Advances in Psychometric Methods for Measuring the Subjective Aspects of Well-Being

Metodi statistici multivariati per la psicometria: scala di qualità di vita e di benessere

Bruno D. Zumbo¹, Alex C. Michalos²

¹ Measurement & Research Methodology Program, and Department of Statistics, University of British Columbia, Vancouver, Canada
   e-mail: bruno.zumbo@ubc.ca

² Institute for Social Research and Evaluation, University of Northern British Columbia, Prince George, Canada

Riassunto: Nel presente lavoro viene introdotta una nuova scala di qualità di vita e di benessere utile negli studi di psicometria. Si mostra, in particolare, come la scala consenta di misurare il benessere, e se la sua performance è equivalente nei maschi e nelle femmine. Inoltre, si studia la relazione della nuova scala con la scala soggettiva di benessere di Deiner, e la sua relazione con vari aspetti della vita come la soddisfazione per le proprie amicizie, la salute e la sicurezza finanziaria.

Keywords: psychometrics, measurement validity, quality of life

1. Introduction

The direct monitoring of key social-psychological states (attitudes, expectations, feelings, aspirations, and values) in the population is necessary for an understanding of social change and the quality-of-life. This approach to social research got a big boost in the 1960s with the space race and the social accountability movement in the U.S., and about the same time program evaluation blossomed. A direct monitoring of this sort is on-going in Northern British Columbia as part of the activities of the Institute for Social Research & Evaluation (ISRE) at the University of Northern British Columbia. Specifically, the ISRE seeks to measure psychological satisfaction, happiness, and life fulfillment by using survey research instruments that ascertain the subjective reality in which people live. The result may aptly be termed life satisfaction, subjective wellbeing, or happiness indicators. The items on this new index of perceived wellbeing introduced in this paper were chosen based on previous quality-of-life surveys (e.g., Michalos, 1996; Michalos & Zumbo, 1999; Michalos & Zumbo, 2000) collected by ISRE.

The current paper has two interconnected goals. First, we wish to describe our new index, the Subjective Wellbeing Index (SWB) that we have developed for the purposes of (a) tracking the perceived and self-reported quality of life, and (b) studying its correlates and determinants with such “domains” as health status, satisfaction with family, job, etc. Second, in the context of our new index, we will demonstrate some new psychometric and statistical methods for investigating the construct validity of self-report quality of life measures. We have used SWB in several studies, e.g., most recently in Michalos, Thommasen, Read, Anderson and Zumbo, 2005. However, this is
the first time we have examined its psychometric properties in some detail with more sophisticated procedures.

2. Methods

2.1. Subjective Wellbeing Index (SWB)

The SWB was developed for the purposes of tracking the perceived quality of life and studying its correlates and determinants with such life domains as health status, satisfaction with family, job, etc. This index consists of four statements, each on a 7-point rating scale (see Figure 1). The first three statements ask respondents to rate their level of satisfaction based on their overall standard of living, whole lives, and overall quality of life, in which ‘1’ indicates respondents are very dissatisfied and ‘7’ indicates respondents are very satisfied. The last statement asks respondents to rate their overall level of happiness, in which ‘1’ indicates respondents are very unhappy and ‘7’ indicates respondents are ‘very happy’. These responses are then combined into a simple composite index (the sum of questions 1-4) ranging from 4 to 28, in which a high score on the index reflects a greater self-perceived wellbeing.

Descriptive wellbeing categories for the SWB total score were derived by dividing respondents’ total score into the top and bottom 10th percentiles, as well as the middle 50% of the distribution of the current sample of survey respondents. These categories correspond, somewhat, with the item anchors: (a) 4-15 Very dissatisfied with one’s life, (b) 16-19 Somewhat dissatisfied, (c) 20-24 Middling, (d) 25-26 Somewhat satisfied, and (e) 27-28 Very satisfied with one’s life.

2.2. Participants – survey respondents

Individuals who were included in this study were obtained from the Health and Health Care Survey carried out by the ISRE in the fall of 2000. Participants were randomly selected from the Dominion telephone list. The sample comprised of 388 community-dwelling adults living in the city of Prince George, British Columbia. The city of Prince George is the main urban settlement in the Northern Interior region of British Columbia, Canada. Of the 383 respondents who identified their gender (5 did not respond to the gender question), 160 were females. Their average age of female respondents was 43 years (SD=13.34), with a range from 18 to 79 years. The average age of the 223 male respondents was 48 years (SD=12.30), with a range from 19 to 85 years.

