An Evaluation of the Precision of Measurement of Ryff's Psychological Well-Being Scales in a Population Sample

Rosemary A. Abbott · George B. Ploubidis · Felicia A. Huppert · Diana Kuh · Tim J. Croudace

Abstract The aim of this study is to assess the effective measurement range of Ryff's Psychological Well-being scales (PWB). It applies normal ogive item response theory (IRT) methodology using factor analysis procedures for ordinal data based on a limited information estimation approach. The data come from a sample of 1,179 women participating in a midlife follow-up of a national birth cohort study in the UK. The PWB scales incorporate six dimensions: autonomy, positive relations with others, environmental mastery, personal growth, purpose in life and self-acceptance. Scale information functions were calculated to derive standard errors of measurement for estimated scores on each dimension. Construct variance was distinguished from method variance by inclusion of method factors from item wording (positive versus negative). Our IRT analysis revealed that the PWB measures well-being most accurately in the middle range of the score distribution, i.e. for women with average well-being. Score precision diminished at higher levels of well-being, and low well-being was measured more reliably than high well-being. A second-order well-being factor loaded by four of the dimensions achieved higher measurement precision and greater score accuracy across a wider range than any individual dimension. Future development of well-being scales should be designed to include items that are able to discriminate at high levels of well-being.

1 Introduction

Over recent years there has been an increasing scientific, populist and political interest in notions of positive mental health and well-being (Dolan and White 2007; Huppert et al.

R. A. Abbott (✉) · G. B. Ploubidis · F. A. Huppert · T. J. Croudace
University of Cambridge, Cambridge, UK
e-mail: raa25@medschl.cam.ac.uk

D. Kuh
MRC Unit for Lifelong Health and Ageing, London, UK

2005; Kahneman et al. 1999; Seligman 2002; Snyder et al. 2001). This widening interest in positive aspects of mental health has led to the development of new scientific constructs and also questionnaire items designed to measure positive well-being.

One of the most widely used survey instruments is Ryff's multi-dimensional Psychological Well-being scales (PWB). It was specifically defined to measure positive aspects of psychological functioning along six theoretically-motivated dimensions: independence and self-determination (autonomy); having satisfying, high quality relationships (positive relations with others); the ability to manage one's life (environmental mastery); being open to new experiences (personal growth); believing that one's life is meaningful (purpose in life); and a positive attitude towards oneself and one's past life (self-acceptance), (Ryff 1989a, b; Ryff and Keyes 1995).

Various versions of Ryff's PWB have been extensively used in a variety of samples and settings. PWB items have been administered in large population-based samples such as the US National Survey of Families and Households (Sweet and Bumpass 1996), Midlife in the United States, (Brim et al. 2004), the Canadian Study of Health and Aging (Clarke et al. 2001), and the English Longitudinal Study of Ageing (ELSA). The Ryff items are also used as well-being outcomes in smaller surveys, such as the study of body consciousness (McKinley 1999), life challenges (McGregor and Little 1998) or midlife work aspirations (Carr 1997), as well as outcomes of therapeutic interventions (Fava et al. 2005).

There have been a number of psychometric studies of the multi-dimensional structure of the Ryff PWB (Abbott et al. 2006; Burns and Machin 2009; Cheng and Chan 2005; Clarke et al. 2001; Kafka and Kozma 2002; Lindfors et al. 2006; Ryff and Keyes 1995; Springer and Hauser 2006; Van Dierendonck 2004; van Dierendonck et al. 2008). Cumulative knowledge about the performance of the PWB is hard to glean since several versions are in use, varying in length from 120 to 12 items with varying degrees of item overlap. In general however, whilst the above psychometric studies provide some support for the *a priori* six dimensional structure, high inter-factor correlations are typically reported and some item have been found to cross-load on more than one factor (Clarke et al. 2001; Springer and Hauser 2006).

Several of these studies have sought to model the more problematic aspects of the PWB latent structure in some way to improve understanding of the items, and to improve model fit. For example, models have included method factors to address measurement artifacts due to item wording, or excluded items with low factor loadings or high cross-loadings (Abbott et al. 2006; Burns and Machin 2009; Cheng and Chan 2005; Springer and Hauser 2006).

In a recent study in a UK epidemiological sample of women, Abbott et al. (2006) undertook a detailed psychometric assessment of the 42-item PWB using factor analysis procedures appropriate for the ordinal response format of the Ryff items. Their findings broadly confirmed the six-factor structure; but method factors were necessary to achieve acceptable model fit and to ensure that construct variance was not obscured by common response tendencies to similarly worded items (Abbott et al. 2006). Further, in this sample the modelling supported the notion of a second-order general well-being factor defined by loadings from four of the six-first-order dimensions (environmental mastery, personal growth, purpose in life, self-acceptance). The high correlation between these four factors has also been confirmed in subsequent studies (e.g. Burns and Machin 2009).

All psychometric studies to date have been concerned with the construct validity of the PWB scales. Little is known about the precision of measurement of the PWB scales i.e. the accuracy of the scores that are derived from applications of this measure of well-being in different samples. To our knowledge, the reliability (or precision of measurement) of estimated PWB scores has not previously been investigated. Since there are many scales in

existence which measure negative aspects of well-being by assessing symptoms of mental disorder, it is particularly important for a scale which purports to be about positive well-being to show high precision of measurement at high values of the scale. Ideally, for population wide measurement of individual variation in these new constructs, items need to be designed with response wording that enables well-being to be measured accurately across the range from low, through mid-range to high levels, so that the effective measurement range is as wide as possible.

The aim of this paper is to establish the precision of measurement of the 42-item version of Ryff's Psychological Well-being scales. We use the parameters of an ordinal factor analysis model to profile the effective measurement range of the six PWB subscales and a second-order general well-being factor. Our conclusions are derived from graphical representations of score accuracy (precision) at each score value, calculated from the Test Information Functions for each subscale. Unlike a classical test theory based analysis, these plots show variations in score reliability, and determine where precision of measurement is highest or lowest, across the measurement range.

2 Methods

2.1 Sample

The Medical Research Council 1946 National Survey of Health and Development (NSHD), the first British birth cohort study is a stratified sample of singleton births occurring to married parents in England, Scotland and Wales during the week of 3–9 March 1946, (Wadsworth 1991; Wadsworth et al. 2006). The original sample comprised 5,362 individuals (2,547 women). Data have been collected regularly since childhood.

