Routledge
Taylor & Francis Group

Abbreviated Three-Item Versions of the Satisfaction with Life Scale and the Harmony in Life Scale Yield as Strong Psychometric Properties as the Original Scales

Oscar N. E. Kjell[1] and Ed Diener[2]

[1]Department of Psychology, Lund University, Lund, Sweden; [2]Department of Psychology, University of Illinois at Urbana-Champaign, Champaign and Urbana, Illinois

ABSTRACT

The cognitive components of subjective well-being can be measured with the Satisfaction with life scale (SWLS) and the Harmony in life scale (HILS), which both comprise five items each. The aim of this article is to abbreviate these scales and examine their psychometric properties and validity. Three datasets including test-retest data are used (N = 787; N = 860; N = 343). The two first datasets were already collected, whereas the third dataset included delivering the three-item scales (SWLS-3; HILS-3) together (in random order) with one shared instruction. The last study was pre-registered, including open data and code. The SWLS-3 and the HILS-3 demonstrate good psychometric properties, including very high internal consistency and item total correlations, strong test-retest reliability, where two-factor models of cognitive well-being tend to yield very good fit indices. Further, the scales demonstrate measurement invariance across time and gender. In fact, the three-item scales demonstrate as strong psychometric properties as compared with the five-item scales. Additionally, the scales demonstrate similar validity by yielding similar correlations to assessments of well-being, mental health problems and social desirability. Thus, the SWLS-3 and the HILS-3 can efficiently be used together with one shared instruction, without compromising (and in most aspects even yielding small improvements) the psychometric soundness of the scales.

Introduction

The subjective well-being (SWB) approach assesses well-being as a cognitive component and an affective component (Diener, 1984). The cognitive component focuses on life evaluations: how individuals *think* about their lives. To assess the cognitive component, the Satisfaction with life scale (SLWS; Diener, Emmons, Larsen, & Griffin, 1985) is most often used; whilst recent research show that it can meaningfully be complemented with the Harmony in life scale (HILS; Kjell, Daukantaitė, Hefferon, & Sikström, 2016). The two scales have in common that they do not impose a lot of criteria or aspects that respondents are forced to evaluate: They allow *subjective* evaluations, where respondents decide for themselves what they consider meaningful and important in relation to satisfaction with life (SWL) and harmony in life (HIL).

The main aim of this article is to abbreviate these scales, whilst not compromising their psychometric soundness. It is not argued that the original SLWS and HILS are poor, but rather that they more efficiently can be delivered in abbreviated versions, without compromising the psychometric properties. It is valuable to provide evidence supporting the use of abbreviated versions of the SWLS and the HILS. This is because answering long scales with many items may put unnecessary demands on participants and may result in poor data quality in long surveys; whilst efficiently and accurately assessing both SWL and HIL result in a more comprehensive and detailed understanding of well-being (e.g., see Kjell, 2011, 2018). The examination of the psychometric properties in this article includes examining their internal consistency, item total correlation, test-retest reliability, factor structure using confirmatory factor analyses as well as examining measurement invariance across time and gender in three different datasets. In addition, validity is investigated by examining the scales' correlation to other measures of well-being, mental health problems and social desirability.

Advantages of shorter scales

Short scales have been found advantageous in several contexts. Sandy, Gosling, Schwartz, and Koelkebeck (2017) describe several such contexts, including large online studies where, respondents might not have the patience for long questionnaires; longitudinal designs, where respondents are tracked at numerous occasions over a long time; and pre-screenings, where the aim is to quickly identify a number of traits or states before allowing entry to a full study. Further,

"the demand for short scales is currently expanding at an accelerating speed. One reason for the increasing need for short scales could be a changing way to approach psychological research in general. With research questions becoming more and more complex, involving more and more constructs …" (Ziegler, Kemper, & Kruyen, 2014, p. 185). Examples of short measures with satisfactory psychometric properties include the 5 and 10 items scales for the Big-Five personality domains (Gosling, Rentfrow, & Swann, 2003), the Ten Item Values Inventory (Sandy et al., 2017) and the Single-Item Self-Esteem Scale (Robins, Hendin, & Trzesniewski, 2001).

