The scales of general well-being (SGWB)

Ylenio Longo

University of Nottingham

Iain Coyne

Loughborough University

Stephen Joseph

University of Nottingham

Author note

Ylenio Longo, Division of Psychiatry and Applied Psychology, University of Nottingham; Iain Coyne, School of Business and Economics, Loughborough University; Stephen Joseph, School of Education, University of Nottingham.

This research was supported by the Economic and Social Research Council grant ES/J500100/1 to Ylenio Longo. We would like to thank Nick Manning, Tom Farsides, Stavroula Leka, Guy Curtis, Petter Gustavsson, Veljko Jovanović, Ralph van den Bosch, Alexander Gunz, and Jessica Morgan, for their insight and support, although any errors are our own. Portions of this article were presented in Ylenio Longo's PhD Thesis.

Correspondence concerning this article should be addressed to Ylenio Longo, School of Medicine, Division of Psychiatry and Applied Psychology, University of Nottingham, YANG Fujia Building, Jubilee Campus, Nottingham NG8 1BB. E-mail: yleniolongo@gmail.com

Longo, Y, Coyne, I., & Joseph, S. (2017). The Scales of General Well-Being (SGWB). *Personality and Individual Differences*.

Abstract

This paper presents the development and validation of a new well-being questionnaire: the Scales of General Well-Being (SGWB). A review of current measures identified fourteen common constructs as lower-order indicators of well-being: happiness, vitality, calmness, optimism, involvement, self-awareness, self-acceptance, self-worth, competence, development, purpose, significance, self-congruence and connection. Three studies were then conducted. In study 1, the item pool was developed and the adequacy of its content to assess each of the fourteen constructs was evaluated by consulting a panel of six subject expert academics. In study 2, the dimensionality was assessed in an adult North American sample (N = 560). The results supported the hierarchical factor structure. In study 3, further evidence confirmed the factor structure, and provided support for the measure's internal and test-retest reliability, measurement invariance across gender, age and a longitudinal period of 5 weeks, and criterion validity in an adult North American sample (N = 1,101). The SGWB promises to be a useful research tool that provides both a global measure of well-being as well as a collection of fourteen individual health-related scales.

Keywords: scale development; scale validation; bifactor; higher-order; well-being; measurement; quality of life.

This paper aims to develop and validate a new well-being questionnaire: the Scales of General Well-Being (SGWB). The hypothesized factor structure of the questionnaire includes a single higher-order or general factor of well-being as well as several specific indicators included in current measures of well-being. The literature has often defined at least two general factors of well-being, labeled subjective well-being (SWB) and psychological well-being (PWB). More generally, the two have also been respectively labeled positive feeling and positive functioning, hedonia and eudaimonia, or pleasure and personal fulfillment. In the past, these have usually been modeled as distinct concepts. Recently, however, several studies have shown that elements of positive feeling and positive functioning are adequately explained by a single general factor. The adequacy of a single general factor of well-being has been tested using higher-order and bifactor models.

For example, Huppert and So (2013) proposed a measurement model of well-being with ten single-item indicators loading on two factors. However, a recent study showed that the two factors were highly correlated (r = .76), and after including a general factor, thus specifying a bifactor model, the two specific factors were unreliable (Longo, Coyne, Joseph & Gustavsson, 2016).

Keyes' (2002) proposed a model with fourteen lower-order factors loading on three higher-order factors: emotional, psychological and social well-being. However, after testing this model in a student sample (N = 591) and a large representative U.S. sample (N = 4,032), a single higher-order factor was found to adequately explain the variance in the lower-order factors (Gallagher et al., 2009). A higher-order model was also found to be adequate when focusing only on emotional well-being (i.e. positive feeling) and psychological well-being (i.e. positive functioning) components in a large international sample (*N* = 7,617) (Disabato, Goodman, Kashdan, Short, & Jarden, 2016).

Similar results were also found with Seligman's (2011) conceptualization of well-being. Two studies supported a higher-order model in which the lower-order factors, including positive feeling and positive functioning elements, loaded on a single higher-order well-being factor (Coffey, Wray-Lake, Mashek & Branand, 2016).

Additional studies indicated support for a general factor of well-being using a bifactor model (Chen, Jing, Hayes & Lee, 2013; de Bruin & du Plessis, 2015; Jovanović, 2015). In the bifactor model, well-being items showed stronger loadings on the general factor compared to their specific factors, indicating that well-being items reflect primarily a single general factor of well-being.

Taken together, these results suggest that well-being data has a hierarchical structure, where items are influenced by specific factors, as well as a single general factor of well-being, whether directly (in bifactor models) or indirectly (in higher-order models).

If well-being data fit hierarchical models with a single general factor, it is necessary to develop a conceptual notion of well-being as a unidimensional hierarchical construct, by identifying commonalities among the lower-order factors of well-being. Specifically, the new measure is based on commonalities among six conceptualizations of well-being proposed by Diener et al. (2010), Huppert and So (2013), Keyes (2002), Ryan, Huta and Deci, (2008), Seligman (2011), and Waterman et al. (2010), which have been included in recent reviews of the field (Huta, 2013; Kashdan et al., 2008; Keyes, 2015; Ryan & Deci, 2001).

Most of the indicators presented in current conceptualizations and measures share several common properties. First, as noted by Kashdan et al. (2008), with few exceptions, these constructs consist of the same experiential phenomena: subjective feelings and evaluations. Consistent with this notion, general well-being has been recently defined as “the experience of life going well” (Huppert & So, 2013, p. 838). Only 4 of the 49 constructs presented in the six conceptualizations do not share this commonality. Autonomous behavior, intrinsic goals, effort, and emotional resilience are not defined as subjective experiences or evaluations. Instead, intrinsic goals are defined as pursuits (Ryan et al., 2008); autonomous behavior (Ryan et al., 2008) and the “investment of effort in the pursuit of excellence” (Waterman et al., 2010, p. 45) are defined as behaviors; and emotional resilience is defined as the “ability to manage anxiety and worry” (Huppert & So, 2013 p. 856). Beyond these exceptions, the first main point of convergence for well-being constructs is their subjective experiential nature.

Second, each of these subjective experiences is interpreted only as an indicator of good mental health. Well-being is not defined by any single indicator, but rather by a collection of indicators (e.g. Huppert & So, 2013; Keyes, 2013; Seligman, 2011). Just like each item loading on a factor is an imperfect measure of that factor, each well-being indicator, or lower-order factor, is merely an imperfect indicator of general well-being. A high score on a single lower-order factor does not guarantee high well-being. Therefore, multiple indicators are taken into consideration to provide a comprehensive account of well-being.

Third, these indicators are relatively stable, with test-retest reliabilities approximately ranging from .60 to .85 over a period of 4 to 8 weeks (e.g. Pavot & Diener, 2008; Ryff, 1989; Watson, Clark & Tellegen, 1988).

Fourth, in all but one case, well-being indicators involve the *presence* of a healthy experience (e.g. positive affect). Only one construct included in one conceptualization involves the *absence* of an unhealthy experience: negative affect. Negative affect strongly overlaps with psychopathology constructs, such as depression and anxiety, with correlations as high as .77 (e.g. Brown, Chorpita & Barlow, 1998; Watson et al., 1988). Consistent with the majority of conceptualizations reviewed here, the current conceptualization focuses on the presence of healthy experiences.

1.1. Summary

A single general factor of well-being has been found to adequately explain the variance in well-being data. Based on similarities among well-being indicators presented in current models, the general factor was conceptualized as a collection of relatively stable subjective feelings and evaluations, which represent symptoms of good mental health.

In study 1, the aim was to identify the common constructs in the well-being literature to guide scale development. Items were then developed based on construct definitions. Content evidence of validity was then assessed by asking a group of experts to evaluate the correspondence between the items and the definition of the construct they attempt to measure.

In study 2, the questionnaire was administered to a North American sample. Response distributions, item clarity, dimensionality and reliability of the responses were assessed. Based on these findings, well-performing items were selected.

Study 3 replicated these findings in a new sample. The study also assessed measurement invariance and criterion evidence of validity by correlating the questionnaire scores with other measures of conceptually similar and different constructs.

Study 1: Item development and content validation

Researchers have included a variety of constructs to measure well-being. In this case, the choice of constructs was informed by the conceptual definition of general well-being presented above, apparent commonalities in the conceptual definitions of the lower-order constructs, as well as a review of the psychometric properties and limitations of previous measures of these constructs (e.g. Chen et al., 2013; Springer & Hauser, 2006). A summary of the following fourteen constructs and their conceptual relationships with constructs presented in other well-being models is presented in the supplementary material.

Happiness consists of moderate-arousal pleasant feelings, such as feeling happy, cheerful and pleased. *Vitality* consists of high-arousal pleasant feelings, such as feeling energetic and lively. *Calmness* consists of low-arousal pleasant feelings, like serenity and peacefulness. *Optimism* is defined as a positive outlook on and expectations about the future. *Involvement* describes the flow state: an absorbing experience in which the individual is completely focused on the task at hand. *Self-awareness* consists in knowing oneself and experiencing a state of mindful awareness. *Self-acceptance* consists in experiencing different aspects of oneself (e.g. one's past, personality, thoughts, and feelings) in a tolerant, receptive and non-judgmental way. *Self-worth* consists in positive evaluations and feelings about oneself. *Competence* consists of feeling and perceiving oneself as effective and able to overcome challenges and achieve desired outcomes. *Development* consists in experiencing continuous growth and improvement. *Purpose* consists in having clear goals, a sense of direction and a larger aim in life. *Significance* is the feeling that what we do is worthwhile, rewarding and valuable. *Self-congruence* is the perception that our actions are compatible with our interests, values, and beliefs. *Connection* involves a feeling of belonging, mutual caring, love, and closeness.

