items are relatively close to zero in Class 1, which indicates a poor discrimination among the five SWLS items. Such a result could be due to random responses. Moreover, the threshold parameters presented in Table II indicate a tendency for members of this class to apply the extreme response options of the scale (i.e., strongly disagree or strongly agree). This is so because the relation between the threshold parameter and the observed responses is not ordered (for an illustration of ordered response category, see Figure 3). A response set is ordered when an increasing level of the threshold parameter indicates that every response category is representative for an interval of the individual parameter dimension (von Davier, 2000). For Class 1, however, the responses are highly unordered because the thresholds for the Categories 2–4 are lower than that for Category 1. Such a situation is also referred to as ''quasi-categorization'' because the number of response categories in the scale does not reflect the quantitative structure of the data. The middle categories operate in such a way that the probability of someone in Class 1 to apply one of them is never greater than the

TABLE I

Ordinal (raw score) means and interval (person parameter) means for the Satisfaction With Life Scale in Northern Greenland and Northern Norway

| | Raw scores | | | | Parameter scores | | | |
|---|---|---|---|---|---|---|---|---|
| | W. S. (*n*) | Green. (*n*) | Norw. (*n*) | *t* | W. S. (*n*)[a] | Green. (*n*) | Norw. (*n*) | *t* |
| All classes | 2.75 (568) | 2.76 (162) | 2.64 (406) | 1.50 | 1.01 (568) | 0.62 (162) | 1.17 (406) | $-4.1^{***}$ |
| Latent Class 1 | | 2.74 (131) | 2.92 (91) | −1.48 | 0.37 (222) | 0.32 (131) | 0.44 (91) | $-1.99^{*}$ |
| Latent Class 2 | | 3.08 (18) | 3.00 (192) | 0.65 | 2.56 (210) | 2.78 (18) | 2.54 (192) | 0.89 |
| Latent Class 3 | | 2.58 (12) | 1.95 (112) | $3.10^{**}$ | −0.02 (124) | 1.26 (12) | −0.15 (112) | $3.5^{**}$ |
| Latent Class 4 | | 0.6 (1) | 1.3 (11) | – | −3.48 (12) | −7.07 (1) | −3.15 (11) | – |

*Notes:* W.S: = Whole sample; Green = Greenland, Norw. = Norway.
$^{*}p < 0.05$.
$^{**}p < 0.01$.
$^{***}p < 0.001$.
[a]All latent class means are significantly different from each other (at $p < 0.01$ with the Bonferonni ad hoc test).

probability of applying either a 0 (*Strongly disagree*) or a 5 (*Strongly agree*). In other words, for Class 1 the SWLS does not discriminate well between individuals high and low on life satisfaction. This is further illustrated by the Warm's modified likelihood estimate (WLE) curve in Figure 2a. This curve indicates that even if the SWLS raw scores increase, the latent trait parameter is rather stable around the zero-point. Furthermore, the raw score distribution in this class is negatively skewed, with most scores in the range of 15–18 on the 0–20 point scale (i.e., indicating high proportion of extreme scores at the upper end of the scale). Recall that in the interest of reducing the number of cells in the contingency table, the original 5–35 SWLS range is transformed to a 0–20 scale. Combining the information from Figures 1 and 2 with the threshold parameters in Table II, one has the impression that the members of Class 1 have randomly reported responses to the SWLS items, except for their tendency to apply the strongly agree option.

*Class 2*. Members of Class 2 avoid both the low and the high extreme scores. This can be seen because the threshold of passing to the second category (which represents threshold 1) is generally very