The representativeness of the study sample has been well documented at various points during their adult life follow-up (Wadsworth 1991; Wadsworth et al. 2006). For example, a comparison of the sample retained at age 43 and 53 with population census data has shown that the NSHD survey members are generally representative of the national population of a similar age (Wadsworth et al. 2003).

The sample for this study comprised women participants who participated in a health survey in midlife. This sub-study of women's health in midlife (Women's Health Survey (WHS)) was undertaken annually, by postal questionnaire, between the ages of 47–54. This study included 1,778 (70%) of the original cohort of women; the others had died (6%), previously refused to take part (12%) or lived abroad and were not in contact with the study or could not be traced (13%).

2.2 Measures

A 42-item version of Ryff's Psychological Well-being scales were included in the WHS at age 52 and sent to 1,421 women who had completed at least one WHS survey in the previous 2 years. The 42-item version of the PWB was selected for use in this sample on the personal recommendation of Ryff. The PWB was not administered to men in this cohort. The response format for all PWB items comprised six ordered categories labelled from 'disagree strongly' to 'agree strongly'. Twenty PWB items comprised positive item content and 22 had negative item content. Prior to analysis, items with negative content were reverse scored so that high values indicated well-being. Full question wording of the 42-items is shown in Table 1.

Table 1 Frequencies of response categories and descriptive statistics, Ryff psychological well-being scales ($N = 1,179$)

	Response categories[a]						Missing (%)	Mean	SD	Skew	Kurtosis	CITC
	1 (%)	2 (%)	3 (%)	4 (%)	5 (%)	6 (%)						
Autonomy												
A1+ I am not afraid to voice my opinions even when they are in opposition to the opinions of most people	4.0	11.5	10.7	19.9	34.8	16.7	2.7	4.2	1.4	−0.7	−0.5	0.45
A2+ My decisions are not usually influenced by what everyone else is doing	4.0	8.8	12.9	20.6	34.9	16.9	2.0	4.3	1.3	−0.7	−0.3	0.43
A4+ I have confidence in my opinions even if they are contrary to the general consensus	2.6	5.4	10.8	24.1	35.8	18.9	2.4	4.5	1.2	−0.8	0.3	0.51
A6+ Being happy with myself is more important than having others approve of me	1.5	4.1	10.0	17.9	29.4	35.2	1.9	4.8	1.3	−1.0	0.3	0.36
A3− I tend to worry what other people think of me	*10.1*	*14.3*	*29.4*	*9.6*	*15.0*	*19.8*	*1.8*	*3.7*	*1.6*	*0.1*	*−1.2*	*0.45*
A5− I often change my mind about decisions if my friends and family disagree	*3.5*	*13.8*	*28.1*	*19.4*	*20.3*	*12.1*	*2.9*	*3.8*	*3.8*	*0.2*	*−0.9*	*0.38*
A7− It is difficult for me to voice my own opinions on controversial matters	*6.8*	*11.8*	*18.7*	*15.1*	*21.3*	*23.6*	*2.7*	*4.1*	*4.1*	*−0.3*	*−1.0*	*0.51*
Positive relations with others												
R1+ Most people see me as loving and affectionate	1.6	2.5	5.3	15.1	43.0	29.9	2.6	4.9	1.1	−1.3	2.1	0.32
R3+ I enjoy personal and mutual conversations with family members or friends	1.0	1.6	2.1	8.3	25.8	59.7	1.4	5.4	1.0	−2.1	5.1	0.41
R6+ People would describe me as a giving person, willing to share my time with others	1.1	1.1	3.5	17.2	37.2	37.6	2.2	5.1	1.0	−1.3	2.3	0.42
R7+ I know that I can trust my friends and they know that they can trust me	0.9	1.1	2.4	8.7	28.9	55.8	2.1	5.4	0.9	−2.0	4.8	0.43
R2− I often feel lonely because I have few close friends with whom to share my concerns	*7.1*	*7.1*	*13.5*	*9.6*	*18.6*	*42.8*	*1.3*	*4.6*	*1.6*	*−0.9*	*−0.6*	*0.47*
R4− I don't have many people who want to listen when I need to talk	*4.1*	*9.3*	*12.2*	*11.2*	*25.0*	*36.4*	*1.7*	*4.6*	*1.5*	*−0.8*	*−0.5*	*0.58*
R5− It seems to me that most other people have more friends than I do	*6.0*	*10.0*	*16.0*	*12.3*	*20.4*	*32.5*	*2.8*	*4.3*	*1.6*	*−0.6*	*−0.9*	*0.60*
Environmental mastery												
E2+ I am quite good at managing the many responsibilities of my daily life	1.7	1.6	2.3	10.2	39.9	42.5	1.8	5.2	1.0	−1.8	4.4	0.51
E4+ I generally do a good job of taking care of my personal finances and affairs	2.4	2.4	3.0	11.9	36.1	42.0	2.1	5.1	1.1	−1.7	3.1	0.39