Satisfaction with life and Harmony in life

Research demonstrate that SWL and HIL complement each other in providing a comprehensive understanding of subjective well-being (e.g., see Kjell, 2011, 2018). SWL involves "a global assessment of a person's quality of life according to his [or her] chosen criteria" (Shin & Johnson, 1978, as cited in Diener et al., 1985, p. 71). In contrast, HIL "is by its very nature relational. It is through mutual support and mutual dependence that things flourish" (Li, 2008, p. 427). That is, "harmony encourages a holistic world view that incorporates a balanced and flexible approach to personal well-being that takes into account social and environmental contexts" (Kjell et al., 2016, p. 894). In accordance to these definitions, individuals describe their SWL with words such as *happy*, *content*, *fulfilled*, *pleased* and *gratified*; and their HIL with words such as *peaceful*, *balanced*, *calm*, *unity* and *agreement* (Kjell, Kjell, Garcia, & Sikström, 2018). Further, in a large cross-cultural investigation where individuals where allowed to freely describe what happiness is for them, the responses concerned both harmony and psychological balance (25% of the responses) as well as satisfaction (7% of the responses; Delle Fave, Brdar, Freire, Vella-Brodrick, & Wissing, 2011, see also similar results in Delle Fave et al., 2016). Hence, together SWL and HIL capture central and complementary aspects of well-being.

Items of the SWLS-3 and the HILS-3

The original SWLS and HILS comprise five items each (SWLS-5; HILS-5); and, here it is suggested that the first three items of each scale (SWLS-3; HILS-3) are most apt to form abbreviated versions. From a psychometric property perspective, the three first items of the SWLS yielded the strongest factor loadings and item-total correlations (Diener et al., 1985); and research has identified the last item showing less convergence with the other items (Pavot & Diener, 2009; see also Vittersø, Biswas-Diener, & Diener, 2005). Similarly, the first three items of the HILS also yielded the strongest item-total correlations (Kjell et al., 2016). Further, in a two-factor solution of the SWLS-5 and the HILS-5, the first three items of the scales yielded the strongest factor loadings at two separate measurement occasions (with the same participants).

The first three items of each scale also make most sense to select from a theoretical perspective as they arguably are most directly tapping into the targeted constructs. The first three items in the SWLS-5 concern being *satisfied*, having an *ideal life* or *excellent conditions*; whereas the last two items tap into evaluating one's past (as in *So far I have gotten the important things I want in life*), and have gotten important things (as in *If I could live my life over, I would change almost nothing*). In terms of the HILS-5, the first three items focus on the most central aspects of HIL where the items include the words *harmony* or *balance* as opposed to the last two items that focus on *accept* (as in *I accept the various conditions of my life*) and *fitting in* (as in *I fit in well with my surroundings*).

Psychometric properties

Previous research indicates that the five-item scales of SWL and HIL demonstrate good psychometric properties in terms of internal consistency, test-retest, item-total correlations as well as test-retest reliability (e.g., see Diener et al., 1985; Kjell et al., 2016). Confirmatory factor analyses have further demonstrated that the SWLS-5 and the HILS-5 form a two-factor model with good fit (Kjell et al., 2016). Although, to our knowledge, measurement invariance has not been examined for the HILS-5. Whereas, for the SWLS-5, research has found that factor loadings, unique variances and factor variance are invariant across sexes (Shevlin, Brunsden, & Miles, 1998; for a review see Emerson, Guhn, & Gadermann, 2017) and time (using a spanish version in an adolescent sample; Esnaola, Benito, Antonio-Agirre, Axpe, & Lorenzo, 2019). This article focuses on examining the psychometric properties of the three-item scales in regard to internal consistency, item total correlation, test-retest reliability, factor structure using confirmatory factor analyses as well as measurement invariance across gender and time. Further, as comparison, the article presents these aspects of psychometric properties for the five-item scales as well, which are based on two of the three datasets (i.e., the last dataset only comprises the three-items scales).

Hypotheses

The following hypotheses were pre-registered *after* having analyzed the first two datasets (already collected from Kjell et al., 2016; 2018) but *before* collecting the third dataset. The pre-registered hypotheses include:

H_1. The SWLS-3 and the HILS-3 yield *good* internal consistency and *strong* or *very strong* item total correlations.

H_2. The SWLS-3 and the HILS-3 yield a good fit in a two-factor solution using confirmatory factor analyses.

H_3. The SWLS-3 and the HILS-3 yield *strong* longitudinal measurement invariance (strong measurement invariance is further described in the Statistical methods section).

H_4. The SWLS-3 and the HILS-3 yield *strong* or *very strong* test-retest correlations after two weeks follow up.