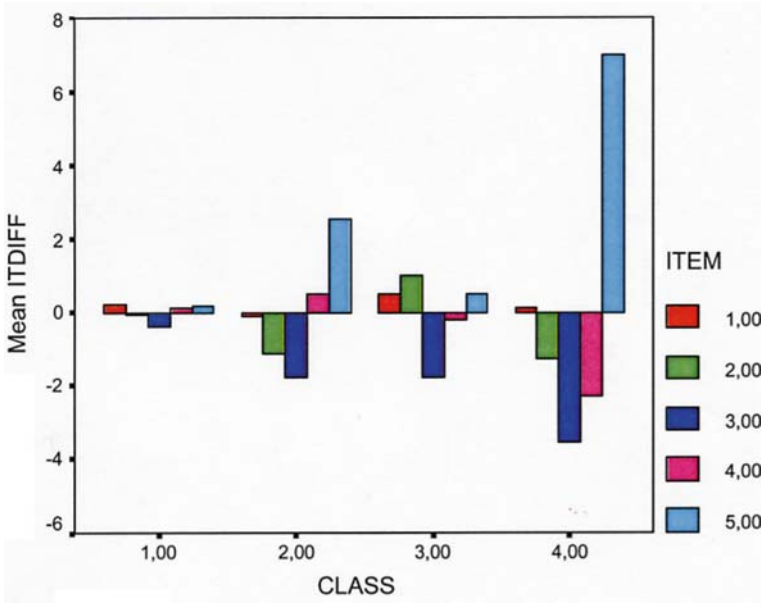


*Figure 1*. Item difficulties for the five SWLS items in four latent classes.

TABLE II

Threshold parameters and item location for the five SWLS items across the four latent classes

| Class | Item | Item location | Threshold | | | |
|---|---|---|---|---|---|---|
| | | | 1 | 2 | 3 | 4 |
| 1 | 1* | 0.21 | 1.54 | −0.30 | −0.98 | −0.98 |
| 1 | 2* | −0.08 | 0.38 | 0.80 | −0.55 | −0.96 |
| 1 | 3* | −0.41 | 0.31 | −0.24 | −0.50 | −1.21 |
| 1 | 4* | 0.12 | 0.75 | 0.10 | −0.16 | −0.21 |
| 1 | 5* | 0.17 | 1.73 | −0.81 | 0.28 | −0.53 |
| 2 | 1 | −0.12 | −7.12 | 0.58 | 1.03 | 5.02 |
| 2 | 2* | −1.12 | −8.12 | −0.05 | −0.24 | 3.93 |
| 2 | 3* | −1.79 | −8.79 | 3.21 | −3.69 | 2.11 |
| 2 | 4* | 0.50 | −0.68 | −0.84 | 0.77 | 2.74 |
| 2 | 5 | 2.53 | 0.60 | 1.57 | 1.57 | 6.40 |
| 3 | 1 | 0.49 | −3.45 | −1.48 | 2.46 | 4.40 |
| 3 | 2 | 1.01 | −2.92 | −2.06 | 1.99 | 7.04 |
| 3 | 3 | −1.77 | −8.10 | −2.43 | 0.35 | 3.11 |
| 3 | 4 | −0.22 | −3.04 | −1.41 | 1.61 | 1.96 |
| 3 | 5 | 0.48 | −1.02 | −0.19 | 1.37 | 1.81 |
| 4 | 1 | 0.13 | −6.87 | 0.13 | 0.137 | 7.13 |
| 4 | 2 | −1.28 | −8.28 | −1.28 | −1.28 | 5.72 |
| 4 | 3 | −3.53 | −10.53 | −3.53 | −3.53 | 3.47 |
| 4 | 4 | −2.29 | −9.29 | −2.29 | −2.29 | 4.71 |
| 4 | 5 | 6.97 | 1.63 | 6.46 | 8.33 | 11.45 |

*Notes*: * = Unordered response categories.

easy, whereas the threshold of passing to the last category is very difficult. This pattern indicates that only people with extremely low levels of life satisfaction would apply the *Strongly disagree* option. Here is an example. For Item 3, only people with a latent life satisfaction score lower than 8.8 standard deviations below the mean would check this option. Moreover, Figure 1 reveals that this class easily agrees with statements such as ''The conditions of my life are excellent'' (Item 2) and ''I am satisfied with my life'' (Item 3) on the one hand, but tends to disagree with the statement ''If I could live my life over, I would change almost nothing'' (Item 5). So, even if the members of Class 2 agree that their life conditions are excellent, they would consider changing their life if they were able to do so. The threshold 4 on Item 5 is very high for Class 2. This means that it is very rare to give high scores on this item.