Table 1 continued

	Response categories[a]						Missing (%)	Mean	SD	Skew	Kurtosis	CITC
	1 (%)	2 (%)	3 (%)	4 (%)	5 (%)	6 (%)						
E5+ I am good at juggling my time so that I can fit everything in that needs to be done	2.9	6.0	7.0	14.9	35.1	32.0	2.1	4.7	1.3	-1.1	0.6	0.43
E7+ I have been able to build a home and a lifestyle for myself that is much to my liking	2.7	2.5	5.6	16.3	33.5	37.2	2.1	4.9	1.2	-1.3	1.7	0.55
E1- I do not fit very well with the people and the community around me	2.5	4.8	6.7	6.9	23.4	53.6	2.1	5.1	1.3	-1.5	1.5	0.27
E3- I often feel overwhelmed by my responsibilities	4.7	9.8	19.0	15.0	21.4	28.0	2.1	4.3	1.5	-0.4	-0.9	0.41
E6- I have difficulty arranging my life in a way that is satisfying to me	4.9	10.0	16.6	15.7	25.5	25.2	2.3	4.3	1.5	-0.5	-0.8	0.53
Personal growth												
G3+ I think it is important to have new experiences that challenge how you think about the world	2.2	4.3	8.4	24.3	29.7	29.5	1.6	4.7	1.2	-0.9	0.4	0.48
G5+ I have the sense that I have developed a lot as a person over time	2.5	3.4	7.9	22.8	34.4	26.5	2.5	4.7	1.2	-1.0	0.8	0.37
G1- I am not interested in activities that will expand my horizons	2.0	6.8	10.3	11.9	24.7	41.9	2.4	4.8	1.4	-1.0	0.1	0.50
G2- I don't want to try new ways of doing things — my life is fine the way it is	8.5	15.6	11.9	20.4	26.5	15.2	1.9	3.9	1.5	-0.4	-1.0	0.48
G4- When I think about it, I haven't really improved much as a person over the years	4.5	10.0	9.6	13.2	27.5	33.3	1.8	4.5	1.5	-0.8	-0.4	0.41
G6- I do not enjoy being in new situations that require me to change my old familiar ways of doing things	8.6	12.9	23.3	16.8	19.5	16.2	2.6	3.7	1.5	-0.1	-1.0	0.52
G7- There is a truth in the saying that you can't teach an old dog new tricks	8.6	11.9	16.1	16.0	21.1	23.9	2.5	4.1	1.6	-0.4	-1.0	0.43
Purpose in life												
P5+ I am an active person in carrying out the plans I set for myself	1.8	2.9	7.6	24.3	35.0	26.1	2.3	4.7	1.2	-0.9	0.9	0.43
P7+ I enjoy making plans for the future and working to make them a reality	2.4	4.3	11.5	22.2	30.6	25.7	3.2	4.6	1.5	-0.8	0.1	0.47
P1- I tend to focus on the present, because the future nearly always brings me problems	6.4	12.3	13.1	12.2	22.5	28.9	4.6	4.2	1.6	-0.5	-1.0	0.35
P2- My daily activities often seem trivial and unimportant to me	5.2	9.4	15.3	10.5	23.0	34.7	2.0	4.4	1.6	-0.7	-0.8	0.48
P3- I don't have a good sense of what it is I am trying to accomplish in life	5.5	11.5	17.4	11.8	22.7	27.9	3.1	4.2	1.6	-0.5	-1.0	0.51
P4- I used to set goals for myself but that now seems a waste of time	2.3	7.2	16.0	15.7	25.0	31.2	2.6	4.5	1.4	-0.6	-0.6	0.56

Table 1 continued

	Response categories[a]						Missing (%)	Mean	SD	Skew	Kurtosis	CITC
	1 (%)	2 (%)	3 (%)	4 (%)	5 (%)	6 (%)						
P6− *I sometime feel I have done all there is to do in life*	3.5	4.4	7.6	10.8	20.9	50.7	2.1	5.0	1.4	−1.4	0.9	0.36
Self-acceptance												
S2+ I have made some mistakes in the past, but feel that all in all everything has worked out for the best	4.5	4.9	8.3	17.5	33.0	29.2	2.5	4.6	1.4	−1.0	0.4	0.37
S5+ The past had its ups and downs, but in general I wouldn't want to change it	6.9	6.5	10.7	14.8	30.3	28.0	2.7	4.4	1.5	−0.9	−0.2	0.52
S6+ When I compare myself with friends and acquaintances, it makes me feel good about who I am	4.0	6.0	12.2	23.3	30.3	21.2	3.0	4.4	1.3	−0.7	−0.1	0.55
S7+ In general, I feel confident and positive about myself	5.4	6.8	10.0	14.1	39.5	21.0	3.2	4.4	1.4	−0.1	0.1	0.57
S1− *I feel that many of the people I know have got more out of life than I have*	6.4	9.0	14.3	10.6	21.8	36.7	1.2	4.4	1.6	−0.7	−0.8	0.58
S3− *In many ways, I feel disappointed about my achievements in life*	5.3	8.7	16.8	11.5	21.6	34.3	1.8	4.4	1.6	−0.6	−0.8	0.62
S4− *My attitude about myself is probably not as positive as most people feel about themselves*	6.1	17.4	23.6	13.8	18.3	18.5	2.3	3.8	1.5	−0.1	−1.2	0.52

No. of items with positive content are denoted by a + sign after the item number. Items with negative content and denoted by −sign and presented in italics

Items with negative content are reverse coded. Items are listed by subscale and not in the order in which they were presented

[a] Response categories (1) strongly disagree; (2) moderately disagree; (3) slightly disagree; (4) slightly agree; (5) moderately agree; (6) strongly agree

CITC Corrected item total correlation

2.3 Factor Analysis Modelling of Ordinal Responses (Latent Variable Probit IRT)

An underlying variable approach to ordinal data factor analysis implements a probit (normal ogive) item response model with thresholds defined on a latent continuum that capture changes in response level. Factor analysis of these underlying variables can then be carried out using traditional linear factor analysis methods applied to polychoric correlations. This approach has been popularised by Muthén (1984) and Jöreskog (1990), with related work by McDonald (1999). Software for binary and ordinal item factor analysis (equivalent to an item response model) is now widely available for example, LISREL (Jöreskog and Sörbom 2004), *Mplus* (Muthén and Muthén 1998–2004) and NOHARM (binary items only) (Baker 2001).

Here we applied the ordinal probit item response model to a 42-item version of Ryff's Psychological Well-being scales using *Mplus* software, Version 4.2 (Muthén and Muthén 1998–2004). Parameters were estimated using robust, weighted least squares (rWLS) methods that rely only on the bivariate associations among items. This method was undertaken rather than one based on full information maximum likelihood (FIML) as a FIML model would have been computationally unmanageable due to the large number of factors (6 construct + 2 method) (See: http://www.statmodel.com/discussion). We would not expect our conclusions to differ if a full information approach had been used since many other studies have compared results under these approaches with fewer factors (Bartholomew and Knott 1999; Takane and de Leeuw 1987).

Parameter estimates for the factor loadings, which quantify the discriminating power of each item, and probit thresholds from *Mplus* output are used to calculate Test Information Curves (TICs), that represent the accuracy of each estimated score value across the measurement range of the instrument. TICs are usually associated with item response theory perspectives based on FIML estimation procedures, but are also calculated by the ordinal factor analysis procedures in *Mplus* and can be plotted using graphics commands.

TICs are also commonly referred to as Scale Information Functions (SIFs) and are derived from the amount of Fisher information i.e. the reciprocal of the square root of the posterior standard deviation of the estimated score (posterior mean), (Baker and Kim 2004). The Scale Information Function is generally displayed graphically, along with associated plots, such as the conditional standard error of measurement (SEM) calculated from the inverse of the square root of the information function. These plots are important in the interpretation of measurement precision across the latent trait continuum. They identify the point on a scale, or the range of scale values, which are measured with high precision i.e. where standard errors of measurement are low, and score accuracy therefore high. The point on the measurement continuum where standard errors start to increase indicates less precise measurement and lower score reliability or accuracy. This can occur at either end, and conceivably (but rarely) in the middle range of scores.