In terms of the education level of the respondents, thirty-four percent (n=132) of the respondents had had either some university courses or had completed their university degree, 44% (n=169) had either courses toward or had completed a diploma or certificate from trade, technical, or business; and 13% (n=52) had only completed secondary school. The remaining 9% (n=33) of respondents had either attended or completed elementary school. Overall, as expected given a mail-out survey, this is a

– 376 –
Figure 1: Subjective Wellbeing Index (SWB)

Please indicate how satisfied you are with each of the following aspects of your life. If the item does not apply to you, just skip over it. Indicate whether you are (1) very dissatisfied, (2) somewhat dissatisfied, (3) a little dissatisfied, (4) about evenly balanced, (5) a little satisfied, (6) somewhat satisfied, (7) very satisfied with each of the following:

<table>
<thead>
<tr>
<th>How satisfied are you with (circle your response)</th>
<th>Very Dissat.</th>
<th>Even Balance</th>
<th>Very Satis.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Your overall standard of living ..................</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Your life as a whole ..................................</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Your overall quality of life ........................</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4. Considering your life as a whole, would you describe it as (1) very unhappy, (2) somewhat unhappy, (3) a little unhappy, (4) an even mix of happiness and unhappiness, (5) a little happy, (6) somewhat happy, or (7) very happy?

Very Unhappy Even Happy
1 2 3 4 5 6 7

slightly more educated sample compared to the 2001 Statistics Canada census data of Prince George. A cross-tabulation of gender with education level revealed nearly twice as many males than females had completed a diploma or certificate, whereas nearly twice as many females than males had completed some university courses. The average household income was $69,880 CND (females $63,891 and males $73,973).

3. Psychometric item analyses

In contemporary psychometric theory (e.g., Zumbo, 2007a; Zumbo, Gelin & Hubley, 2002) a central question in developing an index is how the items come together to reflect the phenomenon of interest. This question is addressed through item analysis. The item analysis technology of tests and measures was developed to help answer the following practical questions faced by researchers: (1) given that the items are combined to create one scale score, do they measure just one latent variable? (2) how much of the observed variation is true variation and therefore how precisely do the items measure? (3) how does this precision change across the levels of the continuous latent variable? and (4) does the index measure differently for sub-populations of interest? The results will be presented around these four psychometric questions.

Because the SWB items are ordinal in nature (i.e., in our case a seven-point response scale, and hence not continuous), where appropriate a polychoric correlation matrix was used as input for the analyses. As Zumbo (2007a) notes, using a polychoric matrix is an underlying variable approach to modeling ordinal data (as opposed to an item response theory approach). We also make use of model the item responses directly via item
response theory, which given the sample size and the number of items is of the nonparametric variety.

For a polychoric correlation matrix, an underlying continuum for the polytomous scores (i.e., the 7-point rating scale) is assumed and the observed responses are considered manifestations of respondents exceeding a certain number of latent thresholds on that underlying continuum. Conceptually, the idea is to estimate the latent thresholds and model the observed cross-classification of response categories via the underlying latent continuous variables. Formally, for item $j$ with response categories $c = 0, 1, 2, \ldots, C-1$, define the latent variable $y^*$ such that

$$y_j = c \quad \text{if} \quad \tau_c < y_j^* < \tau_{c+1},$$

where $\tau_c, \tau_{c+1}$ are the latent thresholds on the underlying latent continuum, which are typically spaced at non-equal intervals and satisfy the constraint $-\infty = \tau_0 < \tau_1 < \cdots < \tau_{C-1} < \tau_C = \infty$. It is worth mentioning at this point that the latent distribution does not necessarily have to be normally distributed, although it is commonly assumed so due to its well understood nature and beneficial mathematical properties, and that one should be willing to believe that this model with an underlying latent dimension is actually realistic for the data at hand.

### 3.1. Do the items measure one latent variable?

Given that the item responses were on a rating scale, a principal component analysis of the polychoric correlation matrix was computed using PRELIS 2.30 (Jöreskog & Sörbom, 2000). The first component accounted for 76.5% of the generalized multivariate variance. We therefore fit a unidimensional factor analysis model (using MINRES estimation) and the loadings on this first factor were all above 0.73, indicating that the index is strongly unidimensional.

### 3.2. How much of the observed variation is true variation and therefore how precisely do the items measure?

Continuing with the theme of the ordinal nature of the items, we used the recently developed ordinal coefficients theta (Zumbo, Gadermann, & Zeisser, 2007) to quantify the measurement error. Ordinal coefficient alpha was 0.897 – not surprisingly given the unidimensionality, ordinal theta was 0.898. Therefore, at least 89.7% of the observed variation is attributable to individual differences among survey respondents on the latent variable, the variable being measured – i.e., the total score on the scale.

### 3.3. How does this precision of measurement change across the levels of the continuous latent variable?

Coefficient alpha (and theta) is a marginal reliability estimate; akin to an average over the continuum of the index, therefore, we investigated how the reliability of the index scores varies along the continuum of variation. That is, it is possible (actually, quite likely) that the index measures with more precision at different levels of the total score.
We used nonparametric item response theory (Ramsay, 1991, 1997) embodied in the software TestGraf (Ramsay, 2001) to estimate the conditional reliability function.