A problem with Class 2 is the unordered thresholds. However, the quasi-categorization is not as severe as it was for Class 1. First, it is only items 2, 3, and 4 which are unordered. Second, for items 2 and 4, the thresholds are only mildly reversed (−0.05 vs. −0.24 for thresholds 2 and 3 on item 2; −0.68 vs. −0.84 for thresholds 1 and 2 on item 4). The main problem for Class 2 is related to item 3. As indicated by thresholds 2 and 3 (Table II), for no level of life satisfaction on this item is the probability for answering 2 (*Neither agree nor disagree*) higher than answering either 1 (*Disagree*) or 3 (*Agree*). The trouble with the midpoint category (labelled ''Undecided'', ''Neutral'', ''Neither nor'', etc.) is a puzzle applying to many data sets in the survey literature (e.g., Andrich et al., 1997). Consulting Figure 2b, it can be observed that the WLE curve for Class 2 discriminates better than it did for Class 1. The latent trait midpoint is also somewhat more consistent with the raw score means, although the latter is still somewhat skewed.

*Class 3*. The members of this class seem more reluctant to change their life conditions than Class 2. Compared with the latter, Class 3 also has lower latent life satisfaction ($M_{C2} = 2.56$, $SD_{C2} = 1.32$ vs. $M_{C3} = -0.02$, $SD_{C3} = 1.66$, $p < 0.001$). The differences between the two classes with regard to the willingness to change their life circumstances is also interesting in terms of how they respond to the second SWLS item (''The conditions of my life are excellent''). Class 2 tends to agree with this statement whereas Class 3 tends to disagree with it, although they nevertheless agree about not changing anything, given that they could live their lives over (cf. Figure 1). Class 3 has ordered categories because participants with the highest latent life satisfaction trait are also giving the highest responses on all the five SWLS-items (Table II). Similarly, the lowest ratings given on the five items are from persons with the lowest latent life satisfaction. For example, the first threshold parameter for SWLS1 is −3.45 and the fourth threshold parameter is 4.40. The second and third thresholds are located between these values, at −1.48 and 2.46, respectively. The first threshold on item three is low, which means that it is very easy to pass this threshold, which again indicates that people avoid the *Strongly disagree* option for this item unless they are extremely unhappy. The ordered responses are further illustrated in Figure 3, where the Item Characteristic Curves for Item 3 in Class 3 is depicted. The horizontal