2.1. Item development

The item development process was adapted from DeVellis (2012). Items were written based on the conceptual definitions of the fourteen well-being constructs presented previously.

The main goals here were to generate items that only tapped into the single construct under focus, and were as clear and concise as possible. To accomplish this, the first items generated closely resembled the definition of the construct they were designed to measure (e.g. “I feel happy” for happiness). Then, previous scales attempting to measure each construct, or closely related constructs, were examined and key terms were extrapolated from these scales to develop new items. Additionally, a thesaurus (Oxford Dictionaries, 2014) was used to identify related terms describing each construct, while omitting terms that were too close in meaning to avoid any redundancy (Cattell, 1978). In few cases, closely related items (e.g. energetic and full of energy) were included in the initial item pool, so that the one exhibiting better psychometric properties may later be retained (DeVellis, 2012).

The initial item pool contained a minimum of 10 items for each construct (see Table 1). The aim for the final scale was to include approximately 5 items to measure each construct, with 4 items as a minimum. This target was chosen so that the final scales would minimize participant burden while maintaining adequate internal reliability. Whenever a single construct (e.g. self-awareness) comprised several facets (e.g. awareness of one's thoughts, feelings, bodily sensations, etc.), a grid was used to ensure at least two items were generated to tap into each facet (Kline, 2000).

A 5-point response format was used: 1 = not at all true, 2 = a bit true, 3 = somewhat true, 4 = mostly true, 5 = very true (adapted from Liao, 2014; Weinstein, Przybylski & Ryan, 2012). Well-being data tends to be negatively skewed: most people report average to high levels of well-being (e.g. Springer & Hauser, 2006). Consequently, a scale that provides more options covering moderate to high levels of well-being has been recommended (Saris & Gallhofer, 2014) and has been found to produce data with a near-normal distribution (Liao, 2014). The final draft is presented in the supplementary material.

All items were phrased as declarative statements. Theoretically, reversed well-being items were expected to tap into ill-being constructs (Huppert & So, 2013). From a psychometric perspective, reversed items can be confusing to respondents, particularly in long questionnaires (DeVellis, 2012; van Sonderen et al., 2013); they can produce factor structure problems (Schmitt & Stults, 1985; Woods, 2006); they were less effective in measuring well-being constructs like flow (Jackson & Eklund, 2004) and vitality (Bostic, Rubio & Hood, 2000); and, while they are normally included to prevent acquiescent responding, they have been found to be ineffective in doing so (van Sonderen et al., 2013).

2.2. Content validation

Six academics, each of whom had published several articles on well-being research, were contacted through their university email address and asked to evaluate whether the items were clear and relevant to the construct they were designed to measure. For each construct, the experts were presented with a definition, followed by a list of items. Then, they assessed whether each item was a good measure of the construct on a Likert response format (1 = inadequate, 2 = needs revision, 3 = adequate). Additionally, they were asked to provide comments if they thought any item was inadequate or needed revision. Based on these recommendations, of the 147 items submitted for content validation, 18 were excluded from further analyses and 6 were edited.

Items were generally retained if they had a mean rating of at least 2.30 / 3, which corresponds approximately to a rating of 7.5 / 10 (Hardesty & Bearden, 2004). The comments were also taken into consideration. For instance, the development item "I feel I am growing" had an adequate rating, but one reviewer argued that it may be interpreted too literally. Indeed, when presented on its own, the item may be misinterpreted as physical growth. Therefore, it was excluded from further analyses.

Finally, the academics were asked whether all aspects of each construct were represented in the item pool. Nearly all agreed that they were. One exception mentioned that the competence scale may also tap into constructs like self-efficacy and resilience. Indeed, the chosen definition of competence closely resembles that of self-efficacy and resilience, although it focuses more on one's perception of confidence (e.g. Deci & Ryan, 2000). Therefore, this was not deemed problematic, especially after considering that all competence items received high ratings. After content validation, each scale included 8 or more items and the total item pool comprised 129 items (Table 1).

Table 1 Number of items per scale at different stages of scale development

Scales	Item development	Content validation	CFA (study 2)
Happiness	10	8	4
Vitality	11	8	4
Calmness	10	8	4
Optimism	10	9	5
Involvement	10	9	4
Self-awareness	10	9	4
Self-acceptance	12	11	4
Self-worth	10	8	5
Competence	10	10	6
Development	10	9	5
Purpose	10	9	5
Significance	10	8	5
Congruence	12	11	5
Connection	12	12	5
Total	147	129	65

Study 2: Item selection and scale dimensionality

Study 2 examined the factor structure of the questionnaire and the extent that selected items exhibited desirable psychometric properties. Each of the fourteen scales was expected to fit a unidimensional model. The 14 scales together were expected to fit a model with 14 correlated factors as well as higher-order and bifactor models with a single higher-order or general factor explaining most of the variance.

3.1. Method

3.1.1. Sample. Data were collected from the Amazon Mechanical Turk, whose U.S. samples have been found to be more heterogeneous than student samples (Buhrmester, Kwang, & Gosling, 2011), with demographics comparable to a large U.S. stratified survey (Huff & Tingley, 2015). In total, 573 responses from U.S. residents were recorded. Since few responses were unfinished, the final sample included 560 responses. Ages ranged from 19 to 77 years (M = 37.18, SD = 13.10), and 60.9% were female. A small monetary incentive was used to encourage participation (approx. $0.50).

To ensure high quality of the data, only people whose previous work on MTurk had been rated as adequate at least 95% of the time were able to access the questionnaire (Peer, Vosgerau & Acquisti, 2014). Additionally, a screening question was included toward the end of the survey: "I read instructions carefully. To show that you are reading these instructions, please leave this question blank." (Maniaci & Rogge, 2014). Around 8% of participants failed this screening question and were therefore omitted from further analyses. Finally, seven cases were omitted because they provided the same response to all items (Cummins, 2013). The final sample included 507 participants.

3.1.2. Procedure. Items were grouped based on the construct they measured, so that each construct would be measured in a single page of the questionnaire. The grouped format was chosen because it has been shown to improve convergent and discriminant validity (Harrison & McLaughlin, 1996) and participants find it easier to understand the meaning of a construct when its items are presented in a grouped format (De Vaus, 2002). Within each page, item order was randomized to minimize the effect of question order (De Vaus, 2002; Saris & Gallhofer, 2014).

3.2. Results

3.2.1. Preliminary analyses. Overall, a very small total amount of missing values was found (below 1%). The highest amount of missing values for a single item was less than 1%. Additionally, 84% of cases were complete listwise and, 98% of cases were complete pairwise. No problematic nonresponse patterns were identified, and missing values were estimated using full information maximum likelihood (Graham, 2009).

Items were generally found to be clear, with no item reported as confusing by 2% or more participants.

Response distributions were then examined using the psych package version 1.5.8 (Revelle, 2015) in R version 3.2.3 (R Core Team, 2015). All items produced responses covering the entire range of the 5-point response format. On average, means approximated the middle value of the 5-point format (i.e. 3), although there was a tendency to report slightly higher values (M = 3.62, range = 2.85 – 4.24). Standard deviations indicated sufficient variability in the responses (mean SD = 1.11, range = 0.88 – 1.30). Absolute values of univariate skewness (M = -.61, range = -1.38 and 0.02) and kurtosis (M = -.29, range = -1.04 and 1.55) were relatively low, compared to the absolute cutoff points of 2 and 7, respectively (Fabrigar & Wegener, 2012). Mardia's multivariate skewness and kurtosis values were above five, indicating a significant deviation from multivariate normality (Bentler, 2006). Therefore, robust maximum likelihood (MLR) estimation was used during factor analyses.

3.2.2. Dimensionality of each scale. Analyses were conducted using the lavaan package version .5–20 (Rosseel, 2012) and the semTools package version .4-12 (semTools contributors, 2016). First, the dimensionality of each scale was assessed. Since each scale was based on a conceptual framework and previous measures of the same construct, a confirmatory factor analysis (CFA) was conducted (Fabrigar & Wegener, 2012).

Items were selected iteratively based on a combination of statistical and theoretical considerations. From a statistical perspective, items exhibiting the highest modification indices, expected parameter change and residual correlations were candidates for exclusion. From a theoretical and psychometric perspective, when choosing between two items, the simpler, clearer, less redundant item, which better reflected the definition of the construct was selected (DeVellis, 2012). For example, a residual correlation between the optimism items "I'm optimistic" and "I look on the bright side" significantly reduced model fit, indicating that one of them should be omitted from the scale. In this case, the item "I look on the bright side" was omitted, because it was lengthier: it conveyed the same idea as "I'm optimistic" but with greater complexity. Thus, items were selected until each scale exhibited good fit and included items covering the entire content range of each construct's definition.

Several fit indices were used to evaluate the model fit of the final scales: the Yuan-Bentler chi-square (YB χ^2), the Comparative Fit Index (CFI), the Standardized

Root Mean Square Residual (SRMR), and the Root Mean Square Error of Approximation (RMSEA). CFI and RMSEA were based on the Yuan-Bentler chi-square. A non-significant chi-square indicates excellent fit, which becomes rarer to achieve as sample size and model complexity increase; CFI values above .90 indicate acceptable fit and values above .95 indicate good fit; SRMR values below .06 indicate good fit; RMSEA values below .08 indicate acceptable fit and values below .06 indicate good fit, and its upper confidence interval should not be above .10 (Brown, 2015; Kline, 2016). However, RMSEA was found to perform poorly in models with few degrees of freedom, as it over-rejects well-fitting models (Kenny, Kaniskan & McCoach, 2015). Therefore, RMSEA was not taken into consideration when each individual scale was factor analyzed. In addition to fit indices, correlation residuals were examined to identify specific areas of misfit and these should not exceed a value of .10 (Kline, 2016).