H_5. The SWLS-3 and the HILS-3 yield *strong* measurement invariance across gender.

In addition to these hypotheses, the validity for the three-item scales is investigated by examining their correlation to constructs relating to well-being (i.e., happiness, and psychological well-being), mental health problems (i.e., depression, anxiety and stress) and social desirability; where the focus is to compare the correlations with the five-item scales. We did not pre-register specific hypotheses for these analyses, but generally anticipated the correlations of the three- and five-item scales to be similar.

Materials and methods

Participants

Participants in *Dataset 1* were taking part in Kjell et al. (2016) second study including a range of other well-being related instruments. Participants in *Dataset 2* were taking part in Kjell et al.'s (2018) seventh study, which also included several other well-being related instruments. *Dataset 3* was specifically collected for the purpose of this article, where the material is described under Instruments below.

Dataset 1 were collected using Mechanical Turk (Mturk) and included a test-retest procedure (M = 57.2; SD = 5.6 days between Time 1 [T1] and Time 2 [T2]). At T1, 787 participants completed the survey and control questions correctly (360 Females and 427 Males, with a mean age of 30.8 [SD = 9.8] years, 141 failed the control questions); at T2, 535 participants completed the survey and the control questions correctly (252 females and 283 males, with a mean age of 31.2 [SD = 9.8] years, 60 failed the control questions). Most participants came from India, followed by the USA and other countries; for more detailed information see Kjell et al. (2016).

Dataset 2 were also collected on Mturk including test-retest (M = 30.8; SD = 2.0 days between T1 and T2). At T1, 860[1] participants completed the survey and control questions correctly (439 Females and 421 Males, with a mean age of 32.8 [SD =10.1] years, 42 failed the control questions); at T2, there were 477 participants (261 Females and 216 Males, with a mean age of 34.1 [SD = 10.4] years, 42 failed the control questions). More than 90% of the participants reported coming from the USA, followed by other countries; see Kjell et al. (2018) for more details.

Dataset 3 were collected for this article. Participants were recruited from Prolific, using the following pre-screeners: Fluency in English, nationality from the UK and the minimum age of 18 years. Participants were paid £0.3 to partake at T1, and the study took 1.02 (SD = 1.5) mins to complete. Three-hundred-fifty participants completed the study, but 7 answered the control question incorrectly and were removed from the analyses. The final sample comprises 343 participants (236 Females, 106 Males, and 1 Other, with a mean age of 34.4 [SD = 11.9] years).

After two weeks, the 343 participants were invited to partake again for £0.3. As pre-registered, those who had not

answered were sent a reminder two days later; the survey was closed one week after the first invitation for T2. Three-hundred participants answered but one was removed for not answering the control item correctly. The final T2 sample comprised 299 participants (87.2% of the T1 sample; 214 Females, 84 Males, and 1 Other, with a mean age of 35.0 [SD = 12.1] years). The study took on average 1.03 (SD = 1.60) minutes to complete, and there were on average 14.8 (SD = 1.40) days between T1 and T2.

Instruments

For Dataset 1 and 2 we only present the instruments that are employed in the analyses of this article; whereas for Dataset 3 we describe all measures that were included in the data collection.

Dataset 1

The Satisfaction with Life Scale (SWLS; Diener et al., 1985) assesses life satisfaction with five items (e.g., *In most ways my life is close to my ideal*) answered using a 7-point rating scale ranging from 1 = *Strongly Disagree* to 7 = *Strongly Agree*. See the Results section for psychometric information.

The Harmony in life Scale (HILS; Kjell et al., 2016) measures harmony in life with five items (e.g., *My lifestyle allows me to be in harmony*). The closed-ended items are answered on the same scale as the SWLS and the psychometric properties are presented in the Results section.

The Subjective Happiness Scale (SHS; Lyubomirsky & Lepper, 1999) measures happiness as a cognitive construct; i.e., how the respondent thinks about their life in terms of happiness. The measure comprises four items answered on closed-ended Likert-type scales that range from 1 to 7; with different scales that are specific to each item (e.g., the item: *In general, I consider myself*; is coupled with the following scale: 1 = *Not a very happy person* to 7 = *A very happy person*). The McDonald's omega was .87 and Cronbach's alpha was .82.