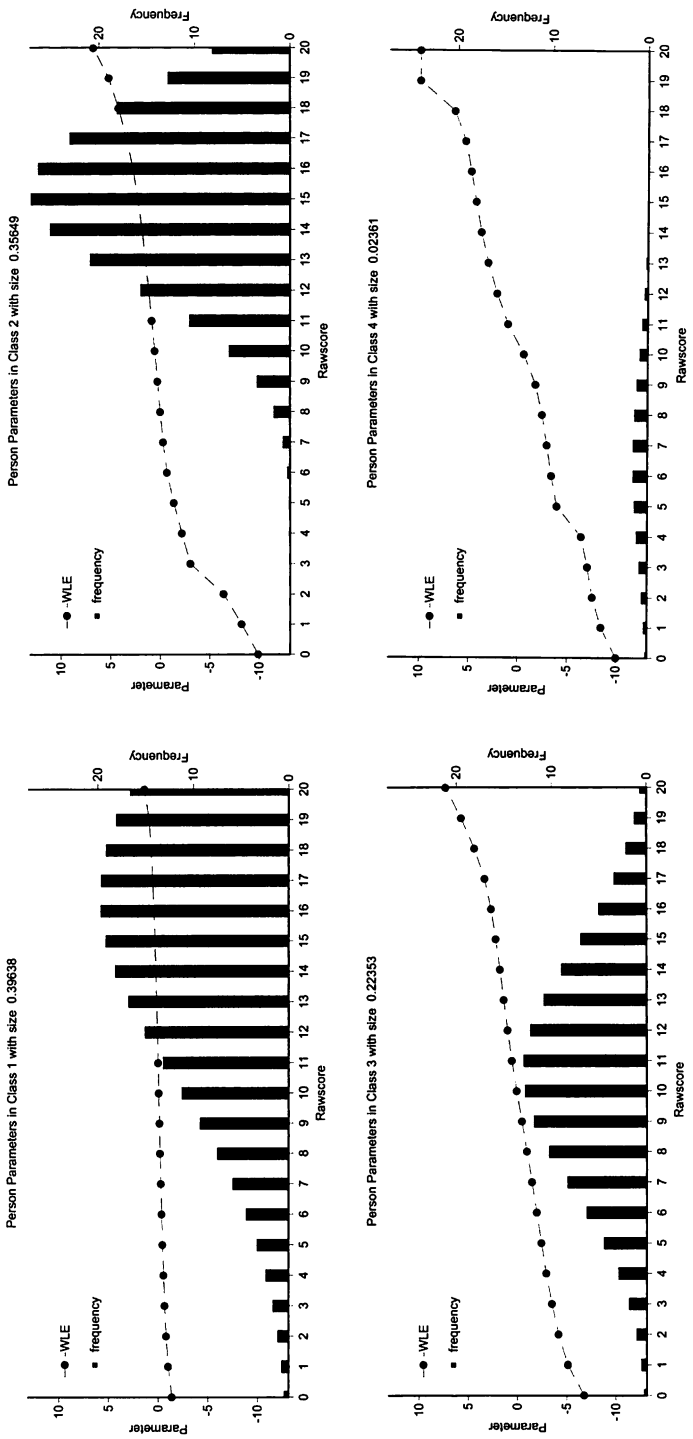


*Figure 2.* Raw score and person parameter graphs for the four latent classes, shown simultaneously with the person parameter plot (the Warm's modified likelihood estimate-WLE) at the left side of the $y$-axis (dots), and the expected frequency at the right side of the $y$-axis (bars).

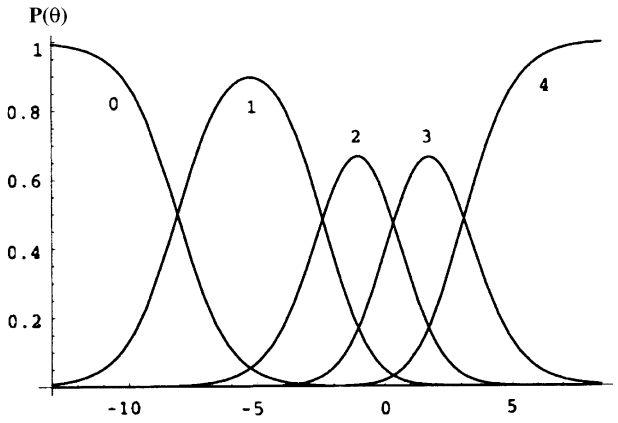


*Figure 3*. Item Characteristic Curves for SWLS3 in class-3. The trace lines give the probability (*y*-axis) of observing each response categories (the peaks numbered 0–4) at increasing level of the latent life satisfaction (in *z*-score units along the *x*-axis).

axis represents the latent trait of life satisfaction in z-score units. The vertical axis represents the probability of observing each response category (numbered 0–4). Thus, for Item 3 most respondents with life satisfaction values greater than 3 (in z-score units) *Strongly agree* (i.e., response category 4) that they are satisfied with their lives. Respondents with life satisfaction values between 0 and 3 *Agree* with this item, whereas respondents with life satisfaction values between −2 and 0 are neutral. Respondents with latent scores between −8 and −2 *Disagree* and respondents scoring lower than −8 *Strongly disagree* with this item. In other words, only those with respectively extremely low or extremely high levels of life satisfaction use the response options *Strongly disagree* or *Strongly agree* in this class. Finally, Figure 2c shows that for this class the SWLS discriminates nicely between individuals with varying degrees of life satisfaction, and that the responses are fairly normally distributed.