Before item selection, fit indices showed generally acceptable or marginally acceptable fit. After item selection, fit substantially improved (see Table 2). All indices showed good fit and residual correlations never exceeded .10. Interestingly, despite the large sample size, χ^2 was not significant. Such good fit may be due to several factors: the initial item pool was derived from previously validated measures, the items were then subjected to content validation and further refined, well-fitting items were then selected during CFA, and the final models were relatively simple, as they included only 4 to 6 items.

Table 2 Fit indices of individual scales after item selection (Study 2)

Scale	YB χ^2	*df*	*p*	CFI	SRMR
Happiness	4.58	2	.101	.998	.006
Vitality	0.22	2	.897	1.00	.002
Calmness	1.89	2	.389	1.00	.004
Optimism	7.98	5	.157	.998	.008
Involvement	0.03	2	.983	1.00	.001
Self-awareness	1.28	2	.528	1.00	.008
Self-acceptance	0.37	2	.830	1.00	.004
Self-worth	2.77	5	.735	1.00	.004
Competence	6.31	9	.708	1.00	.007
Development	3.17	5	.674	1.00	.006
Purpose	3.39	5	.641	1.00	.006
Significance	2.37	5	.796	1.00	.004
Congruence	1.25	5	.941	1.00	.005
Connection	6.19	5	.288	.999	.009

3.2.3. Dimensionality of the entire questionnaire. After selecting items in each scale, a CFA was conducted on the entire well-being questionnaire. The final questionnaire is presented in the supplementary material.

A variety of models were tested, including alternatives to the hypothesized correlated 14-factor model (see Table 3). In the first model, all items loaded on a single factor. If this model exhibited good fit, it would indicate that different scales have no discriminant validity from each other. The model exhibited poor fit and high residual correlations. This suggests that the fourteen factors are likely distinguishable.

The second model specified 14 independent (i.e. uncorrelated) factors: all items loaded on their hypothesized factor, but all factor correlations were fixed to 0. If the model adequately fit the data, it would indicate that the 14 scales are independent of each other. This would suggest that, unlike previous well-being measures, the scales do not measure a common construct, which would threaten the validity of the questionnaire. However, the model showed poor fit and very high residual correlations, which would particularly affect the SRMR (see Table 3). This suggests that the factors are likely not independent.

The third model was the same as the second model, but the factors were correlated. Model fit was satisfactory, thus supporting the hypothesized factor structure. Considering the complexity of the model, it was highly unlikely for the χ^2 to be non-significant. Nevertheless, all other fit indices indicated good fit. Furthermore, residual correlations were very low, the highest being .14. This was slightly higher than .10, but considering the large number of items and factors it was not deemed problematic. In short, the hypothesized model, with 14 correlated factors, fit the data best.

Table 3 Fit indices of the entire questionnaire (Study 2)

Model	YB χ^2	*df*	*p*	CFI	SRMR	RMSEA (90% CI)	BIC
One factor	13,198.05	2,015	.000	.581	.081	.105 (.103-.106)	82,156
14 independent factors	6,819.42	2,015	.000	.820	.419	.069 (.067-.070)	74,361
14 correlated factors	2,751.51	1,924	.000	.969	.033	.029 (.027-.031)	70,053
Higher-order	3,218.26	2,001	.000	.954	.051	.035 (.033-.037)	70,141
Bifactor	3,005.88	1,950	.000	.960	.044	.033 (.031-.035)	70,214

Factor loadings were strong (M = .86, range = .61 - .94; see also supplementary material). All fourteen scales exhibited adequate reliability, with McDonald's omega hierarchical (ωh) coefficients ranging from .81 to .91 (see Table 4). McDonald's ωh was used to measure reliability because this was found to outperform Cronbach's α (Zinbarg, Revelle, Yovel & Li, 2005). Factor correlations were moderate to strong and all of them were positive, indicating that the covariance among the 14 factors may be explained by a higher-order or general factor.

To further assess the dimensionality of the scale, the average variance explained (AVE) by each factor was compared with the variance shared between factors (see Table 4). AVE corresponds to the average factor loading squared for each scale, while variance shared corresponds to the squared correlations between factors. AVE should ideally be .50 or greater, indicating that 50% or more of the variance in the items is reliable variance captured by the factor (Fornell & Larker, 1981; Hair et al., 2010), thus supporting convergent validity. Indeed, this was the case for every scale. The variance explained by each factor should also be greater than the variance shared with other factors. In other words, the AVE of a factor should exceed the squared correlations between that factor and all other factors (Fornell & Larker, 1981; Hair et al., 2010), which would indicate that different factors are not measuring the same construct, thus supporting discriminant validity. This was true in all cases except one: the self-acceptance factor explained 63% of the variance in its items, but shared 66% of variance with the self-worth factor.

Table 4 Reliability and discriminant validity analysis (Study 2)

	Hap	Vit	Cal	Opt	Inv	Aw	Acc	Wor	Cop	Dev	Pur	Sig	Cog	Con
ωh	.90	.89	.90	.88	.89	.81	.84	.90	.88	.88	.89	.91	.85	.87
AVE	.81	.77	.81	.77	.74	.56	.63	.81	.74	.76	.75	.81	.66	.71
MVS	.61	.61	.52	.60	.36	.39	.66	.66	.50	.45	.56	.56	.40	.39

Note: ωh = McDonald's omega hierarchical; AVE = average variance explained by the factor; MVS = maximum variance shared with other factors

An alternative to Fornell and Larker's method, is the recently developed heterotrait-monotrait method (HTMT; Henseler, Ringle & Sarstedt, 2015). HTMT values show the ratio of the inter-item correlations between scales to the inter-item correlations within scales. HTMT values are interpreted in the same way as factor correlations, with values below .85 providing strong evidence of discriminant validity (Henseler et al., 2015). In this sample, the highest HTMT value was .80 between self-acceptance and self-worth, supporting the discriminant validity of all scales.

Since the Fornell and Larker method showed an overlap between self-acceptance and self-worth, a further analysis was carried out to investigate whether the two factors could be merged together. A one-factor model of self-acceptance and self-worth fit the data poorly compared to a two-factor model, which showed good fit (see Table 5). Therefore, despite the overlap between self-acceptance and self-worth they were maintained separate.

Table 5 Competing models for self-worth and self-acceptance (Study 2)

Model (no. factors)	YB χ^2	*df*	*p*	CFI	SRMR	RMSEA (90% CI)	BIC
Worth-acceptance (one)	239.01	27	.000	.919	.066	.124 (.113-.136)	10,579
Worth-acceptance (two)	32.00	26	.193	.998	.017	.021 (.000-.039)	10,268

Two hierarchical models were then tested. Firstly, a higher-order model was similar to the independent factors model, yet with the addition of a single higher-order factor on which the fourteen lower-order factors loaded. Model fit was slightly worse compared to the correlated factors model, as evidenced by a higher BIC (see Table 3). However, this was expected, as fewer parameters were estimated, and more parsimonious models tend to have lower fit indices. Indeed, 91 factor correlations were now explained by the more parsimonious higher-order loadings. Despite this, the higher-order model fit the data adequately. Loadings on the higher-order factor were high (*M* = .76, range = .57 - .88). Additionally, factor correlations were used to compute McDonald's omega hierarchical (ωh = .86), which supported the reliability of the higher-order factor. The omega analysis also indicated that, after controlling for a single general factor, additional factors had unacceptable reliability estimates, below .20 (see supplementary material).

Secondly, a bifactor model was also similar to the independent factors model, with the addition of a general factor on which all items loaded. Again, this model did not fit as well as the correlated factors model, perhaps due to its increased parsimony. Nevertheless, model fit was adequate. Loadings on the specific factors remained significant, indicating that, even after controlling for the general factor, each scale provided significant additional variance. Furthermore, the general factor saturation was high (ωh = .94), again supporting its reliability.

Two additional indices were recently found to be useful when evaluating bifactor models: explained common variance (ECV) and the percent of uncontaminated correlations (PUC). These indices indicate the extent to which a single score from *g* is unbiased in representing the variance in the items (Rodriguez Reise & Haviland, 2016). ECV represents the proportion of common variance explained by a factor (e.g. *g*). The general factor explained the majority (i.e. 60%) of the common variance in the items (ECV = .60).

The ECV is moderated by the PUC, which represents the percentage of correlations among items in the model that can be attributed solely to the general factor. When PUC is high, an overall *g* score will be unbiased even when ECV is relatively low (e.g. .50) (Bonifay, Reise, Scheines & Meijer, 2015). In the present bifactor model, the PUC was .94, indicating that 94% of the correlations reflected only the general factor.

In short, the higher-order and bifactor models supported the use of a single general factor of well-being.

The results provide preliminary support for the conceptualization of well-being, with fourteen lower-order (or specific) factors and a single general factor. To ensure that the current results are replicable, these findings needed to be supported in a new sample, which was recruited in the next study.

Study 3: Further dimensionality and criterion validity testing

The purpose of this study was to further test the dimensionality of the new well-being questionnaire. It was expected that the factor structure supported in study 2 would also be supported in a new sample. Test-retest reliability and longitudinal invariance were also assessed. Furthermore, since differences in well-being have been found across gender and age (e.g. Meade & Dowswell, 2015), the measurement invariance of the questionnaire across gender and age was assessed.

Additionally, criterion validation was conducted by correlating scale and general well-being scores with external measures. The correlation between each scale and previous measures of the same (or similar) constructs was examined. Strong correlations were expected between each scale of GWB and its respective comparison measures (e.g. between the new and previous measures of happiness or pleasant affect). Comparatively weaker relationships were expected between each scale and measures of different constructs (e.g. between the new happiness scale and a comparison measure of self-awareness).