The Scales for Psychological Well-Being (SPWB; Ryff, 1989; Ryff & Keyes, 1995) abbreviated version comprises 18 items, which cover six subscales/dimensions involving (McDonald's omega/Cronbach's alpha are presented after each example item): Autonomy (e.g., *I judge myself by what I think is important, not by the values of what others think is important*; .48/.42), Environmental mastery (e.g., *In general, I feel I am in charge of the situation in which I live*; .67/.61), Personal growth (e.g., *For me, life has been a continuous process of learning, changing, and growth*; .53/.40), Positive relations with others (e.g., *People would describe me as a giving person, willing to share my time with others*; .61/.58), Purpose in life (e.g., *Some people wander aimlessly through life, but I am not one of them*; .52/.18), and Self-acceptance (e.g., *I like most aspects of my personality*; .71/.69). There are three items per dimension/subscale, and items are answered on a Likert-type scale that ranges from 1 = *Strongly disagree* to 6 = *Strongly agree*.

---

[1]Kjell et al., (2018) report 854 participants since 6 participants had not written any words in other questions and were thus removed; but are used here since they completed the SWLS and the HILS.

The *Depression, Anxiety and Stress Scale the short version* (DASS-21; Sinclair et al., 2012; shortened from Lovibond & Lovibond, 1995) includes 21 items (i.e., 7 items/construct) including Depression (e.g., *I felt down-hearted and blue*; McDonald's omega = .93; Cronbach's alpha = .91), Anxiety (e.g., *I felt I was close to panic*; McDonald's omega = .90; Cronbach's alpha = .87) and Stress (e.g., *I found it hard to wind down*; McDonald's omega = .89; Cronbach's alpha = .87). The items are answered using a 4-point scale referring to severity/frequency, which ranges from 0 = *Not at all* to 3 = *Very much, or most of the time*.

The *Marlowe-Crowne Social Desirability Scale* (Crowne & Marlowe, 1960) *the shorter version Form A* (Reynolds, 1982) comprises 11 items that capture social desirability (e.g., *I am always courteous, even to people who are disagreeable*). Respondents are required to answer whether the statement is personally *True* or *False* for them. McDonald's omega was .70 and the Cronbach's alpha was .65.

Dataset 2

The HILS and the SWLS as previously described for Dataset 1.

The *Patient Health Questionnaire-9* (Kroenke & Spitzer, 2002) measures Depression with nine items (e.g., *Feeling down, depressed or hopeless*), coupled with rating scales ranging from 0 = *Not at all* to 3 = *Nearly every day*. Both McDonald's omega and Cronbach's alpha were .93.

The *Generalized Anxiety Disorder Scale-7* (Spitzer, Kroenke, Williams, & Löwe, 2006) assesses Anxiety with seven items (e.g., *Worrying too much about different things*) answered on the same rating scale as the PHQ-9. Both McDonald's omega and Cronbach's alpha were .94.

The *Marlowe-Crowne Social Desirability Scale* (Crowne & Marlowe, 1960) *the shorter version Form A* (Reynolds, 1982) as previously described, was also included in this dataset, where both McDonald's omega and Cronbach's alpha were .73.

Dataset 3

The *Abbreviated Version of the Satisfaction with Life Scale* (SWLS-3) comprises the three first items from the original SWLS developed by Diener et al. (1985). The items (e.g., *I am satisfied with my life*) are answered on the same scale as the SWLS. Internal consistency statistics for the scale in the current study are presented in the Results section.

The *Abbreviated Version of the Harmony in Life Scale* (HILS-3) includes the three first items from the full version of the HILS as developed by Kjell et al. (2016). The items (e.g., *I am in harmony*) are answered using the same rating scale format as described for the SWLS. For internal consistency statistics see the Results section.

The *Control Question* included the following attention check question: *Please answer the alternative '4 neither agree nor disagree' below*. Participants that failed to answer it correctly were removed from the analyses as pre-registered. This kind of attention checks has been shown to increase the statistical power and quality of data sets (Oppenheimer, Meyvis, & Davidenko, 2009).

Procedure

All three datasets were collected online, and participants were first informed about the study, that participation is voluntary, anonymous, and that they can withdraw at any time without giving a reason. For more detailed procedural information about the collection of Dataset 1 and 2 see Kjell et al. (2016, 2018), respectively.

In the collection of Dataset 3, participants were also informed that they will be asked to partake in two weeks' time, and that their data would be open upon publication of the article. Then participants were asked to enter their Prolific ID and to fill out the SWLS-3 and the HILS-3 in randomized order (i.e., the scales were presented together, on the same webpage, only showing the instructions once). Lastly, participants answered the demographic questions and were debriefed. After two weeks, participants were contacted again and asked to complete the same survey again as specified above; after two days, those participants that had not answered the survey were reminded, and after a week the survey was closed.