*Class 4*. When it comes to Class 4, it looks somewhat similar to the second class, but for these members, high scores on item five were extremely rare. Relative to their life satisfaction, members of Class 4 disagree strongly with this item. However, this class is very small and unstable. For example, most of the aberrant participants removed from the analyses belonged to this class. Due to the low number of

TABLE III

Latent class membership and nationality

| Class | Nation | | |
|---|---|---|---|
| | Greenland | Norway | Total |
| 1 | 131 (80.9%) | 91 (22.4%) | 222 (39.1%) |
| 2 | 18 (11.1%) | 192 (47.3%) | 210 (37.0%) |
| 3 | 12 (7.4%) | 112 (27.6%) | 124 (21.8%) |
| 4 | 1 (0.6%) | 11 (2.7%) | 12 (2.1%) |

*Notes:* Pearson chi-square (3) = 166.2, $p < 0.001$.

respondents in this class, we find it difficult to make a meaningful interpretation of the responses.

Before closing, a few comments regarding the distribution of nationality across the latent classes may be warranted. Table III reveals that most of the Greenlanders (81.0%) belong to the first class, whereas only 22.4% of the Norwegians were ascribed Class 1 membership. Almost half of the Norwegians and about 11% Greenlanders belong to the second class. Seven per cent of the Greenlanders and 22% of the Norwegians were ascribed membership in Class 3. Clearly, class inclusion is highly dependent on cultural background (Pearson chi-square with 3 degrees freedom was 166.2, $p < 0.001$)

## DISCUSSION

In the current article, a mixed Rasch model was applied in order to understand whether members of different cultures ascribe the same meaning to the five items of the SWLS. The results indicate that survey based comparisons with ordinal raw score data may be misleading, particularly if the issue of response bias is ignored. For instance, when comparing the Norwegian raw scores on the SWLS with those from Greenlanders, no mean difference appeared. However, after transforming the ordinal scores into an interval scale (the trait estimates), the Norwegians were in general more satisfied than Greenlanders. This happens because the trait estimates are corrected for response bias, which turned out to be culturally dependent in our data. Broadly speaking, this means that Norwegians are more sat-

isfied with their lives than Greenlanders. However, there exists a subgroup of Greenlanders with a higher SWLS mean compared with their Norwegians counterparts. It is a challenge for further research to understand why the particular response pattern revealed in this subpopulation is more beneficial to Greenlanders than it is to Norwegians.

Members of the latent Class 1 frequently use the extreme scores of the original SWLS scale, whereas extreme scores are avoided in the other classes. In transforming ordinal scores to interval scores, persons with extreme scores are given a lower score relative to persons who avoid the extreme options of the original scale. Hence, relative to people with an extreme response set, participants who avoid extreme ratings are given a lower (higher) interval score in the rare occasions in which the extreme score options actually are chosen. Because the Greenlanders are over-represented in the first class (with an extreme score bias), they will on average get lower interval scores after transformation, and as a consequence have lower mean scores compared to the Norwegians (Table I).

Based on raw score estimates, Class 1 seems to have a relatively high level of life satisfaction. However, this result is partly caused by the tendency to an extreme score bias, and after controlling for such a bias, the class mean was much reduced. Relative to the Greenlanders, the Norwegians members of this class had a higher level of life satisfaction. Class 2 and Class 3 show the most interesting response pattern in our sample, with the former somewhat more satisfied than the latter. There is no difference in the latent level of life satisfaction between Greenlanders and Norwegians in Class 2. Another feature of this class is that they are more prone to change the circumstances of their lives, even if they are satisfied with their current situation.

A small proportion of the Greenlandic subsample belongs to Class 3. These individuals are generally more satisfied with their lives compared with their Norwegian counterparts. A salient characteristic of the responses given to the SWLS by the members of Class 3 is that they do not consider their life conditions as being excellent, but they would nevertheless not change much in their lives, even if they could do so. It would be an interesting task for future research to better understand the somewhat paradoxical result that people who in general agree that their life conditions are excellent would like to change them if they could (i.e., Class 2),

whereas people who are less satisfied with their life circumstances seem more hesitant to change them (i.e., Class 3).