Our modelling of the Ryff PWB scales was based on a previously reported confirmatory factor analysis of the Ryff items which evaluated alternate factor structures of the scale (Abbott et al. 2006). This study suggested that the addition of method factors to address methodological artefacts due to positive versus negative item content significantly improved model fit compared to the *a priori* six-factor structure. In addition, four of the six dimensions of well-being (environmental mastery, personal growth, purpose in life, and self-acceptance) were sufficiently highly correlated to warrant a second-order general well-being factor. The remaining two factors, autonomy positive relations with others were more independent. Other minor model modifications included the omission of two poor-fitting items from personal growth (resulting in a 40-item scale), moving one item from environmental mastery onto
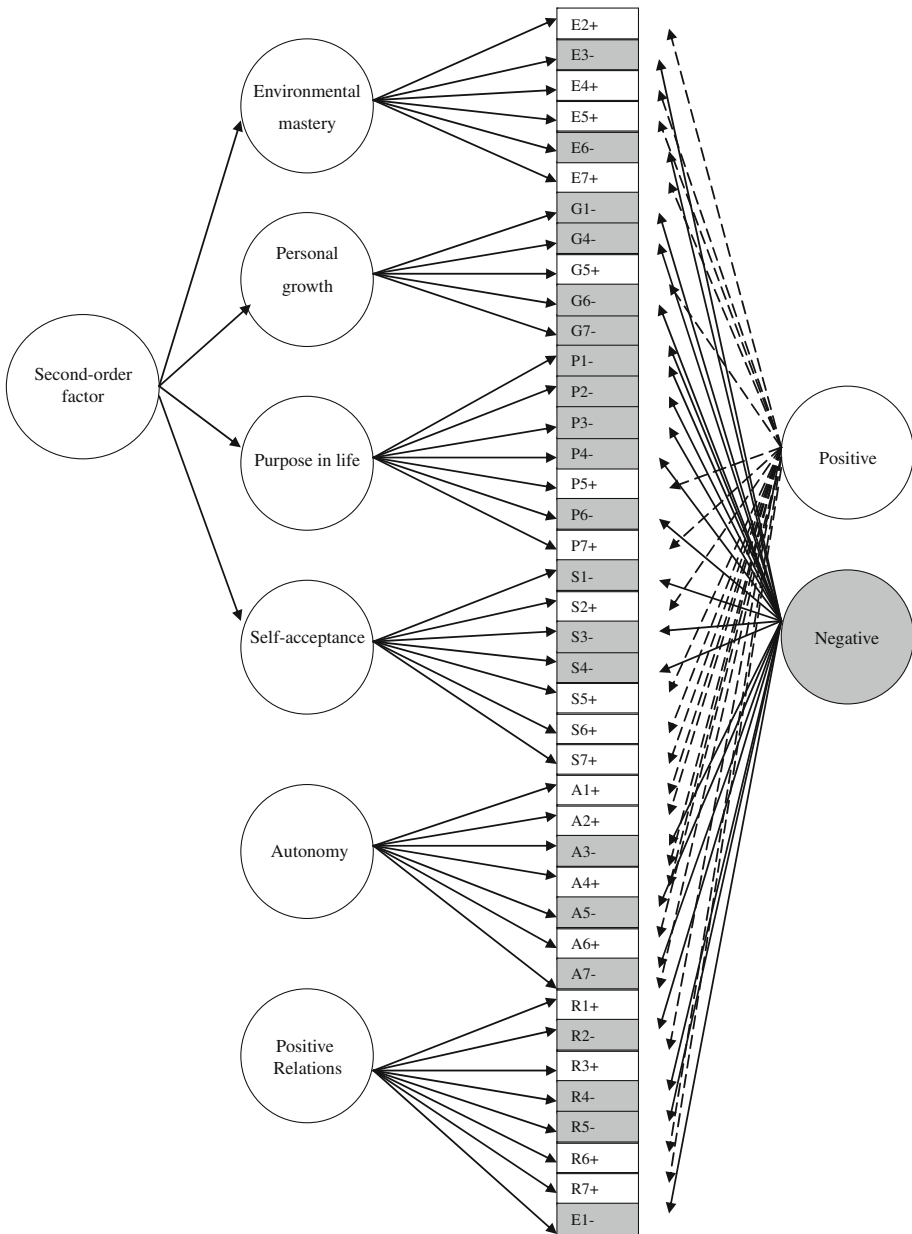
Fig. 1 Psychological Well-being modified 40-item model with second-order factor and method factors. Residual correlations between the six PWB latent variables ranged from 0.25 to 0.85 (not displayed due to model complexity). Goodness of fit: Chi Square: 2.46 (df = 255), TLI = 0.94, RMSEA = 0.086, WRMR = 2.01

positive relations, and allowing correlated residuals between two items from positive relations (see Abbott et al. 2006 for further details). This model is graphically represented in Fig. 1. It should be noted that all six-first-order factors are correlated (with residual

correlations ranging from 0.25 to 0.85). The method factors are uncorrelated with the latent constructs. Small amounts of missing item level data were present (957/1,179 had complete data on all items) but not in great enough numbers to influence the results. rWLS estimation in *Mplus* includes partially missing data under a MCAR assumption.

3 Results

The analysis sample included 1,179 respondents who completed at least 85% of PWB items (36 out of 42 questions); 957 had complete data on all items. Table 1 shows the percentage of participants responding to each of the six-point Likert style response categories. Negative items have been reverse scored so all items measure well-being in the same direction, i.e. one indicates the lowest level of well-being and six the highest. There were few responses to the lowest levels of well-being, particularly for questions with positive item content. For eight of the items, over 40% of respondents endorsed the highest (most positive) category, indicating ceiling effects on these items. Treating responses as continuous, interval scores, as in traditional factor analysis approaches to psychometric analysis, yields mean scores and corrected item-total correlations (CITC) by subscale, these are also reported in Table 1.

3.1 Item Factor Loadings and Thresholds

Factor loadings estimated simultaneously for the six-first-order factors and then for the general well-being factor, are reported in Table 2. These loadings identify the relative discriminating power of each item as a measure of the intended latent construct. In factor analysis models for Likert scored items, an item's discriminating power is captured by the magnitude of the factor loading (the correlation of the factor with the underlying variable).

Only ten of the forty Ryff items included in the analysis were highly discriminating indicators of their latent factor i.e. loaded above or equal to 0.70 (Table 2; column 3 'standardised'). This indicates that the latent constructs explained around 50% of the variation in responses for only a quarter of the items (A7−, R2−, R4−, R5−, E6−, G4−, G5 + , P3−, S1−, S3−). We note that these discriminating items were distributed across all six-factors, and with the exception of one from personal growth, were characterised by negative item content.