It should be noted that item response theory (unlike the polychoric correlation methods described above) directly models the item responses as a function of a latent (unobserved) total scale score. We chose to use an item response model that is flexible with respect to modeling the probability of choosing or observing a response. Rather than assume that a logistic response function is appropriate, which is assumed with the conventional parametric item response methods, we adopt an approach that estimates response curves directly, thereby escaping the restrictions imposed by parametric models such as the Rasch and other polytomous item response models (Wu & Zumbo, 2007). In contrast to logistic response models, we make no a priori assumptions about the underlying distribution of responses and their relation to the latent variable they purport to measure. This approach is extremely flexible and is consistent with an analytic strategy that emphasizes exploratory data analysis. Nonparametric kernel-smoothing techniques were developed by Ramsay (1991, 1997) for use in analyzing such measurement data.

Figure 2 depicts the conditional reliability plot for the SWB. One can see in Figure 2 that the scale reliability varies from approximately 0.70 to 0.92, and that the index, therefore, measures less precisely for respondents who report being very dissatisfied to somewhat dissatisfied with their life and performs best at the high end of the total score continuum. Please note that there are few individuals who scored 28 on the scale; so the curve at that end of the continuum should not be considered accurate.

**Figure 2: Conditional reliability plot of the Subjective Wellbeing Index**

3.4. Does the index measure differently for sub-populations of interest?

Based on recommendations from Zumbo, Sireci, and Hambleton (2003), an exploratory approach based on the polychoric correlation matrix was conducted for male and female respondents separately. For each gender we found essential unidimensionality
with all loadings greater than .70. The factor congruence coefficient was .997 indicating that the factors are similar in males and females. The one minor difference that was apparent is that item #3, the quality of life item, loaded less for females (males = .92, females=.79 factor loadings). Given the component and factor analysis results, we conclude, that we are essentially measuring the same variable for both genders.

Recently Zumbo (2003) showed that scale-level studies of measurement invariance, of the form reported above, must be supplemented with item-level studies of invariance (i.e., differential item functioning; DIF) to detect measurement problems. DIF, in its essence, involves a form of matching of respondents from different groups (see, Zumbo, 2007b for review of DIF). We examined the DIF across genders for the items of the SWB using non-parametric item response modeling. The results indicate that the items performed the same for males and females.

Due to space limitations we do not provide the DIF plots for the four SWB items, but the interested reader is encouraged to write to us for copies. Two points are noteworthy. First, using criteria established by Zumbo and Witarsa (2004) (i.e., for our sample sizes, the composite DIF measure should not exceed 0.0415) the items perform the same for male and female respondents. Therefore, there is no gender bias in item responding. Second, the items perform well. That is, the item response function for each item begins in the far lower left-hand corner of each plot and the item response function increases to the top right-hand corner of each plot. The slope of that function is related to the item discrimination for that item.

### 4. Correlation with life domains

Having demonstrated in Section 3 of this paper that the SWB has strong psychometric properties we turn our attention to how the index correlates to satisfaction with various life domains. In addition, however, we first investigated the correlation of the SWB with a widely used measure of subjective wellbeing, Diener’s Satisfaction with Life Scale, SWLS (Diener et al., 1985). Diener’s SWLS scale is a five-item self-report scale designed to measure global cognitive judgments of one’s life. The SWLS items are, like the SWB items, on a 7-point response scale. We therefore estimated the correlation between the SWB and the SWLS using latent variable structural equation modeling based on the joint polychoric correlation matrix of the items from both scales. The resultant correlation, which is, by statistical design given our method, corrected for the attenuation due to measurement error in each of the SWLS and the SWB is 0.873. Although the correlation between the SWB and the SWLS is a substantial correlation, this indicates that approximately 24% of the variation in the SWB is distinct from Deiner’s scale. Thus, this lends support to the unique contribution of the SWB to the literature.

Finally, we turned to the correlation of the SWB with the reported satisfaction with various life domains. As traditionally done in the research literature the satisfaction with the various life domains is reported on a single question using a 7-point satisfaction response scale. Therefore, we computed the correlation of the SWB score with each life domain from the joint distribution of the SWB items and each life domain item – based on the joint polychoric correlation matrix. The resulting correlation is between a latent variable and the underlying variable (as described above in Section 3) from the domain item response. In essence each correlation involves a polychoric
correlation and the latent variable for SWB. This type of correlation has not, to our knowledge, been discussed in the literature and was developed for our purposes. Please note that the standard practice of correlating the observed total scale score with the 7-point domain satisfaction responses (treating these item responses as continuous) can greatly attenuate the resultant correlation – in our data this was as much as a difference of 0.15 in the correlation magnitude.