Another result worth pondering is the high proportion of Greenlanders in Class 1. The responses in this class are highly unordered and biased towards the extremes. The items are all given a similar meaning, perhaps because of random responses, and they do not discriminate well between differences in the latent trait satisfaction with life. Surveys from non-western samples may thus be ''noisier'' than those generated by a sample which is more familiar with the procedures of self-report data collection. There is of course nothing new in the claim of cultural variations in response style. It is the possibility offered by the Mixed Rasch Model, however, to identify, describe, and eventually isolate persons with different response sets. In our study about 20% of the hunting and gathering population of Northern Greenland behaved reliably in the context of a survey investigation. Whether this is good or bad may be debated, but the ability to better isolate the meaningful responses from noisy ones is promising for a more sophisticated understanding of cultural differences and similarities in life satisfaction. For example, reanalyses of existing data by means of mixed Rasch models may provide valuable new insights in the field of subjective-well-being research.

### *Limitations*

The Mixed Rasch Model estimates the personal parameters corrected for different latent classes if the groups are relatively large. At least for Class 4, this was not the case, and any interpretation of this class should be made cautiously. The requirement for large samples also places some restriction on cross-cultural applications of survey research. There are simply some cultural groups that cannot meet the requirement of large samples. For example, the entire population of Inughuit is, according to Statistics Greenland (Grønlands Statistik, 2000) only 817, among which only about 500 are above 19 years old. Our sample of Inughuit thus includes over one-third of the entire adult population.

Another problem is related to the exclusion of aberrant persons. Aberrant participants are either those who have not filled in the questionnaire carefully, or persons who only marginally belong to the population. In both cases, the removal of such persons will increase

the accuracy of the analysis. However, the practice has at least two important drawbacks. First, the criteria for exclusion are not clear. In our analyses we decided to include only persons revealing scoring patterns that depart no more than two standard deviations from the model's prediction. However, other cut-off criteria may be used, which includes more or fewer participants. Second, by excluding participants one also dichotomises respondents into either normal or aberrant individuals, without any further specification of different types of aberrant individuals (see van den Wittenboer et al., 1997, for a discussion of these issues).

In closing, we will point to some of the promising new perspectives for life satisfaction research that have been revealed by the present analyses. First, our results indicate that the meaning of the SWLS items diverge both within and between cultures. Thus, the common practice of restricting the entire population to one single response format needs to be relaxed. The good news is that this restriction is not made in the Mixed Rasch Model. Second, Item Response Theory offers a unique strategy of dealing with response biases both within and across cultures. The importance of handling response biases in survey research can hardly be overstated (e.g., Watkins and Cheung, 1995; van de Vijver and Leung, 1997; Cheung and Rensvold, 2000; Cheung and Chan, 2002; Byrne and Watkins, 2003), and the use of IRT can aid substantially in correcting for such biases. Finally, as additional latent class analyses are introduced to the field of survey research, a more differentiated pattern of relation between relevant concepts will likely emerge. For instance, the impact of life events, health, income, and so forth, on life satisfaction may vary for different classes of individuals. Similarly, the relations between phenomena such as self-esteem, optimism, self-efficacy, affective style, personal strivings, coping strategies, adaptation and other concepts in the social and personal psychology may vary between distinct groups.

The history of science has repeatedly shown that more precise estimates of the relation between concepts are a prerequisite for significant progress within scientific disciplines (Mayr, 1982; Holland et al., 1989; Thagard, 1992, 1999). Latent class modelling provides a promising foundation for more precise analyses of survey data. The method should be particularly helpful in future comparisons of self-reported quality-of-life indicators across different cultures.

## ACKNOWLEDGEMENTS

We like to thank Dr. Dariush Arai for providing the Item Characteristic Curves in Figure 3 and for commenting on an earlier version of this article.

## REFERENCES

Andrich, D., J. A. L. de Jong and B.E. Sheridan: 1997, 'Diagnostic opportunities with the Rash model for ordered response categories', in J. Rost and R. Langeheine (eds.), Application of Latent Trait and Latent Class Models in the Social Sciences (Waxman, Münster), pp. 59–72.

Bem, D. J. and A. Allen: 1974, 'On predicting some of the people some of the time: the search for cross-situational consistence in behavior', Psychological Review (81), pp. 506–20.

Biswas-Diener, R., J. Vittersø and E. Diener: 2003, 'Most people are pretty happy, but there is cultural variation: The Inughuit, the Amish, and the Maasai', Manuscript submitted for publication.