Conversely, seven items displayed low discriminative ability, with factor loadings of < 0.40. These items were spread across five of the six dimensions, the exception being purpose in life where all items loaded > 0.40. Four of seven low discriminating items had positive item content (R1 + , E4 + , E5 + , S2 +) and three had negative item content (A5−, G1−, G7−).

Overall, items with negative content were found to have higher factor loadings than items with positive content. Three-quarters of the items with negative content had factor loadings of more than 0.50 compared to only half of the items with positive content. Table 2 also shows the five thresholds corresponding to the distinctions between the six ordinal category response options.

3.2 Conditional Standard Error of Measurement of PWB Subscales

Figure 2 shows the conditional standard error of measurement (SEM) along the *y*-axis for each *a priori* subscale of the PWB. Here the *x*-axis represents the population continuum of

Table 2 IRT parameters—psychological well-being, 40-item model ($N = 1,179$)

	Factor loadings λ		Thresholds (Tau) (SE)					Second-order factor loadings/discrimination parameters (a)	
	Unstandardised (SE)	Standardised	1	2	3	4	5	Unstandardised (SE)	Standardised
Autonomy									
A1+	1.00 (0.00)	0.52	−1.78 (0.07)	−1.01 (0.04)	−0.62 (0.04)	−0.07 (0.04)	0.94 (0.04)	n/a	
A2+	0.94 (0.07)	0.49	−1.76 (0.07)	−1.13 (0.05)	−0.65 (0.04)	−0.07 (0.04)	0.95 (0.04)		
A4+	1.16 (0.07)	0.60	−1.95 (0.08)	−1.41 (0.05)	−0.87 (0.04)	−0.15 (0.04)	0.86 (0.04)		
A6+	0.77 (0.07)	0.40	−2.16 (0.09)	−1.59 (0.06)	−1.00 (0.04)	−0.42 (0.04)	0.36 (0.04)		
A3−	1.24 (0.09)	0.64	−1.26 (0.05)	−0.67 (0.04)	0.12 (0.04)	0.37 (0.04)	0.83 (0.04)		
A5−	0.75 (0.07)	0.39	−1.80 (0.07)	−0.93 (0.04)	−0.08 (0.04)	0.44 (0.04)	1.15 (0.05)		
A7−	1.43 (0.09)	0.74	−1.48 (0.06)	−0.88 (0.04)	−0.30 (0.04)	0.10 (0.04)	0.70 (0.04)		
Positive relations with others									
R1+	1.00 (0.00)	0.32	−2.18 (0.10)	−1.75 (0.07)	−1.31 (0.05)	−0.68 (0.04)	0.51 (0.04)	n/a	
R3+	1.57 (0.16)	0.50	−2.39 (0.12)	−1.95 (0.08)	−1.69 (0.06)	−1.13 (0.05)	−0.28 (0.04)		
R6+	1.33 (0.13)	0.42	−2.32 (0.11)	−2.03 (0.08)	−1.57 (0.06)	−0.72 (0.04)	0.30 (0.04)		
R7+	1.55 (0.16)	0.49	−2.35 (0.11)	−2.06 (0.09)	−1.70 (0.06)	−1.11 (0.05)	−0.18 (0.04)		
R2−	2.30 (0.23)	0.73	−1.47 (0.06)	−1.07 (0.05)	−0.59 (0.04)	−0.32 (0.04)	0.17 (0.04)		
R4−	2.48 (0.24)	0.79	−1.72 (0.07)	−1.10 (0.05)	−0.64 (0.04)	−0.32 (0.04)	0.33 (0.04)		
R5−	2.51 (0.24)	0.80	−1.53 (0.06)	−0.98 (0.04)	−0.44 (0.04)	−0.11 (0.04)	0.44 (0.04)		
E1−	1.90 (0.20)	0.60	−1.98 (0.08)	−1.46 (0.06)	−1.08 (0.05)	−0.80 (0.04)	−0.12 (0.04)		
Environmental mastery									
E2+	1.00 (0.00)	0.52	−2.16 (0.09)	−1.87 (0.07)	−1.61 (0.06)	−1.00 (0.04)	0.17 (0.04)	1.00 (0.00)	0.60
E4+	0.63 (0.06)	0.33	−2.01 (0.08)	−1.68 (0.06)	−1.42 (0.05)	−0.84 (0.04)	0.18 (0.04)	0.72 (0.04)	0.43
E5+	0.72 (0.06)	0.37	−1.91 (0.08)	−1.35 (0.05)	−0.99 (0.04)	−0.49 (0.04)	0.45 (0.04)	0.86 (0.04)	0.52
E7+	1.33 (0.07)	0.69	−1.92 (0.08)	−1.62 (0.06)	−1.22 (0.05)	−0.59 (0.04)	0.31 (0.04)	1.16 (0.04)	0.70
E3−	1.17 (0.07)	0.58	−1.68 (0.06)	−1.05 (0.05)	−0.41 (0.04)	−0.02 (0.04)	0.56 (0.04)	0.86 (0.04)	0.51
E6−	1.55 (0.09)	0.81	−1.67 (0.06)	−1.04 (0.05)	−0.47 (0.04)	−0.05 (0.04)	0.65 (0.04)	1.15 (0.04)	0.69