A general well-being score was expected to be very strongly correlated with scores of general well-being (Keyes, 2013) including elements of positive feeling (i.e. subjective well-being) and positive functioning (i.e. psychological and social well-being). Furthermore, the relationship between general well-being, social desirability and personality traits was assessed. If the new questionnaire taps into the same general construct as previous measures of well-being, it should exhibit similar relationships to external variables. Based on previous studies, well-being was expected to exhibit a moderate correlation with social desirability (Fastame, Penna & Hitchcott, 2015; McCrae & Costa, 1983), moderate to strong correlations with neuroticism and extraversion, moderate correlations with conscientiousness and agreeableness and a small correlation with imagination (Donnellan et al., 2006; Steel, Schmidt & Shultz, 2008). Additionally, general well-being was expected to have a small correlation with dark triad traits: narcissism, psychopathy and Machiavellianism (Aghababaei & Błachnio, 2015).

Finally, based on the conceptual definition of well-being, it was expected that each well-being construct would be negatively correlated to measures of psychological ill-being (e.g. depression, anxiety) (e.g. Watson et al., 1988), and would be perceived as having a high subjective value. Subjective value has been previously studied with regards to well-being constructs, like positive affect (Diener, 2000), and to identify the most relevant among several components of well-being (Hogan et al., 2015). Symptoms of good mental health are assumed to be intrinsically valuable. In other words, they are worth experiencing for their own sake, rather than as means to an end (Seligman, 2011). Therefore, each construct measured in the SGWB was expected to be rated with a high perceived value score.

4.1. Method

4.1.1. Design. The study was divided into 3 sections. In the first section, all participants completed the SGWB. Due to the large number of measures included in the survey, the second section was divided into two parts (2a and 2b), and participants were randomly assigned to one of these two conditions. In section 2a, participants completed comparison measures tapping into the fourteen constructs assessed by the new well-being questionnaire. Additionally, participants completed a measure of social desirable responding and a personality questionnaire. In section 2b, participants completed measures of subjective, psychological and social well-being, a psychological ill-being scale, and a measure of the perceived value of each well-being construct.

Five weeks after completing the survey, all participants were asked to complete the questionnaire again. The goal was to receive responses from approximately half of the sample to assess the test-retest reliability of the questionnaire. This short follow-up also included a measure of the dark triad traits to further assess discriminant validity.

4.1.2. Sample and procedure. Data were collected from Amazon Mechanical Turk. In total, 1,101 responses from U.S. residents were recorded. Ages ranged from 17 to 83 years (*M* = 36.93, *SD* = 12.13), and 56.6% were female. A small monetary incentive was used to encourage participation ($0.77). Once again, only individuals whose previous work had been rated as adequate more than 95% of the time were able to access the survey (Peer et al., 2014). Furthermore, the survey could only be accessed by workers who had not previously participated in study 2.

Screening questions were embedded in the survey. Section 1 included the item “please leave this item blank”, while sections 2a and 2b included the item “I have seventeen fingers on my left hand,” which required a negative response (DeSimone, Harms & DeSimone, 2015). Failing any of these screening questions resulted in exclusion from further analyses. Around 8% of participants failed these screening questions and were therefore omitted from further analyses. Furthermore, 18 cases showed no variation in their responses to the SGWB and were therefore omitted (Cummins, 2013). Finally, 3 participants were excluded because they reported not being fluent in English. The final sample included 989 participants, of which 486 completed section 2a and 503 completed section 2b.

In the follow-up conducted five weeks later, a total of 446 responses were recorded. The time gap between the main study (time 1) and the follow-up (time 2) was consistent with previous studies, whose time gap usually ranged between 4 and 8 weeks (Pavot & Diener, 2008; Ryff, 1989; Watson et al., 1988). Ages ranged from 19 to 83 (M = 39.72, SD = 12.65), and 61% were female. Following the same screening procedure as above, 12 cases failed the attention check and 3 showed no response variation. Thus, the final follow-up sample included 431 cases.

4.1.3. Measures. In section 1, participants completed the SGWB. Each of the fourteen scales was presented on a separate page. Furthermore, scale order and item order were randomized. Omega hierarchical coefficients are presented in Table 8. Further information regarding each scale, the number of items, scoring and omega coefficient can be found in the supplementary material.

In section 2a, participants completed measures tapping into the fourteen constructs assessed by the new well-being questionnaire. Happiness was measured with the pleasant affect scale (Feldman Barrett & Russell, 1998) and the happiness scale (Huppert & So, 2013). Vitality was measured with the subjective vitality scale (Ryan & Frederick, 1997; Bostic et al., 2000) and the energy / general activation scale (Thayer, 1986). Calmness was measured with the pleasant deactivated (Feldman Barrett & Russell, 1998) and the calmness / general deactivation scale (Thayer, 1986). Optimism was measured with the life orientation test-revised (LOT-R, Scheier, Carver & Bridges, 1994) and the optimism scale (Huppert & So, 2013). Involvement was measured with the Core flow scale (Martin & Jackson, 2008) and the total concentration subscale of the Dispositional Flow Scale 2 (DFS2; Jackson & Eklund, 2004). Self-awareness was measured with the awareness subscale of the

Cognitive and Affective Mindfulness Scale-Revised (Feldman et al., 2007) and the awareness subscale of the Philadelphia Mindfulness Scale (Cardaciotto et al., 2008). Self-acceptance was measured with the acceptance subscale of the Cognitive and Affective Mindfulness Scale-Revised (Feldman et al., 2007) and the self-kindness subscale of the self-compassion scale (Neff, 2003). Self-worth was measured with the self-esteem scale (Sheldon et al., 2001) and the Rosenberg self-esteem scale (Rosenberg, 1965). Competence was measured with the competence satisfaction subscale of the balanced measure of psychological needs (Sheldon & Hilpert, 2012) and the new general self-efficacy scale (Chen, Gully & Eden, 2001). Development was measured with the intentional behavior subscale of the Personal Growth Initiative Scale – II (PGIS II; Robitscheck et al., 2012) and the love of learning subscale of the VIA-IS (Peterson & Seligman, 2004). Purpose was measured with the purpose in life test-short form (Schulenberg, Schnetzer & Buchanan, 2011) and the purposeful life subscale of the meaningful life measure (Morgan & Farsides, 2009). Significance was measured with an adapted version[1] of the valued life subscale of the meaningful life measure (Morgan & Farsides, 2009) and the meaning in life scale (Huppert & So, 2013). Self-congruence was measured with the Authorship subscale of the index of autonomous functioning (Weinstein et al., 2012) and the Autonomy satisfaction subscale of the balanced measure of psychological needs (Sheldon & Hilpert, 2012). Connection was measured with the relatedness satisfaction subscale of the balanced measure of psychological needs (Sheldon & Hilpert, 2012) and the relatedness satisfaction subscale of the Need Satisfaction and Frustration Scale adapted to one's life in general (Longo, Gunz, Curtis & Farsides, 2016). Consistent with the SGWB, instructions asked the participants to focus on how they felt in their

[1] Adapted to one's activities rather than one's life.

life in general. Additionally, participants completed a measure of social desirable responding, the short from of the Marlowe-Crowne Social Desirability Scale (MC-A; Reynolds, 1982), and a personality questionnaire, the mini-IPIP measure of the big five personality factors (Donnellan et al., 2006).

In section 2b, subjective well-being (SWB) was measured with the 5-item satisfaction with life scale (Diener et al., 1985) and the 20-item positive and negative affect schedule (PANAS, Watson et al., 1988). Psychological well-being was measured with Ryff's 42-item psychological well-being scales (PWB, Ryff, 1989). Social well-being was measured with Keyes' 15-item social well-being scale (Keyes, 1998). A score of flourishing was then generated by combining scores of SWB, PWB and social-well-being (Keyes, 2002). Before combining scores from questionnaires using different response format lengths (e.g. a 5-point scale and a 7-point scale), they were transformed to z-scores (DeVaus, 2002). Psychological ill-being was measured with the 29-item patient-reported outcomes measurement system (PROMIS-29 v2) (Cella et al., 2010). Finally, to test the perceived value of each well-being construct, participants were asked to what extent they find each construct important in their life (see Table 14).

The follow-up study also included a 12-item concise measure of the dark triad: narcissism, psychopathy and Machiavellianism (Jonason & Webster, 2010). The three traits were found to be explained by a general factor (Jonason, Kaufman, Webster & Geher, 2013). Therefore, an overall dark triad score was used.

To attempt to control for the influence of proximity (Weijters, De Beuckelaer & Baumgartner, 2014) the order of the scales in each section was randomized.

4.2. Results

4.2.1. **Preliminary analyses**. A small amount of missing values was found in each section (less than 1%) and no problematic nonresponse patterns were identified. Consistent with study 2, full information maximum likelihood was used during confirmatory factor analyses. Furthermore, in correlation analyses using scale scores, full-information maximum likelihood correlation matrices were estimated using the psych package. P-values in these correlation analyses used Holm (1979) adjustment for multiple tests.

All items in the SGWB produced responses covering the entire range of the 5-point response format. On average, means approximated the middle value of the 5-point format (i.e. 3), although there was a tendency to report slightly higher values (average *M* = 3.67, range = 2.86 – 4.21). Standard deviations indicated sufficient variability in the responses (average *SD* = 1.08, range = 0.83 – 1.26). Absolute values of univariate skewness (*M* = -.60, range = -1.26 and 0.02) and kurtosis (*M* = -.24, range = -0.85 and 0.95) were relatively low. Mardia's multivariate skewness and kurtosis values were above five, indicating a significant deviation from multivariate normality (Bentler, 2006). Therefore, robust maximum likelihood (MLR) estimation was used during factor analyses.

4.2.2. **Dimensionality of each scale**. The study used the same statistical package, model specifications and fit indices used in the previous study. Overall, results were consistent with the previous study, indicating that items in each scale fit a unidimensional model (see Table 6).