Statistical methods

The data was analyzed using frequentist (Neyman-Pearson) statistics, Confirmatory Factor Analyses (CFA), and measurement invariance analyses. The following criteria were used for the two first datasets and pre-registered for the third dataset. The alpha level was set to .05. Cronbach's alpha and McDonald's omega above .70 were considered to indicate *good* internal consistency. Pearson correlations of .2 − .39 were interpreted as *weak*, .40-.59 as *moderate*, .6 − .79 as *strong*, and above .8 as *very strong*.

To identify the CFA models, the factor loading of the first item of the latent variables were set to 1 (which is the default in lavaan; Rosseel, 2012). The robust Maximum likelihood estimator (MLM) was used as individual items were not normally distributed in the datasets (and was thus pre-registered for Dataset 3). *P*-values above .05 are considered to indicate good fit; however, since *p*-values are biased by sample size the following criteria were also used to indicate *good* fit: the comparative fit index (CFI) above .95 and the root mean square error of approximation (RMSEA) below .05 (and below .08 for *acceptable* fit; Schreiber, Nora, Stage, Barlow, & King, 2006). To examine whether the two three-item scales perform similarly across time and gender, measurement invariance analyses were carried out. The following five models using increasingly restrictive parameter specification across time or gender were carried out:

- Model 1 (baseline, configural) constrains the factors to be invariant across time or gender, whereas there are no equality constrains on the parameters.

- Model 2 (metric, referred to as *weak invariance*) includes constrains for the factor loadings so that they are invariant across time/gender.
- Model 3 (scalar, *strong invariance*) adds constrains on the intercepts of the items so that they are invariant across time/gender.
- Model 4 (*strict invariance*) further adds constrains to the residual variances so that they are invariant across groups.
- Model 5 further constrains the means of the factors so that they are invariant across time/gender.

To indicate non-invariance, the analyses tested the difference between the models of increased restrictions (i.e., 2-1, 3-2, 4-3, and 5-4, respectively) using $p > .05$ or a CFI difference cutoff of .01. Importantly, demonstrating strong invariance (i.e., both metric and scalar invariance) justifies comparing the means (i.e., here between time or gender), and will thus be the focus of the analyses and discussion. The data was analyzed using R 3.5.2, specifically, the CFA and the measurement invariance analyses were carried out using lavaan, 0.6-3 (Rosseel, 2012) and semTools 0.5-1 (semTools Contributors, 2016) packages, and other analyses were used using the psych 1.8.10 package (Revelle, 2018), MASS (Venables & Ripley, 2002), Hmisc (Harrell, 2019), effsize (Torchiano, 2019), and questionr (Julien et al., 2018).

Results

Descriptive statistics for items and total scores are presented in Table 1. In Dataset 3, participants who did not participate at T2 statistically differed from those who did. Participants differed in terms of age (Welch two sample t-test; $t_{(62.3)} = 2.71$, $p = .009$, Cohen's $d = 0.39$), gender (Pearson's Chi-squared test; $Chi^2_{(2)} = 8.70$, $p = .013$, Cramer's $V = .16$), and HILS-3 score (Welch Two Sample t-test; $t_{(61.0)} = 2.46$, $p = .017$, Cohen's $d = 0.36$); but not in SWLS-3 score (Welch Two Sample t-test; $t_{(59.4)} = 0.86$, $p = .391$, Cohen's $d = 0.13$).

High internal consistency and item total correlations

Dataset 1
The two three-item scales yield high Cronbach's alphas in Dataset 1 (SWLS-3 = .88; HILS-3 = .90), which are even slightly higher than for the five-item scales (SWLS-5 = .87; HILS-5 = .89). McDonald's omega total is also high for the three-item scales (SWLS-3 = .88; HILS-3 = .90), as they are for the five-item scales (SWLS-5 = .89; HILS-5 = .91). All items, in both scales, demonstrated very strong item total correlations (see Table 1).

Dataset 2
The SWLS-3 and the HILS-3 demonstrated very high Cronbach's alphas in Dataset 2 as well (SWLS-3 = .94; HILS-3 = .96); which, again, is slightly higher than the five-item scales (SWLS-5 = .93; HILS-5 = .94). Further, McDonald's omega total is also high for the three-item

Table 1. Descriptive statistics for the SWLS-3 and the HILS-3 scales and items.