Table 2 continued

	Factor loadings λ		Thresholds (Tau) (SE)					Second-order factor loadings/discrimination parameters (a)	
	Unstandardised (SE)	Standardised	1	2	3	4	5	Unstandardised (SE)	Standardised
Personal growth									
G5+	1.97 (0.20)	0.71	−1.98 (0.08)	−1.57 (0.06)	−1.08 (0.05)	−0.32 (0.04)	0.61 (0.04)	1.17 (0.04)	0.70
G1−	1.00 (0.00)	0.36	−2.04 (0.08)	−1.35 (0.05)	−0.86 (0.04)	−0.48 (0.04)	0.18 (0.04)	0.60 (0.05)	0.36
G4−	2.36 (0.23)	0.84	−1.71 (0.07)	−1.06 (0.05)	−0.70 (0.04)	−0.31 (0.04)	0.41 (0.04)	1.14 (0.05)	0.68
G6−	1.26 (0.12)	0.45	−1.35 (0.05)	−0.77 (0.04)	−0.10 (0.04)	0.34 (0.04)	0.97 (0.04)	0.74 (0.04)	0.44
G7−	0.65 (0.10)	0.23	−1.35 (0.05)	−0.81 (0.04)	−0.32 (0.04)	0.10 (0.04)	0.70 (0.04)	0.49 (0.04)	0.29
Purpose in life									
P5+	1.32 (0.11)	0.55	−2.08 (0.09)	−1.67 (0.06)	−1.15 (0.05)	−0.31 (0.04)	0.62 (0.04)	1.10 (0.04)	0.66
P7+	1.36 (0.11)	0.56	−1.98 (0.08)	−1.49 (0.06)	−0.88 (0.04)	−0.20 (0.04)	0.63 (0.04)	1.05 (0.04)	0.63
P1−	1.00 (0.00)	0.41	−1.50 (0.06)	−0.86 (0.04)	−0.43 (0.04)	−0.10 (0.04)	0.52 (0.04)	0.70 (0.05)	0.42
P2−	1.50 (0.11)	0.62	−1.62 (0.06)	−1.04 (0.05)	−0.51 (0.04)	−0.22 (0.04)	0.38 (0.04)	0.94 (0.04)	0.56
P3−	1.70 (0.13)	0.70	−1.58 (0.06)	−0.93 (0.04)	−0.37 (0.04)	−0.06 (0.04)	0.56 (0.04)	1.06 (0.04)	0.64
P4−	1.62 (0.12)	0.67	−2.01 (0.08)	−1.30 (0.05)	−0.64 (0.04)	−0.19 (0.04)	0.47 (0.04)	1.07 (0.04)	0.65
P6−	1.05 (0.10)	0.44	−1.81 (0.07)	−1.41 (0.05)	−1.01 (0.04)	−0.62 (0.04)	−0.04 (0.04)	0.74 (0.05)	0.45
Self-acceptance									
S2+	0.49 (0.04)	0.38	−1.68 (0.06)	−1.30 (0.05)	−0.91 (0.04)	−0.35 (0.04)	0.53 (0.04)	0.76 (0.04)	0.46
S5+	0.68 (0.03)	0.53	−1.47 (0.06)	−1.10 (0.05)	−0.69 (0.04)	−0.26 (0.04)	0.56 (0.04)	0.97 (0.04)	0.59
S6+	0.70 (0.03)	0.55	−1.75 (0.07)	−1.26 (0.05)	−0.75 (0.04)	−0.08 (0.04)	0.78 (0.04)	1.00 (0.04)	0.60
S7+	0.86 (0.03)	0.67	−1.60 (0.06)	−1.16 (0.05)	−0.76 (0.04)	−0.33 (0.04)	0.78 (0.04)	1.13 (0.04)	0.68

Table 2 continued

	Factor loadings λ		Thresholds (Tau) (SE)					Second-order factor loadings/discrimination parameters (a)	
	Unstandardised (SE)	Standardised	1	2	3	4	5	Unstandardised (SE)	Standardised
S1–	1.00 (0.00)	0.78	−1.52 (0.06)	−1.01 (0.04)	−0.53 (0.04)	−0.24 (0.04)	0.33 (0.04)	1.22 (0.05)	0.74
S3–	1.09 (0.03)	0.85	−1.62 (0.06)	−1.07 (0.05)	−0.49 (0.04)	−0.18 (0.04)	0.39 (0.04)	1.31 (0.05)	0.79
S4–	0.84 (0.03)	0.66	−1.55 (0.06)	−0.71 (0.04)	−0.05 (0.04)	0.31 (0.04)	0.88 (0.04)	1.04 (0.04)	0.63

Based on modified version of Ryff' 42-item PWB with 40-items. $N = 1{,}179$

Under Muthéns normal ogive item response modelling approach, an items discriminating power is captured by the magnitude of the factor loading standardised as a range from −1.0 to 1.0
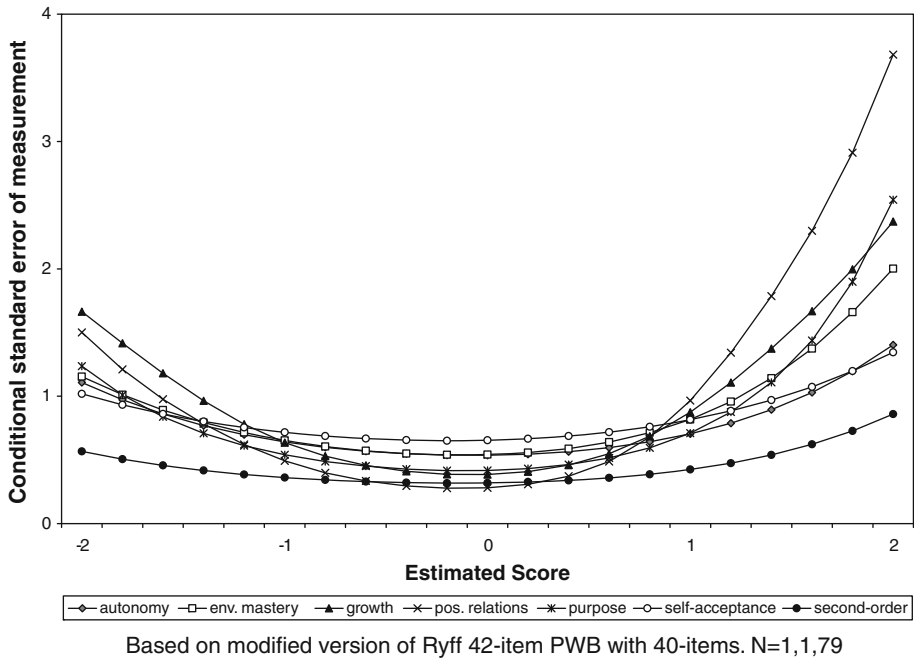
Based on modified version of Ryff 42-item PWB with 40-items. N=1,1,79

Fig. 2 Conditional Standard Error of Measurement—Psychological Well-being (40-item model). Based on modified version of Ryff' 42-item PWB with 40-items. $N = 1,1,79$

estimated latent trait scores in terms of a standardized normal distribution (*Mean* = 0, SD = 1).

The subscales here comprise between five (personal growth) and eight (positive relations) items. Each first-order PWB factor measured well-being with only modest precision. Within the mid-range of the distribution (+1.0 and −1.0 SD) which represents 68% of the population, conditional SEM values were relatively flat for all factors and generally below 0.6 in value. Three of the factors displayed slightly higher levels of measurement precision (positive relations, personal growth and purpose in life). Within the mid-range score accuracy for the self-acceptance subscale was lower than any of the other factors ranging from a conditional SEM of 0.65 at the *Mean* to 0.8 at +1.0 SD.