Table 6 Fit indices of individual scales (Study 3)

Scale	YB χ^2	*df*	*p*	CFI	SRMR
Happiness	1.37	2	.504	1.000	.002
Vitality	0.37	2	.833	1.000	.001
Calmness	1.04	2	.596	1.000	.002
Optimism	13.93	5	.016	.996	.008
Involvement	4.89	2	.087	.998	.007
Self-awareness	5.51	2	.064	.996	.010
Self-acceptance	4.40	2	.111	.998	.008
Self-worth	13.16	5	.022	.997	.007
Competence	24.88	9	.003	.995	.011
Development	15.36	5	.009	.996	.008
Purpose	42.27	5	.000	.982	.016
Significance	18.61	5	.002	.994	.010
Congruence	11.05	5	.050	.996	.012
Connection	34.42	5	.000	.986	.017

N = 989

4.2.3. Dimensionality of the entire questionnaire. The dimensionality of the entire questionnaire was also consistent with the previous study (see Table 7). Specifically, the one-factor model and the independent-factors model did not fit the data, but the hypothesized correlated factors model fit the data well. Standardized factor loadings were generally high (*M* = .86, range = .59 - .94, see also supplementary material) indicating that each item adequately contributed to its factor (Hair et al., 2010).

Table 7 Fit indices of the entire questionnaire (Study 3)

Model	YB χ^2	*df*	*p*	CFI	SRMR	RMSEA (90% CI)	BIC
One factor	23,798.32	2015	.000	.578	.081	.105 (.103-.106)	154,240
14 independent factors	10,954.42	2015	.000	.827	.425	.067 (.066-.068)	137,737
14 correlated factors	3,305.95	1924	.000	.973	.032	.027 (.026-.028)	128,792
Higher-order	4,091.09	2001	.000	.959	.049	.032 (.031-.034)	129,253
Bifactor	3,725.51	1950	.000	.966	.042	.030 (.029-.032)	129,148

N = 989

All fourteen scales exhibited adequate reliability, with ωh coefficients ranging from .82 to .92 (see Table 8). The AVE of each factor exceeded .50, supporting internal convergent validity. Squared correlations among factors did not exceed the AVE of each factor, supporting discriminant validity. Furthermore, the highest HTMT value was .78, which was below the .85 cut-off criterion. In other words, both the Fornell and Larker method and the HTMT method supported the discriminant validity of the scale scores.

Table 8 Reliability and discriminant validity analysis (Study 3)

	Hap	Vit	Cal	Opt	Inv	Aw	Acc	Wor	Cop	Dev	Pur	Sig	Cog	Con
ωh	.92	.91	.92	.86	.89	.82	.84	.91	.88	.91	.88	.90	.85	.86
AVE	.85	.80	.86	.75	.78	.59	.64	.84	.71	.79	.74	.80	.62	.68
MVS	.61	.57	.56	.61	.37	.30	.54	.59	.46	.47	.53	.53	.45	.47

Note: ωh = McDonald's omega hierarchical; AVE = average variance explained by the factor; MVS = maximum variance shared with other factors.

Consistent with study 2, hierarchical models showed adequate fit. The higher-order model exhibited a strong general factor with loadings ranging from .53 (self-awareness) to .87 (optimism) (*M* = .76). A single general factor was highly reliable (ωh = .86) and, after controlling for this, additional factors had unacceptable reliability estimates, below .18 (see supplementary material).

In the bifactor model, loadings on the specific factors remained significant, indicating that, after controlling for the general factor, each item in each scale measures additional construct-specific variance. Furthermore, the general factor was reliable (ωh = .94) and explained the majority of the common variance in the items (ECV = .60, PUC = .94). Overall, the dimensionality of the questionnaire was successfully replicated in the new sample.

4.2.4. Invariance across gender and age. Measurement invariance is met when individuals with the same level of the latent variable, belonging to different groups, answer each item in the scale in same way (e.g. Brown, 2015). Multigroup CFA was conducted on the entire sample (*N* = 989) to assess whether the questionnaire was invariant across gender and age groups. Age groups were derived from splitting the continuous age variable into three categories of approximately equal sizes: group 1 (*n* = 353, *M* = 25.86, range = 17-30, *SD* = 3.12), group 2 (*n* = 310, *M* = 35.08, range = 31-40, *SD* = 2.88), and group 3 (*n* = 329, *M* = 53.20, range = 41-83, *SD* = 7.92).

Multigroup CFA involves several steps (e.g. Brown, 2015). The goal was to assess the measurement invariance of both the lower-order and the general well-being factor. Therefore, the higher-order model was used for measurement invariance tests. As a preliminary step, the measurement model was tested separately in each group, and was found to fit the data adequately in all groups (see Table 9). Therefore, it was possible to proceed with invariance testing.

Table 9 CFA in each gender and age group (higher-order model)

Group	YB χ^2	*df*	*p*	CFI	SRMR	RMSEA (90% CI)
Gender						
Male	3080.21	2001	.000	.950	.037	.037 (.034-.039)
Female	3487.40	2001	.000	.954	.036	.036 (.034-.038)
Age						
Group 1	3119.44	2001	.000	.941	.040	.040 (.037-.042)
Group 2	3158.85	2001	.000	.939	.043	.043 (.041-.046)
Group 3	3120.64	2001	.000	.939	.041	.041 (.039-.044)

Invariance testing proceeds through the following three steps. In step 1, the model is tested in all groups at the same time and the questionnaire is expected to fit the same factor structure across different groups (configural invariance). In step 2, the factor loadings are constrained to be equal in both groups. That is, each item is expected to measure its factor equally well in both groups (metric invariance). In step 3, in addition to the previous constraints, intercepts are constrained to be equal in both groups. In other words, for each level of the latent variable, the observed scores in each item are expected to be the same in both groups (scalar invariance). When scalar invariance is achieved, latent scores can be meaningfully compared across groups. Consistent with best practices using higher-order models, each invariance test was conducted sequentially first on the lower- and then higher-order sections of the model (Byrne & Stewart, 2006). Identification was achieved using the reference-group method (Little, Slegers & Card, 2006).

Invariance is supported when there is no significant difference in fit between one step and the next (Hair et al., 2010). The chi-square difference test can be very conservative, and over-reject adequately invariant scales (Meade, Johnson & Braddy, 2008). Alternatively, the CFI has been recommended to identify practically significant difference in fit. Specifically, reductions in CFI should not be greater than .01 (Cheung & Rensvold, 2002) and ideally should not be greater than .002 (Meade et al., 2008).

Results of multigroup CFA show that CFI changes (i.e. ΔCFI) never exceeded the value of .002 (see Table 10). The chi-square difference was sometimes significant. However, based on the ΔCFI these differences are minor and do not bias the comparison of overall well-being scores.

Table 10 Multigroup CFA for gender and age groups (higher-order model)

Model	YB χ^2	*df*	*p*	SB$\Delta\chi^2$	Δ*df*	Δ*p*	CFI	ΔCFI
Gender								
Configural invariance	6567.59	4002	.000	-	-	-	.953	-
Metric invariance (lower)	6627.64	4053	.000	52.76	51	.406	.952	.001
Metric invariance (higher)	6648.01	4066	.000	20.25	13	.089	.952	.000
Scalar invariance (lower)	6760.63	4117	.000	118.00	51	.000	.951	.001
Scalar invariance (higher)	6829.93	4130	.000	78.31	13	.000	.950	.001
Age								
Configural invariance	9400.11	6003	.000	-	-	-	.940	-
Metric invariance (lower)	9532.81	6105	.000	126.11	102	.053	.939	.001
Metric invariance (higher)	9580.11	6131	.000	48.21	26	.005	.939	.000
Scalar invariance (lower)	9758.47	6233	.000	181.38	102	.000	.938	.001
Scalar invariance (higher)	9846.67	6259	.000	95.87	26	.000	.936	.002

N = 989

4.2.5. Longitudinal invariance and test-retest reliability. Next, the study examined whether the questionnaire performed consistently over time via test-retest reliability and longitudinal invariance.

A CFA was conducted by specifying two correlated factors corresponding to the scale at time 1 and time 2. Residuals for each item at time 1 were correlated with residuals of the same item at time 2. Individual scales were analyzed, instead of a larger higher-order model, because of the reduced sample size and increased model complexity from the inclusion of correlated residuals among time 1 and time 2 items. The test-retest correlations based on CFA on each of the fourteen scales ranged from .58 to .86 (*M* = .74), while the general factor was highly stable with a correlation of .88 (see Table 11).

Additionally, the metric and scalar longitudinal invariance of each scale was supported. The scales fit the data adequately both at time 1 and time 2 (see supplementary material). In subsequent steps, reductions in model fit were generally smaller than .002 and, in the only case slightly greater than .002 (i.e. self-awareness), ΔCFI never exceeded .01. To facilitate the interpretation of the results, only CFI values are presented in Table 11. This indicates that changes in responses over time reflect actual changes in the construct rather than measurement bias. Taken together, the results supported the longitudinal invariance of the SGWB.

Table 11 Test-retest reliability and longitudinal invariance

	r (CFA)	CFI Configural	CFI Metric	CFI Scalar
Happiness	.77	.998	.998	.998
Vitality	.75	.999	1.00	.999
Calmness	.68	.999	.999	1.00
Optimism	.86	.991	.991	.991
Involvement	.62	.999	.999	1.00
Awareness	.58	.997	.998	.994
Acceptance	.74	.986	.985	.985
Self-worth	.86	.995	.994	.993
Competence	.76	.989	.989	.988
Development	.65	.998	.998	.998
Purpose	.83	.983	.984	.984
Significance	.75	.989	.989	.989
Congruence	.68	.988	.989	.989
Connection	.79	.995	.995	.994
GWB	.88	.930	.930	.928

n = 431

4.2.6. Criterion validation. Study 3 also examined the correlations between each scale of SGWB, and measures of similar constructs. Each comparison measure was expected to correlate most strongly with its respective new measure. These hypothesized correlations are presented in bold in Table 12.