Bond, T. G. and C. M. Fox: 2001, Applying the Rash model: Fundamental measurement in the human sciences (Lawrence Erlbaum, Mathwah, NJ).

Byrne, B. M. and D. Watkins: 2003, 'The issue of measurement invariance revisited', Journal of Cross-Cultural Psychology 34, pp. 155–175.

Campbell, D. T. and D. W. Fiske: 1959, 'Convergent and discriminant validation by the multitrait-multimethod matrix', Psychological Bulletin 56, pp. 81–105.

Cheung, G. W. and R. B. Rensvold: 2000, 'Assessing extreme and acquiescence response set in cross-cultural research using structural equation modeling', Journal of Cross-Cultural Psychology 21, pp. 187–212.

Cheung, M. W. L. and W. Chan: 2002, 'Reducing uniform response bias with ipsative measurement in multiple-group confirmatory factor analysis', Structural Equation Modeling 9, pp. 55–77.

Diener, E. and R. E. Lucas: 2000, 'Explaining differences in societal levels of happiness: Relative standards, need fulfilment, culture, and evaluation theory', Journal of Happiness Studies 1, pp. 41–78.

Eid, M.: 1997, 'Happiness and satisfaction: An application of a latent state-trate model for ordinal variables', in J. Rost and R. Langeheine (eds.), Application of Latent Trait and Latent Class Models in the Social Sciences (Waxman, Münster), pp. 148–154.

Eid, M. and E. Diener: 2001, 'Norms for Experiencing Emotions in Different Cultures: Inter- and Intranational Differences', Journal of Personality and Social Psychology 81, pp. 869–885.

Embretson, S. E. and S. P. Reise: 2000, Item Response Theory for Psychologists (Lawrence Erlbaum Ass, London).

Fiske, A. P.: 1995, 'The cultural dimensions of psychological research: Method effects imply cultural mediation', in P. Shrout and S. Fiske (eds.), Personality Research, Methods, and Theory. A Festschrift Honoring Donald W. Fiske (Lawrence Erlbaum, Hillsdale, N.J), pp. 271–294.

Gigerenzer, G.: 1998, 'Surrogates for theories', Theory & Psychology 8, pp. 195–204.
Grønlands Statistik: 2000, Grønland 2000. Statistisk Årbog (Grønlands hjemmestyre, Nuuk, Grønland).
Hayduk, L.: 2003, 'Pearl's D-separation: One more step into causal thinking', Structural Equation Modeling 10, pp. 289–311.
Holland, J. H., K. J. Holloak, R. E. Nisbett and P. R. Thagard: 1989, Induction. Processes of Inference, Learning, and Discovery (The MIT Press, London).
Jackson, M. L.: 1984, Self-esteem and meaning. A life-historical investigation (State University of New York Press, Albany).
Landauer, T. K. and S. T. Dumais: 1997, 'A solution to Plato's problem. The Latent Semantic Analysis theory of acquisition, induction, and representation of knowledge', Psychological Review 104, pp. 211–240.
Langeheine, R., J. Pannekoek and F. van der Pol: 1996, 'Bootstrapping goodness-of-fit measures in categorical data analysis', Sociological Methods and Research 24, pp. 492–516.
Lazarsfeld, P. F.: 1950, 'The logical and mathematical foundation of latent structure analysis', in S. A. Stouffer, L. Guttman, E. A. Suchman, P.F. Lazardsfeld, S. A. Star, and J. A. Clausen (eds.), Studies in Social Psychology in World War II (Princeton University Press, Princeton, N.J), pp. 362–412.
Lucas, R. E., A. E. Clark, Y. Georgellis and E. Diener: 2003, 'Re-examining adaptation and the setpoint model of happiness: Reactions to changes in marital status', Journal of Personality and Social Psychology 84, pp. 527–539.
Mayr, E.: 1982, The Growth of Biological Thought. Diversity, Evolution, and Inheritance (Harvard University Press, Cambridge).
Michell, J.: 1990, An Introduction to the Logic of Psychological Measurement (Lawrence Erlbaum, Hillsdale, NJ).
Michell, J.: 1999, Measurement in Psychology: Critical History of a Methodological concept (Cambridge University Press, New York).
Murphy, G. L.: 2002, The Big Book of Concepts (The MIT Press, London).
Osgood, C. E.: 1969, 'On the whys and wherefores of E, P, and A', Journal of Personality and Social Psychology 12, pp. 194–199.
Osgood, C.E., G. J. Suci and P. H. Tannenbaum: 1957, The Measurement of Meaning (University of Illinois Press, Urbana, Ill,).
Pavot, W. and E. Diener: 1993, 'Review of the Satisfaction With Life Scale', Psychological Assessment 5, pp. 164–172.
Poortinga, Y. H.: 1989, 'Equivalence of cross-cultural data: An overview of basic issues', Internartional Journal of Psychology 2, pp. 737–756.
Rost, J., C. Carstensen and M. von Davier: 1997, 'Applying the mixed rash model to personality questionnaires', in J. Rost and R. Langeheine (eds.), Application of Latent Trait and Latent Class Models in the Social Sciences (Waxman, Münster), pp. 324–332.
Rost, J. and R. Langeheine: 1997, 'A guide through latent structure models for categorical data', in J. Rost and R. Langeheine (eds.), Application of Latent Trait and Latent Class Models in the Social Sciences (Waxman, Münster), pp. 13–37.
Scheff, T. J.: 1997, Emotions, The social Bond, and Human Reality. Part/Whole Analysis (Cambridge University Press, Cambridge, IK).
Schwarz, N. and F. Strack: 1999, 'Reports of subjective well-being: Judgmental processes and their methodological implications', in D. Kahneman, E. Diener, and