At the negative end of the well-being spectrum (low well-being) reliability of scale scores was lower for all factors, compared to scores in the mid-range (average well-being). Measurement precision for positive relations and personal growth diminished rapidly once *x*-axis values exceeded −1.5. At the positive end of the continuum (i.e. high well-being) estimated well-being scores had lower precision than for scores across the rest of the range. Measurement precision for positive relations with others, which was relatively high around the mean, diminished rapidly with increasing values on the *x*-axis. A similar, but slightly less extreme pattern was apparent for purpose in life, personal growth and environmental mastery. Measurement precision for autonomy and self-acceptance also declined at both ends of the spectrum but less rapidly than the other factors. Of the six-first-order factors, self-acceptance appeared to be the dimension measuring well-being with the greatest accuracy across the widest range of the population well-being continuum, with conditional SEM values displaying a relatively flat profile from 2.0 SD below to 2.0 SD above the

mean (a wide effective measurement range). However these score SEMs are not small enough for precise statements to be made about individual scores, only for populations or groups.

Finally, we examined the impact of capturing well-being at a more general level of analysis, by presenting the conditional SEM based on a second-order well-being factor. It can be seen from Fig. 2, that at the population mean (x-axis = 0), this composite general well-being factor achieved slightly better precision, than each of its four component factors. Importantly, the combination of the items from these four factors increases the precision of measurement beyond the mid-range of the well-being continuum, yielding a relatively flat profile from +2.0 to −2.0 SD of the mean. Examination of individual latent trait scores shows that on the general well-being factor only one individual scored at the lowest estimated latent trait level (−1.90), and 4 individuals at the highest level (+1.71). This suggests that the score distribution estimated from this more general factor model is not compromised by either a floor or ceiling effect despite responses to individual Ryff items often concentrated at the largest score values (see Table 1).

4 Discussion

Ideally, survey instruments used to measure psychological constructs in general population surveys and epidemiological studies should contain item phrasing, response wording and sufficient response categories that enable accurate and precise scores to be estimated across a wide measurement range. New measures designed to evaluate well-being in populations, rather than clinical samples, should differentiate between individuals with high levels of well-being, as well as between those with medium and low levels. The Ryff scales of psychological well-being (Ryff 1989b; Ryff and Keyes 1995) were designed to measure a continuum of positive psychological functioning (Ryff et al. 2006) but no psychometric study has commented on score accuracy or effective measurement range. Psychometric models specifically calculate these properties and our graphs display them for the first time.

We expected that plots of the conditional standard error of measurement would yield relatively flat curves and span a wide range of the x-axis values, indicating precise measurement right across the population continuum. However, our results were not so clear cut. We used item response theory to report the measurement range of the PWB. An ordinal (normal ogive) item response model using an underlying variable approach was applied to the six PWB subscales (autonomy; positive relations with others, environmental mastery, personal growth, purpose in life and self-acceptance) to quantify and plot the conditional standard error of measurement for each score value. We based our IRT analysis on recommended model modifications which have been shown to result in a parsimonious solution for the 42-item version of the PWB (Abbott et al. 2006). This analysis demonstrated that for each subscale, information was concentrated in the middle of the measurement range, i.e. around the population average. Score precision diminished at high and low levels of well-being, but low well-being was measured more reliably than high well-being. Only a quarter of the individual items displayed high discrimination (factor loadings > 0.7). All but one of these items contained negative item content, in other words, the most discriminating items measured well-being by questions about its absence and response disagreeing with such statements.

In previous work in this national epidemiological sample, we have shown that four factors (environmental mastery, personal growth, purpose in life and self-acceptance) were highly inter-related and could be parsimoniously modelled by a second-order factor with

autonomy and positive relations remaining as more independent (Abbott et al. 2006). Our IRT modelling of this second-order well-being factor revealed that it had higher measurement precision across a wider range than the component subscales. This is to be expected because the number of items was four times greater, but importantly, measurement precision for the second-order factor only diminished at the extreme ends of the well-being continuum (beyond +2.0, −2.0 SD) and there were no ceiling or floor effects.

Our second-order measurement model provides some support to claims that there could be less than six dimensions under-pinning psychological well-being. The items included within the second-order factor cover aspects of goal orientation and self-direction which could be attributed to a more motivational dimension of well-being. The resulting structure of a motivational dimension of well-being (covered by our second-order factor) together with the two independent factors of autonomy and positive relations is in many ways analogous to the work of Deci and Ryan (1985, 2000) which proposes that well-being results from the fulfillment of three basic psychological needs – autonomy, relatedness and competence. However, it should be emphasised that our interpretation of the PWB structure should be seen in terms of the hierarchical organisation of the six-factors, which span two conceptual levels and not three independent factors.

The NSHD sample used for this analysis was homogeneous for sex (all women) and age (52 years). This sample represents the surviving members of this general population cohort study of health and development who have been studied from birth and also completed postal questionnaires in midlife (Wadsworth et al. 2006). Comparable data were not available for male cohort members since the PWB scales were administered as part of a study of women's health around the menopause.

4.1 Recommendations

In summary, our findings suggest that for more reliable measurement of 'high well-being' the PWB requires the identification and addition of questions which tap into the more positive end of the well-being continuum. We propose that the second-order factor offers a potentially more reliable measure for capturing variations in well-being across the continuum than derived from the component subscales. Further work on the validity of the second-order measure is now warranted, particularly across other population samples which include men, and also different age groups. Finally, more in-depth theoretical work on the underlying dimensional structure of psychological well-being is required, particularly in light of the similarities identified between the second-order PWB factor model and Deci & Ryan's three dimensional well-being model.

5 Conclusion

In light of the growing interest in the measurement of well-being amongst researchers, practitioners and policy makers (Dolan and White 2007; Huppert et al. 2009; Layard 2005; Marks and Shah 2005) there is a pressing need for scales which can measure well-being effectively across the full spectrum. Our analysis has shown that the subscales of the Ryff's six dimensional PWB, adequately measure average levels of well-being, but have low precision of measurement at high levels. Whilst we support the use of the second-order factor as a general measure of well-being, we recommend that future well-being scales should be designed to include items that discriminate more reliably at high levels along the well-being continuum.

References

Abbott, R. A., Ploubidis, G. B., Huppert, F. A., Kuh, D., Wadsworth, M. E., & Croudace, T. J. (2006). Psychometric evaluation and predictive validity of Ryff's psychological well-being items in a UK birth cohort sample of women. *Health and quality of life outcomes, 4*(76).