For example, in the first row, a criterion measure of pleasant affect correlated most strongly with happiness (*r* = .83), and its second highest correlation was with calmness (*r* = .74). This provided some support for the convergent and discriminant validity of the happiness scale. In addition, a stricter test was used: Steiger's test of significant difference between the two dependent correlations was carried out with the psych package. This test compared the highest correlation between criterion and SGWB scale (r12) with the second highest correlation between the criterion and another SGWB scale (r13), while taking into account the correlation between two scales in the SGWB (r23) (Steiger, 1980). Results of this test are presented in the last column of Table 12. In all cases but two, the correlation between each well-being scale and its respective comparison measure was the highest. In most cases, the difference between this high correlation and the second highest for each row was significant. Overall, the hypothesized correlations were supported.

Table 12 Correlations between each WB scale and external criteria

	Hap	Vit	Cal	Opt	Inv	Aw	Acc	Wor	Cop	Dev	Pur	Sig	Cog	Con	*p*
Pleasant affect	**.83**	.67	.74	.72	.49	.34	.54	.71	.52	.56	.53	.61	.54	.65	.000
ESS-happy	**.80**	.66	.67	.72	.46	.34	.56	.74	.57	.56	.55	.63	.54	.63	.002
Subjective vitality	.75	**.82**	.67	.71	.54	.33	.52	.71	.60	.62	.62	.65	.51	.55	.001
Energy	.71	**.85**	.63	.63	.48	.35	.45	.65	.54	.61	.56	.58	.48	.52	.000
Pl. deactivation	.66	.56	**.83**	.57	.42	.31	.52	.59	.43	.46	.37	.48	.43	.48	.000
Calmness	.42	.34	**.59**	.38	.30	.27	.38	.37	.24	.30	.25	.30	.29	.33	.000
LOTR	.65	.55	.54	**.76**	.35	.30	.48	.67	.48	.47	.50	.52	.41	.53	.000
ESS-Optimism	.66	.59	.56	**.75**	.36	.28	.46	.65	.49	.53	.56	.55	.42	.55	.000
Core flow	.53	.55	.51	.51	**.64**	.35	.43	.50	.53	.51	.50	.55	.47	.46	.012
Concentration	.44	.48	.44	.44	**.62**	.36	.40	.45	.50	.40	.42	.41	.55	.41	.030
CAMSR-Aw	.38	.43	.35	.38	.48	**.49**	.45	.43	.40	.39	.41	.39	.44	.40	.697
PHLMS-Aw	.27	.32	.25	.38	.42	**.60**	.37	.31	.41	.34	.38	.33	.40	.38	.000
CAMSR-Ac	.48	.43	.45	.49	.39	.41	**.63**	.53	.47	.35	.38	.41	.44	.45	.000
Self-kindness	.52	.42	.49	.53	.36	.36	**.66**	.61	.40	.39	.38	.42	.40	.42	.060
Sheldon SE	.72	.61	.62	.75	.47	.40	.67	**.87**	.64	.52	.57	.63	.58	.59	.000
Rosenberg SE	.72	.59	.61	.71	.48	.35	.60	**.83**	.62	.50	.54	.62	.56	.59	.000
BMPN-CS	.53	.56	.43	.53	.51	.33	.42	.54	**.75**	.56	.50	.51	.48	.40	.000
NGSE	.64	.59	.53	.70	.49	.37	.48	.68	**.78**	.57	.59	.58	.48	.51	.000
PGIS II-IB	.47	.56	.38	.51	.49	.40	.32	.46	.59	**.63**	.57	.56	.49	.42	.235
VIA-Learn	.22	.22	.16	.22	.26	.29	.18	.25	.41	**.34**	.28	.25	.35	.25	N/A
PILSF	.65	.61	.52	.68	.47	.36	.47	.67	.59	.60	**.77**	.72	.53	.55	.010
MLM-Purpose	.47	.46	.37	.53	.39	.26	.35	.50	.48	.53	**.73**	.58	.37	.41	.000
MLM-Valued	.60	.56	.52	.61	.56	.36	.50	.62	.53	.55	.63	**.71**	.54	.55	.001
ESS-Significant	.65	.56	.53	.68	.47	.30	.47	.67	.49	.55	.63	**.76**	.47	.55	.000
Authorship	.53	.50	.47	.53	.53	.51	.52	.55	.56	.49	.51	.54	**.77**	.53	.000
BMPN-AS	.54	.53	.52	.55	.55	.39	.51	.57	.52	.50	.47	.54	**.55**	.51	N/A
BMPN-RS	.55	.44	.44	.54	.42	.32	.46	.53	.44	.39	.49	.55	.46	**.73**	.000
NSFS-RS	.59	.54	.51	.57	.44	.35	.53	.57	.46	.42	.47	.53	.48	**.82**	.000

Note: Correlations hypothesized to be the highest are in bold; 2nd highest correlations underlined. All correlations were significant ($p < .01$). The names of criterion measures are shortened here but are presented in the same order and their complete names in the Measures section. $n = 486$.

Scale scores were then averaged into a general well-being composite score. The relationship between general well-being and previous measures of well-being was strong and significant ($p < .001$, $n = 503$): .77 with SWB, .81 with PWB, .61 with social well-being, and .86 with flourishing (Keyes, 2013). Additionally, the relationship between general well-being, social desirability and personality traits was explored. As hypothesized, well-being exhibited a moderate correlation with social desirability ($r = .29$), strong correlations with extraversion ($r = .46$) and neuroticism ($r = -.57$), moderate correlations with agreeableness ($r = .32$) and conscientiousness ($r = .37$), and a small correlation with imagination ($r = .16$). All correlations were significant ($p < .001$, $n = 486$). Additionally, a small correlation was found with the general dark triad score (r = -.22, p < .001, n = 431). In other words, the general well-being score was related to external measures as predicted, further supporting its validity as a well-being measure.

To measure psychological ill-being, five subscale scores from the patient-reported outcomes measurement system (PROMIS-29 v2) (Cella et al., 2010) were used. The three remaining PROMIS subscales were omitted because they focused on physical aspects (e.g. physical pain), which were not the focus of the study. Generally, correlations were moderate to strong in magnitude (see Table 13). All correlations were significant (p < .05), with only one exception between self-awareness and sleep disturbance (r = -.07, p = .112). In short, well-being scales related to ill-being scales as predicted, although self-awareness seemed to have more modest correlations to ill-being.

Table 13 Correlations between well-being and ill-being scales

	Hap	Vit	Cal	Opt	Inv	Aw	Acc	Wor	Cop	Dev	Pur	Sig	Cog	Con
Anxiety	-.52	-.44	-.59	-.46	-.38	-.15	-.46	-.52	-.38	-.28	-.30	-.41	-.33	-.34
Depression	-.67	-.55	-.60	-.59	-.45	-.18	-.49	-.64	-.47	-.40	-.43	-.54	-.43	-.47
Fatigue	-.52	-.51	-.51	-.43	-.39	-.16	-.40	-.47	-.34	-.25	-.29	-.35	-.30	-.39
Sleep disturbance	-.39	-.37	-.42	-.34	-.27	-.07	-.32	-.37	-.25	-.18	-.19	-.30	-.21	-.32
Ability to participate	-.45	-.43	-.44	-.41	-.35	-.15	-.34	-.41	-.31	-.27	-.29	-.38	-.30	-.34

n = 503

Construct value. Finally, the perceived value of each well-being construct was assessed. Symptoms of good mental health are assumed to be worth experiencing for their own sake (Seligman, 2011). Therefore, each construct measured in the SGWB was expected to exhibit a high subjective value rating. Specifically, on a 7-point scale, a score above the middle point (i.e. 3.50) was expected. Participants reported mean scores around 6 (Table 14), indicating that every construct included in the SGWB had a high perceived value in this sample, which further supported the inclusion of these constructs in the questionnaire.

Table 14 Perceived value of each well-being construct

Construct	Item	M	95% CI	
Happiness	Feeling happy and cheerful	6.04	5.94	6.14
Vitality	Feeling energetic / full of energy	5.80	5.69	5.91
Calmness	Feeling calm / relaxed	6.02	5.92	6.11
Optimism	Being optimistic and hopeful	6.12	6.02	6.22
Involvement	Feeling completely involved and engaged in what you do	5.89	5.79	5.98
Awareness	Being in touch with how you feel	5.75	5.64	5.86
Acceptance	Accepting yourself the way you are	6.10	6.00	6.21
Self-worth	Liking yourself a lot	5.99	5.88	6.10
Competence	Feeling highly effective at what you do	6.05	5.96	6.15
Development	Feeling you're consistently improving, developing and advancing	6.04	5.94	6.14
Purpose	Having a purpose and a mission in life	6.02	5.91	6.14
Significance	Feeling that what you do is important and worthwhile	6.13	6.03	6.23
Congruence	Feeling that what you do is consistent with how you see yourself	5.89	5.78	6.00
Connection	Feeling close and connected with the people around you	5.94	5.83	6.06

Instructions: rate how important each of the following things are in your life. Do not focus on whether these are currently present in your life. Instead, think whether you would want these things in your ideal life.
Rating scale: (1) no importance whatsoever, (7) extraordinarily important and valuable (Diener, 2000).

General discussion

The SGWB is supported by consistent evidence of dimensionality, measurement invariance, reliability and validity. Not only does it measure many constructs included in a variety of measures of well-being, but it does so while overcoming some of the dimensionality and discriminant validity issues found in previous measures (e.g. Chen et al., 2013; Springer & Hauser, 2006). Based on current evidence, the use of the SGWB can be deemed adequate in adult North American samples drawn from the general population. The present studies also provide evidence supporting the use of hierarchical models with a single general factor to study scores from the SGWB.