Table 1 lists the correlations of the SWB and the satisfaction with ten commonly studied life domains. The domains are listed in decreasing magnitude of the correlations. All the correlations are statistically significant. The largest correlations are with satisfaction with financial security, and with the respondents own satisfaction with their self-esteem. The smallest correlations (although they are still statistically significant) are with satisfaction with personal safety and satisfaction with family relations.

Table 1: Correlation of the SWB and satisfaction with various life domains

<table>
<thead>
<tr>
<th>Variable</th>
<th>Correlation with the SWB</th>
</tr>
</thead>
<tbody>
<tr>
<td>...your financial security</td>
<td>0.827</td>
</tr>
<tr>
<td>...your self-esteem</td>
<td>0.824</td>
</tr>
<tr>
<td>...your living partner</td>
<td>0.709</td>
</tr>
<tr>
<td>...your health</td>
<td>0.686</td>
</tr>
<tr>
<td>...your job</td>
<td>0.663</td>
</tr>
<tr>
<td>...your recreation activities</td>
<td>0.662</td>
</tr>
<tr>
<td>...your friendships</td>
<td>0.625</td>
</tr>
<tr>
<td>...your religion or spiritual fulfilment</td>
<td>0.529</td>
</tr>
<tr>
<td>...your family relations, generally</td>
<td>0.462</td>
</tr>
<tr>
<td>...your personal safety</td>
<td>0.379</td>
</tr>
</tbody>
</table>

5. Concluding remarks

Our purpose was to showcase current psychometric developments in the context of describing our new Subjective Wellbeing Index. A psychometric theme throughout the paper is the use of modern psychometric methods that are appropriate for the ordinal (rating scale) response formats found in most quality of life and survey studies. These new methods involve non-linear techniques incorporating polychoric correlations as well as direct non-linear modeling of item responses.

For an index to be appropriate for use in social policy and social research, the inferences made from it must be fair and accurate. What that translates to in terms of measurement is tolerable error and equal functioning in key groups (such as gender). The index we introduced today shows these features. On-going research in B.C. is investigating the usefulness of this index in tracking communities over time.
References


Zumbo B.D. (2007b) Three generations of differential item functioning (DIF) analyses: Considering where it has been, where it is now, and where it is going, *Language Assessment Quarterly*, 4, 223-233.


However, instruments for measuring new conceptualizations of well-being associated with these models have not been proposed. The hedonic approaches to conceptualizing well-being focus on pleasure and happiness (Ryan & Deci, 2001). Some instruments were originally presented without psychometric data; in these cases, data reported are from the Mental Measurements Yearbook (Farmer, 2005; Lonborg, 2007) or from the earliest publication that reported psychometric data for an instrument (e.g., the Wellness Inventory; Palombi, 1992). Health and psychological well-being are the most commonly studied outcomes of the retirement transition and bridge employment. Although the subjective nature of self-report questionnaires is not without criticism, their convenience and utility in tapping into psychological health have been validated. For clinical diagnoses of psychiatric disorders, interviews and questionnaires are always used for evaluating the psychological state of the patient. In general, well-validated and reliable questionnaires are highly useful if the targeted outcomes are related to different aspects of psychological well-being. Carsten Wrosch, Michael F. Scheier, in Advances in Motivation Science, 2020. 5 Conclusions. The measure of well-being shows psychometric promise for community surveys. Psychological well-being is not exactly the opposite end of the continuum to psychological distress, but more debate is needed about whether and when, research participants need to be asked questions about both. Background. Measurement of psychological well-being utilises various instruments without any having gained dominance as a “gold standard.” The scoring system used is based on the method developed by the Clinical Research Unit for Anxiety and Depression at the University of New South Wales. In this method, five points are given for “all of the time” to one point for “none” of the time. This results in individual K10 scores being restricted to a range of 10–50. Procedure. This review article deals with the methods for measurement of subjective well-being in children and adolescents. The first part of the review presents an outlook of several psychometric instruments that demonstrate various approaches to measuring subjective well-being: multidimensional and one-dimensional multiple-item scales as well as single item scales. The examples of techniques for establishing validity of developing instruments are provided. The experience of adaptation of psychometric instruments for measuring well-being to use them in younger samples is discussed. The article also highlights Subjective well-being (SWB) is one way of understanding what it means to distinct individuals. Likewise, different things are associated with positive affect for distinct people; this is the subjective aspect of affective appraisal. SWB thus encompasses a vast array of different concepts, from fleeting experiences in our day-to-day, to much broader global judgments that we make about our lives as a whole (Kim-Priesto et al., 2005). It is typically considered a hedonic as opposed to a eudaimonic concept (Deci & Ryan, 2008; Huta & Waterman, 2014). Diener and SWB Theory. Professor Ed Diener is one of the world’s foremost SWB researchers, coining the construct in his seminal 1984 a