Baker, F. (2001). *The basics of item response theory*. College Park: University of Maryland: ERIC Clearinghouse on Assessment & Evaluation.

Baker, F., & Kim, S. (2004). *Item response theory: parameter estimation techniques* (2nd ed.). New York: Taylor & Francis.

Bartholomew, D. J., & Knott, M. (1999). *Latent variable models and factor analysis*. London: Arnold Publishers.

Brim, O., Ryff, C. D., & Kessler, R. C. (Eds.). (2004). *How healthy are we?*. Chicago: The University of Chicago Press.

Burns, R. A., & Machin, M. A. (2009). Investigating the structural validity of ryff's psychological well-being scales across two samples. *Social indicators research*, (in press).

Carr, D. (1997). The fulfillment of career dreams at midlife. Does it matter for women's mental health? *Journal of Health and Social Behaviour, 38*(33), 1–344.

Cheng, S. T., & Chan, A. C. (2005). Measuring psychological well-being in the Chinese. *Personality and Individual Differences, 38*, 1307–1316.

Clarke, P. J., Marshall, V. W., Ryff, C. D., & Wheaton, B. (2001). Measuring psychological well-being in the Canadian study of health and aging. *International Psychogeriatrics, 13*(Suppl. 1), 79–90.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum.

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11*, 227–268.

Dolan, P., & White, M. (2007). How can measures of subjective well-being be used to inform public policy? *Perspectives on Psychological Science, 2*(1), 71–85.

ELSA. English longitudinal study of ageing. From www.ifs.org.uk/elsa.

Fava, G. A., Ruini, C., Rafanelli, C., Finos, L., Salmaso, L., Mangelli, L., et al. (2005). Well-being therapy of generalized anxiety disorder. *Psychotherapy and Psychosomatics, 74*(1), 26–30.

Huppert, F. A., Keverne, B., & Bayliss, N. (Eds.). (2005). *The science of well-being*. Oxford: Oxford University Press.

Huppert, F. A., Marks, N., Clark, A., Siegrist, J., Stutzer, A., Vittersø, J., et al. (2009). Measuring well-being across Europe: Description of the ESS well-being module and preliminary findings. *Social Indicators Research, 91*(3), 301–315.

Jöreskog, K. G. (1990). New developments in LISREL: Analysis of ordinal variables using polychoric correlations and weighted least squares. *Quality & Quantity, 24*, 387–404.

Jöreskog, K. G. & Sörbom, D. (2004). LISREL 8.7. Lincolnwood, IL: Scientific Software International, Inc.

Kafka, G. J., & Kozma, A. (2002). The construct validity of Ryff's scales of psychological well-being (SPWB) and their relationship to measures of subjective well-being. *Social Indicators Research, 57*, 171–190.

Kahneman, D., Diener, E., & Schwarz, N. (Eds.). (1999). *Well-being: the foundations of hedonic psychology*. New York: Russell Sage Foundation.

Layard, R. (2005). *Happiness: lessons from a new Science*. London: Penguin.

Lindfors, P., Berntsson, L., & Lundberg, U. (2006). Factor structure of Ryff's psychological well-being scales in Swedish female and male white-collar workers. *Personality and Individual Differences, 40*(6), 1213–1222.

Marks, N., & Shah, H. (2005). A well-being manifesto for a flourishing society. In F. A. Huppert, N. Bayliss, & B. Keverne (Eds.), *The science of well-being*. Oxford: Oxford University Press.

McDonald, R. P. (1999). *Test theory: a unified treatment*. Manwah: Lawrence Erlbaum Associates.

McGregor, I., & Little, B. R. (1998). Personal projects, happiness and meaning: On doing well and being yourself. *Journal of Personality and Social Psychology, 74*, 494–512.

McKinley, N. M. (1999). Women and objectified body consciousness: Mothers' and daughters' body experience in cultural, developmental, and familial context. *Developmental Psychology, 35*(3), 760–769.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika, 49*, 115–132.

Muthén, L., & Muthén, B. (1998–2004). *Mplus user guide.* (Third ed.). Los Angeles, CA: Muthén & Muthén.

Ryff, C. D. (1989a). Beyond Ponce de Leon and life satisfaction: New directions in quest of successful aging. *International Journal of Behavioural Development, 12*, 35–55.

Ryff, C. D. (1989b). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology, 57*, 1069–1081.

Ryff, C. D., Dienberg Love, G., Urry, H. L., Muller, D., Rosenkranz, M. A., Friedman, E. M., et al. (2006). Psychological well-being and ill-being: Do they have distinct or mirrored biological correlates? *Psychotherapy and Psychosomatics, 75*, 85–95.

Ryff, C. D., & Keyes, C. L. (1995). The structure of psychological well-being revisited. *Journal of Personality and Social Psychology, 69*(4), 719–727.

Seligman, M. (2002). Authentic happiness: Nicholas Brealey.

Snyder, C. R., & Lopez, S. J. (Eds.). (2001). *Handbook of positive psychology*. USA: Oxford University Press, Inc.

Springer, K. W., & Hauser, R. M. (2006). An assessment of the construct validity of Ryff's scales of psychological well-being: Method, mode and measurement effects. *Social Science Research, 35*, 1079–1101.

Sweet, J. A., & Bumpass, L. L. (1996). The national survey of families and households—Waves 1 and 2: Data description and documentation. From http://www.ssc.wisc.edu/nsfh/home.htm.

Takane, Y., & de Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika, 52*, 393–408.

Van Dierendonck, D. (2004). The construct validity of Ryff's scales of psychological well-being and it's extension with spiritual well-being. *Personality and Individual Differences, 36*, 629–643.

van Dierendonck, D., Dario Dı′az, D., Rodrı′guez-Carvajal, R., Blanco, A., & Moreno-Jimenez, B. (2008). Ryff's six-factor model of psychological well-being, a Spanish exploration. *Social Indicators Research, 87*(3), 473–479.

Wadsworth, M. E. (1991). *The imprint of time: childhood, history, and adult life*. Oxford: Clarendon Press.

Wadsworth, M. E., Butterworth, S. L., Hardy, R. J., Kuh, D., Richards, M., Langenberg, C., et al. (2003). The life course prospective design: An example of benefits and problems associated with study longevity. *Social Science and Medicine, 57*(11), 2193–2205.

Wadsworth, M. E., Kuh, D., Richards, M., & Hardy, R. J. (2006). Cohort profile: The 1946 national birth cohort (MRC National Survey of Health and Development). *International Journal of Epidemiology, 31*(1), 49–54.