5.1. The use of general and specific scores

The SGWB can produce both a general well-being score as well as fourteen specific scale scores. A general well-being score would indeed adequately reflect response variations in the questionnaire. This single score may be intuitively appealing because of its simplicity. However, it may also be beneficial to focus on lower order constructs, as these can be specifically targeted during interventions.

At the same time, individual scale scores are not without limitations. Each lower-order well-being scale is only intended as an imperfect indicator of general well-being. While these factors are often perceived as valued and are related to health outcomes, they may sometimes lead to undesirable consequences. For example, experiencing too little, but also too much, positive affect can undermine people's general well-being (Fredrickson & Losada, 2005). Similarly, optimism has been linked to a lower likelihood to disengage from gambling even after losses (Gibson & Sanbonmatsu, 2004), and lower marital satisfaction when one's partner does not meet one's optimistic expectations (McNulty & Karney, 2004). These findings suggest that any well-being dimension need not be experienced all the time, or at its highest level, to be beneficial. Instead, each construct is likely to contribute to one’s psychological health to some extent and in some contexts.

In short, it may be desirable to study well-being using both a general well-being score and individual scales scores, as both approaches exhibit strengths and limitations. However, further research is needed to explore these strengths and limitations in greater depth.

5.2. Limitations and future directions

The main limitation of the studies is a degree of subjectivity in the choice of the conceptualizations reviewed and the choice of lower-order factors. This limitation is engrained in the process of concept and scale development (Kline, 2000). However, it is somewhat reduced by the fact that the chosen conceptualizations were repeatedly presented in recent reviews of the field (e.g. Kashdan et al., 2008; Keyes, 2015), and by choosing lower-order factors that were repeatedly presented in these previous conceptualizations. Because of this strong overlap among previous conceptualizations, the reader will likely notice that including or omitting any one previous conceptualization would not substantially alter the final choice of lower-order factors. Additionally, the SGWB produced an overall well-being score and lower-order scale scores that related to previous measures in way consistent with their hypothesized relationships, indicating that the questionnaire taps into the same constructs as previous measures.

Based on the sample size and heterogeneity in the present studies, it is expected that the factor structure of well-being identified here will generalize to other adult North American samples. Indeed, samples from the Amazon Mechanical Turk have been shown to be more heterogeneous than student samples (Buhrmester et al., 2011) and to show demographics comparable to large U.S. stratified surveys (Huff & Tingley, 2015). However, the structure of well-being constructs has been shown to vary across cultures or contexts (e.g. Russell, 1980). Therefore, if the structure of well-being varies across cultures, the SGWB may need to be modified to account for this variation.

Criterion validation was carried out with data from other self-report measures. The use of self-report measures was deemed adequate because the constructs under study were *subjective* experiences and evaluations (Spector, 2013). That said, for the purpose of questionnaire validation, future studies could further validate the SGWB by assessing the questionnaire's associations with non-self-report criteria. For example, future studies could investigate the relationships between well-being constructs and biological correlates such as REM sleep duration, daily salivary cortisol and cardiovascular risk (e.g. Ryff, Singer & Love, 2004).

The present studies adopted mainly a cross-sectional design, with a short follow-up study to establish longitudinal invariance. The cross-sectional design was adequate to provide concurrent evidence of criterion validity. However, longitudinal or experimental designs could be used in future studies to allow for causal inferences while exploring long-term outcomes of well-being over several years.

5.3. Conclusion

The paper contributes to the current literature by presenting a psychometrically sound questionnaire measuring general well-being as well as fourteen specific well-being constructs. Overall, questionnaire scores exhibited adequate content validity ratings, factor structure, internal consistency, test-retest reliability, invariance across age, gender and a 5-week period, and relationships with external criteria that were consistent with the hypothesized pattern. While some scales exhibited stronger evidence than others at different stages of the validation process, all scales exhibited adequate psychometric properties across the studies.

References

Aghababaei, N., & Błachnio, A. (2015). Well-being and the Dark Triad. *Personality and Individual Differences, 86*, 365-368.

Bentler, P. M. (2006). *EQS 6 Structural Equations Program Manual*. Encino, CA: Multivariate Software, Inc.

Bonifay, W. E., Reise, S. P., Scheines, R., & Meijer, R. R. (2015). When are multidimensional data unidimensional enough for structural equation modeling? An evaluation of the DETECT multidimensionality index. *Structural Equation Modeling, 22*(4), 504-516.

Bostic, T. J., Rubio, D. M., & Hood, M. (2000). A validation of the subjective vitality scale using structural equation modeling. *Social Indicators Research, 52*(3), 313-324.

Brown, T.A. (2015). *Confirmatory factor analysis for applied research*. London: The Guilford Press.

Brown, T. A., Chorpita, B. F., & Barlow, D. H. (1998). Structural relationships among dimensions of the DSM-IV anxiety and mood disorders and dimensions of negative affect, positive affect, and autonomic arousal. *Journal of Abnormal Psychology, 107*(2), 179-192.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*(1), 3-5.

Byrne, B. M., & Stewart, S. M. (2006). Teacher's corner: The MACS approach to testing for multigroup invariance of a second-order structure: A walk through the process. *Structural Equation Modeling, 13*(2), 287-321.

Cattell, R.B. (1978). *The Scientific Use of Factor Analysis*. New York, Plenum.

Cella, D., Riley, W., Stone, A., Rothrock, N., Reeve, B., Yount, S., ... & Group, P. C. (2010). The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *Journal of Clinical Epidemiology, 63*(11), 1179-1194.

Chen, G., Gully, S. M., & Eden, D. (2001). Validation of a new general self-efficacy scale. *Organizational Research Methods, 4*(1), 62-83.

Chen, F. F., Jing, Y., Hayes, A., & Lee, J. M. (2013). Two concepts or two approaches? A bifactor analysis of psychological and subjective well-being. *Journal of Happiness Studies, 14*(3), 1033-1068.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233-255.

Coffey, J. K., Wray-Lake, L., Mashek, D., & Branand, B. (2016). A multi-study examination of well-being theory in college and community samples. *Journal of Happiness Studies, 17*, 187-211.

Cummins, R. A. (2013). Measuring happiness and subjective well-being. In S. A. David, I. Boniwell, & A. C. Ayers (Eds.), *The Oxford handbook of happiness*, (pp. 185-200). Oxford: Oxford University Press.

de Bruin, G. P., & du Plessis, G. A. (2015). Bifactor analysis of the mental health continuum–short form (MHC–SF). *Psychological Reports, 116*(2), 438-446.

De Vaus, D. (2002). *Surveys in social research.* Crows Nest: Allen & Unwin.

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*(4), 227-268.

DeSimone, J. A., Harms, P. D., & DeSimone, A. J. (2015). Best practice recommendations for data screening. *Journal of Organizational Behavior, 36*(2), 171-181.

DeVellis, R. F. (2012). *Scale development: Theory and applications*. London: Sage Publications.

Diener, E. (2000). Subjective well-being: The science of happiness and a proposal for a national index. *American Psychologist*, *55*(1), 34-43.

Diener, E. D., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, *49*(1), 71-75.

Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D. W., Oishi, S., & Biswas-Diener, R. (2010). New well-being measures: Short scales to assess flourishing and positive and negative feelings. Social Indicators Research, 97(2), 143-156.

Disabato, D.J., Goodman F.R., Kashdan T.B., Short J.L., & Jarden A. (2016). Different types of well-being? A cross-cultural examination of hedonic and eudaimonic well-being. *Psychological Assessment*, *28*, 471-482.

Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The mini-IPIP scales: Tiny-yet-effective measures of the Big Five factors of personality. *Psychological Assessment, 18*(2), 192-203.

Fabrigar, L. R., & Wegener, D. T. (2012). *Exploratory factor analysis*. Oxford: Oxford University Press.

Fastame, M. C., Penna, M. P., & Hitchcott, P. K. (2015). Life satisfaction and social desirability across the late life span: What relationship? *Quality of Life Research, 24*(1), 241-244.

Feldman, G., Hayes, A., Kumar, S., Greeson, J., & Laurenceau, J. P. (2007). Mindfulness and emotion regulation: The development and initial validation of the Cognitive and Affective Mindfulness Scale-Revised (CAMS-R). *Journal of Psychopathology and Behavioral Assessment, 29*(3), 177-190.

Feldman Barrett, L., & Russell, J. A. (1998). Independence and bipolarity in the structure of current affect. *Journal of Personality and Social Psychology, 74*(4), 967-984.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, *18*(1), 39-50.

Fredrickson, B. L., & Losada, M. F. (2005). Positive affect and the complex dynamics of human flourishing. *American Psychologist, 60*(7), 678-686.

Gallagher, M. W., Lopez, S. J., & Preacher, K. J. (2009). The hierarchical structure of well-being. *Journal of Personality, 77*(4), 1025-1050.

Gibson, B., & Sanbonmatsu, D. M. (2004). Optimism, pessimism, and gambling: The downside of optimism. *Personality and Social Psychology Bulletin, 30*(2), 149-160.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, *60*, 549-576.

Hair, J. F., Black, W. C., Babin, B. J, Anderson, R. E. (2010). *Multivariate data analysis*: *A global perspective*. New York: Pearson.

Hardesty, D. M., & Bearden, W. O. (2004). The use of expert judges in scale development: Implications for improving face validity of measures of unobservable constructs. *Journal of Business Research, 57*(2), 98-107.

Harrison, D. A., & McLaughlin, M. E. (1996). Structural properties and psychometric qualities of organizational self-reports: Field tests of connections predicted by cognitive theory. *Journal of Management, 22*(2), 313-338.