Dataset	Variable	Mean	SD	Skew	Kurtosis	Alpha	Item tot. r
1	SWLS-3 total	14.90	4.28	−0.92	0.18	.88	–
	SWLS 1	4.75	1.59	−0.78	−0.18	.85	.88
	SWLS 2	4.93	1.58	−0.79	−0.27	.82	.90
	SWLS 3	5.19	1.60	−1.06	0.30	.81	.91
	HILS-3 total	15.30	3.85	−1.1	0.89	.90	–
	HILS 1	5.20	1.32	−1.12	0.96	.85	.91
	HILS 2	5.12	1.45	−0.98	0.35	.90	.90
	HILS 3	5.04	1.45	−1.00	0.45	.83	.93
2	SWLS-3 total	14.41	5.00	−0.83	−0.41	.94	–
	SWLS 1	4.63	1.78	−0.68	−0.74	.90	.95
	SWLS 2	4.82	1.71	−0.80	−0.45	.93	.93
	SWLS 3	4.97	1.80	−0.93	−0.36	.91	.95
	HILS-3 total	15.03	4.84	−0.95	−0.21	.96	–
	HILS 1	5.12	1.58	−1.00	0.09	.95	.95
	HILS 2	4.99	1.70	−0.92	−0.29	.94	.97
	HILS 3	4.92	1.75	−0.87	−0.45	.94	.97
3	SWLS-3 total	13.09	4.20	−0.47	−0.72	.88	–
	SWLS 1	4.10	1.62	−0.36	−1.10	.83	.90
	SWLS 2	4.43	1.50	−0.44	−0.73	.85	.88
	SWLS 3	4.56	1.57	−0.59	−0.71	.80	.91
	HILS-3 total	13.10	4.15	−0.44	−0.75	.92	–
	HILS 1	4.48	1.41	−0.47	−0.65	.90	.92
	HILS 2	4.42	1.56	−0.47	−0.95	.89	.93
	HILS 3	4.20	1.50	−0.37	−0.85	.86	.94

Note. N = 787 in Dataset 1; N = 860 in Dataset 2; and N = 343 in Dataset 3. Alpha = Cronbach's alpha for the entire scales or item level Cronbach's alpha for when item was dropped. Item tot. r = Item total correlation; SWLS = Satisfaction with life scale; HILS = Harmony in life scale; SWLS 1 = *In most ways my life is close to my ideal*; SWLS 2 = *The conditions of my life are excellent*; SWLS 3 = *I am satisfied with my life*; HILS 1 = *My lifestyle allows me to be in harmony*; HILS 2 = *Most aspects of my life are in balance*; HILS 3 = *I am in harmony*.

scales (SWLS-3 = .94; HILS-3 = .96), as they are for the five-item scales (SWLS-5 = .95; HILS −5 = .95). Again, item total correlations for all items in both scales were very strong.

Dataset 3
Importantly, the Cronbach's alphas are still very high when the scales are delivered as three-item scales (SWLS-3 = .88; HILS-3 = .92), and so are also McDonald's omega total (SWLS-3 = .88; HILS-3 = .92). Item total correlations were also very strong in this dataset.

Correlations among the scales

Dataset 1
The intercorrelation between the SWLS-3 and the SWLS-5 is very strong (r = .95; Table 2); as well as between the HILS-3 and the HILS-5 (r = .96). Further, the correlation between the three-item scales is very similar to the correlation between the five-item scales (i.e., SWLS-3 and HILS-3: r = .73; SWLS-5 and HILS-5: r = .74).

Dataset 2
The intercorrelations between the SWLS-3 and the SWLS-5 as well as between the HILS-3 and the HILS-5 are also very strong in Dataset 2 (r = .97 and r = .98, respectively). Further, the correlation between the three-item scales is similar to the correlation between the five-item scales (i.e., r = .85 and r = .84, respectively).

Table 2. Pearson's correlations among the scales.

Dataset		SWLS-3	HILS-3	SWLS-5	HILS-5
1	1. SWLS-3	–			
	2. HILS-3	.73	–		
	3. SWLS-5	.95	.73	–	
	4. HILS-5	.75	.96	.74	–
2	1. SWLS-3	–			
	2. HILS-3	.85	–		
	3. SWLS-5	.97	.84	–	
	4. HILS-5	.85	.98	.84	–
3	1. SWLS-3	–	.74	NA	NA
	2. HILS-3	.74	–	NA	NA

Notes. $p < .001$ for all correlations. N = 787 in Dataset 1; N = 860 in Dataset 2; N = 343 in Dataset 3.