N. Schwarz (eds.), Well-Being: The Foundations of Hedonic Psychology (Russell Sage Foundation, New York), pp. 61–84.
Schyns, P.: 1998, 'Cross-national differences in happiness: Economic and cultural factors explored', Social Indicators Research 43, pp. 3–26.
Shipley, B.: 2000, Cause and Correlation in biology. A Users Guide to Path Analysis, Structural Equations and Causal Inference (Cambridge University Press, Cambridge, UK).
Smedslund, J.: 1988, 'What is measured by a psychological measure?', Scandinavian Journal of Psychology 29, pp. 148–151.
Thagard, P.: 1992, Conceptual Revolutions (Princeton University Press, Princeton, N. J.).
Thagard, P.: 1999, How Scientists Explain Disease (Princeton University Press, Princeton, N. J.).
van de Vijver, F. and K. Leung: 1997, Methods and Data Analyisis for Cross-Cultural Data (Sage, London).
van de Vijver, F. J. R. and K. Leung: 2000, 'Methodological issues in psychological research on culture', Journal of Cross-Cultural Psychology 31, pp. 33–51.
van den Wittenboer, G., J. J. Hox and E. D. de Leeuw: 1997, 'Aberrant response pattern in elderly respondents: Latent class analysis of respondent scalability', in J. Rost and R. Langeheine (eds.), Application of Latent Trait and Latent Class Models in the Social Sciences (Waxman, Münster), pp. 155–162.
Veenhoven, R.: 1996, 'Developments in satisfaction research', Social Indicators Research 37, pp. 1–46.
Vittersø, J.: 2003, 'Qualities of life: Separating goal-orientation and plan-orientation in a multimodal model of well-being', Manuscript submitted for publication.
Vittersø, J. and F. Nilsen: 2002, 'The conceptual and relational structure of subjective well-being, neuroticism, and extraversion: Neuroticism is again found to be the important predictor of happiness', Social Indicators Research 57, pp. 89–118.
von Davier, M.: 2000, WINMIRA 32 User Manual (IPN – Institute for Science Education, Kiel, Germany).
Watkins, D. and S. Cheung: 1995, 'Culture, gender, and response bias', Journal of Cross-Cultural Psychology 26, pp. 490–504.

*University of Tromso* Joar Vittersø
*Department of Psychology*
*Norway*
*E-mail: joarv@psyk.uit.no*

*University of Oregon* Robert Biswas-Diener
*USA*

*University of Illinois and* Ed Diener
*the Gallup Organization*
*USA*