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science, 43*(1), 115-135.

Hogan, M. J., Johnston, H., Broome, B., McMoreland, C., Walsh, J., Smale, B., ... & Groarke, A. M. (2015). Consulting with citizens in the design of wellbeing measures and policies: Lessons from a systems science application. *Social Indicators Research, 123,* 857-877.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*(2), 65-70.

Huff, C., & Tingley, D. (2015). “Who are these people?” Evaluating the demographic characteristics and political preferences of MTurk survey respondents. Research & Politics, 2(3), 1-12.

Huppert, F. A., & So, T. T. (2013). Flourishing across Europe: Application of a new conceptual framework for defining well-being. *Social Indicators Research*, *110*(3), 837-861.

Huta, V. (2013). Eudaimonia. In In S. A. David, I. Boniwell, & A. C. Ayers (Eds.), *The Oxford handbook of happiness*, (pp. 200-213). Oxford: Oxford University Press.

Jackson, S. A., & Eklund, R. C. (2004). *The flow scales manual*. Morgantown, WV: Fitness Information Technology.

Jovanović, V. (2015). Structural validity of the Mental Health Continuum-Short Form: The bifactor model of emotional, social and psychological well-being. *Personality and Individual Differences, 75*, 154-159.

Jonason, P. K., Kaufman, S. B., Webster, G. D., & Geher, G. (2013). What lies beneath the Dark Triad Dirty Dozen: Varied relations with the Big Five. *Individual Differences Research*, *11*(2), 81-90.

Jonason, P. K., & Webster, G. D. (2010). The dirty dozen: a concise measure of the dark triad. *Psychological Assessment*, *22*(2), 420-432.

Kashdan, T. B., Biswas-Diener, R., & King, L. A. (2008). Reconsidering happiness: The costs of distinguishing between hedonics and eudaimonia. *The Journal of Positive Psychology*, *3*(4), 219-233.

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research*, *44*, 486-507.

Keyes, C. L. M. (1998). Social well-being. *Social Psychology Quarterly*, *61*(2), 121-140.

Keyes, C. L. M. (2002). The mental health continuum: From languishing to flourishing in life. *Journal of Health and Social Behavior*, *43*(2), 207-222.

Keyes, C. L. M. (2013). Promoting and protecting positive mental health: Early and often throughout the lifespan. In C. L. M. Keyes (Ed.), *Mental well-being: International contributions to the study of positive mental health*, (pp. 3-28). New York: Springer.

Keyes, C. L. M. (2015). Human Flourishing and Salutogenetics. In Genetics of Psychological Well-Being: The role of heritability and genetics in positive psychology (pp. 3-19). Oxford: Oxford University Press.

Kline, P. (2000). *The handbook of psychological testing*. New York: Routledge.

Kline, R. B. (2016). *Principles and practice of structural equation modeling*. New York: Guilford Press.

Liao, P. S. (2014). More happy or less unhappy? Comparison of the balanced and unbalanced designs for the response scale of general happiness. *Journal of Happiness Studies*, *15*(6), 1407-1423.

Little, T. D., Slegers, D. W., & Card, N. A. (2006). A non-arbitrary method of identifying and scaling latent variables in SEM and MACS models. *Structural Equation Modeling*, *13*(1), 59-72.

Longo, Y., Gunz, A., Curtis, G. J., & Farsides, T. (2016). Measuring need satisfaction and frustration in educational and work contexts: the need satisfaction and frustration scale (NSFS). *Journal of Happiness Studies, 17*(1), 295-317.

Longo, Y., Coyne, I., Joseph, S., & Gustavsson, P. (2016). Support for a general factor of well-being. *Personality and Individual Differences*. *100*, 68-72.

Maniaci, M. R., & Rogge, R. D. (2014). Caring about carelessness: Participant inattention and its effects on research. *Journal of Research in Personality, 48*, 61-83.

McCrae, R. R., & Costa, P. T. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology, 51*(6), 882-888.

McNulty, J. K., & Karney, B. R. (2004). Positive expectations in the early years of marriage: Should couples expect the best or brace for the worst? *Journal of Personality and Social Psychology, 86*(5), 729-743.

Meade, T., & Dowswell, E. (2015). Health-related quality of life in a sample of australian adolescents: Gender and age comparison. *Quality of Life Research, 24*(12), 2933-2938.

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology, 93*(3), 568-592.

Morgan, J., & Farsides, T. (2009). Measuring meaning in life. *Journal of Happiness Studies, 10*(2), 197-214.

Neff, K. D. (2003). The development and validation of a scale to measure self-compassion. *Self and Identity, 2*(3), 223-250.

Oxford Dictionaries. (2014). English Synonyms. Retrieved from http://www.oxforddictionaries.com/

Pavot, W., & Diener, E. (2008). The satisfaction with life scale and the emerging construct of life satisfaction. *The Journal of Positive Psychology, 3*(2), 137-152.

Peer, E., Vosgerau, J., & Acquisti, A. (2014). Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior Research Methods, 46*(4), 1023-1031.

Peterson, C., & Seligman, M. E. (2004). *Character strengths and virtues: A handbook and classification*. New York: Oxford University Press.

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. URL http://www.R-project.org/.

Revelle, W. (2015). psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, http://CRAN.R-project.org/package=psych

Reynolds, W. M. (1982). Development of reliable and valid short forms of the Marlowe‐Crowne Social Desirability Scale. *Journal of Clinical Psychology, 38*(1), 119-125.

Robitschek, C., Ashton, M. W., Spering, C. C., Geiger, N., Byers, D., Schotts, G. C., & Thoen, M. A. (2012). Development and psychometric evaluation of the Personal Growth Initiative Scale–II. *Journal of Counseling Psychology*, *59*(2), 274-287.

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods, 21*(2), 137-150.

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2), 1-36.

Russell, J.A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161-1178.

Ryan, R. M., & Deci, E. L. (2001). On happiness and human potentials: A review of research on hedonic and eudaimonic well-being. *Annual Review of Psychology, 52*(1), 141-166.

Ryan, R. M., & Frederick, C. (1997). On energy, personality, and health: Subjective vitality as a dynamic reflection of well- being. *Journal of Personality*, 65(3), 529-565.

Ryan, R. M., Huta, V., & Deci, E. L. (2008). Living well: A self-determination theory perspective on eudaimonia. *Journal of Happiness Studies*, *9*(1), 139-170.

Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology*, *57*(6), 1069-1081.

Ryff, C. D., Singer, B. H., & Love, G. D. (2004). Positive health: Connecting well-being with biology. *Philosophical Transactions of the Royal Society of London, B Biological Sciences*, *359*, 1383-1394.

Saris, W. E., & Gallhofer, I. N. (2014). *Design, evaluation, and analysis of questionnaires for survey research*. Honboken, NJ: John Wiley & Sons.

Scheier, M. F., Carver, C. S., & Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a reevaluation of the Life Orientation Test. *Journal of Personality and Social Psychology*, *67*(6), 1063-1078.

Schmitt, N., & Stults, D. M. (1985). Factors defined by negatively keyed items: The result of careless respondents? *Applied Psychological Measurement*, *9*(4), 367-373.

Schulenberg, S. E., Schnetzer, L. W., & Buchanan, E. M. (2011). The purpose in life test-short form: Development and psychometric support. *Journal of Happiness Studies*, *12*(5), 861-876.

Seligman, M. E. P. (2011). *Flourish: A new understanding of happiness and well-being – and how to achieve them*. London: Nicholas Brealey.

semTools Contributors. (2016). semTools: Useful tools for structural equation modeling. R package version 0.4-12. Retrived from http://cran.r-project.org/web/packages/semTools/index.html

Sheldon, K. M., Elliot, A. J., Kim, Y., & Kasser, T. (2001). What is satisfying about satisfying events? Testing 10 candidate psychological needs. *Journal of Personality and Social Psychology, 80*(2), 325-339.

Sheldon, K. M., & Hilpert, J. C. (2012). The balanced measure of psychological needs (BMPN) scale: An alternative domain general measure of need satisfaction. *Motivation and Emotion, 36*(4), 439-451.

Spector, P. E. (2013). Survey design and measure development. In T. D. Little (Ed.), *The Oxford Handbook of Quantitative Methods, Volume 1: Foundations*, (pp. 170-188). Oxford: Oxford University Press.

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*(4), 1080-1102.

Steel, P., Schmidt, J., & Shultz, J. (2008). Refining the relationship between personality and subjective well-being. *Psychological Bulletin, 134*(1), 138-161.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, *87*(2), 245-251.

Thayer, R. E. (1986). Activation-deactivation adjective check list: Current overview and structural analysis. *Psychological Reports*, *58*(2), 607-614.

Waterman, A. S., Schwartz, S. J., Zamboanga, B. L., Ravert, R. D., Williams, M. K., Agocha, V. B., Kim, S. Y., & Donnellan, M. B. (2010). The Questionnaire for Eudaimonic Well-Being: Psychometric properties, demographic comparisons, and evidence of validity. *The Journal of Positive Psychology*, *5*(1), 41-61.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063-1070.

Weijters, B., De Beuckelaer, A., & Baumgartner, H. (2014). Discriminant validity where there should be none: Positioning same-scale items in separated blocks of a questionnaire. *Applied Psychological Measurement*, *38*(6), 450-463.

Weinstein, N., Przybylski, A. K., & Ryan, R. M. (2012). The index of autonomous functioning: Development of a scale of human autonomy. *Journal of Research in Personality*, *46*(4), 397-413.

Woods, C. M. (2006). Careless responding to reverse-worded items: Implications for confirmatory factor analysis. *Journal of Psychopathology and Behavioral Assessment*, *28*(3), 186-191.

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α, Revelle's β, and McDonald's ω H: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, *70*, 123-133.