Table 3. CFA results of the SWLS and the HILS show that a 2-factor solution with the shorter scales yield the best fit.

Dataset	Model	Chi^2	p-value	df	CFI	RMSEA
1	**5-items**					
	1-factor	361.647	.000	35	.889	.137
	2-factor	122.407	.000	34	.970	.072
	3-items					
	1-factor	199.414	.000	9	.894	.228
	2-factor	37.827	.000	8	.983	.089
2	**5-items**					
	1-factor	485.641	.000	35	.928	.165
	2-factor	105.228	.000	34	.989	.064
	3-items					
	1-factor	242.719	.000	9	.945	.251
	2-factor	15.424	.051	8	.998	.043
3	**3-items**					
	1-factor	111.480	.000	9	.927	.224
	2-factor	13.156	.107	8	.996	.052

Notes. N = 787 in Dataset 1; N = 860 in Dataset 2; and N = 343 in Dataset 3.
Estimator = robust Maximum Likelihood (MLM)
SWLS = Satisfaction with life scale; HILS = Harmony in life scale.

Dataset 3

When the SWLS-3 and the HILS-3 are delivered on the same page the correlation between them (r = .74) falls between the correlations of Dataset 1 and 2, where the items were presented on different pages (c.f. r = .73 in Dataset 1, and r = .85 in Dataset 2).

Good fit for a two-factor, rather than a one-factor, model

CFA were used to examine whether the six items of the SWLS-3 and the HILS-3 were best captured in a two-factor as compared with a one-factor model. A two-factor model yielded a considerably better fit than a one-factor model across the datasets; and the fit tended to be better for the three-item scales as opposed to the five-item scales (see Table 3 for fit indices and Figure 1 for factor loadings).

Dataset 1

For the three-item scales, a one-factor model yields a poor fit for all fit-criteria; whereas a two-factor model of cognitive SWB including SWL and HIL yields a better fit including a good fit for CFI, but above the acceptable cutoff for RMSEA. This can be compared with the five-item scales, where the two-factor model is acceptable (see RMSEA) to good (see CFI) fit.
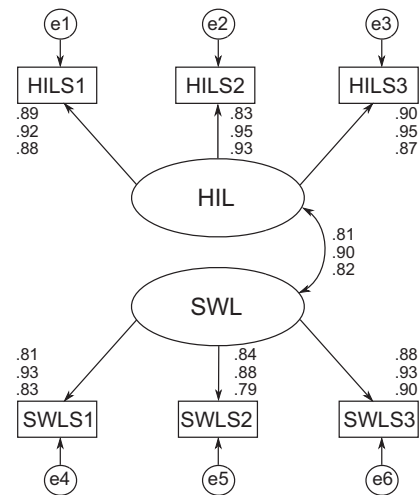

Figure 1. Standardized regression weights for the two-factor model of the Harmony in life (HIL) and the Satisfaction with life (SWL) three-item scales in three different datasets. The first row is Dataset 1 (N = 787); the second row Dataset 2 (N = 860), and the third row Dataset 3 (N = 343). The factor loading of the first item of each latent construct was set to 1.0 to identify the models.

Dataset 2

A two-factor model yields a good fit for the three-item scales, where the p-value is just above .05; and the fit indices indicate good fit. This fit is better than the one-factor model; and it is also worth noting that it is a better fit than for the five-item scales where the fit is only acceptable to good, and the p-value is below the .05 threshold.

Dataset 3

Presenting all items together with one shared instruction did not disturb the two-factor fit in Dataset 3; where a two-factor model yields a good fit for the three-item scales, which is better than for the one-factor model. It is notable that the p-value is above .05, the CFI indicates good fit and the RMSEA acceptable fit.

Longitudinal measurement invariance

To assess psychometric equivalence of the SWLS-3 and the HILS-3 across time, analyses of longitudinal measurement invariance were carried out. Overall, both scales demonstrated *strict* invariance across time (see Table 4 for the SWLS, and Table 5 for the HILS) for all three datasets.

Dataset 1

The configural model showed an acceptable (see RMSEA) to good (see CFI) fit for the SWLS-3; and a good fit for the HILS-3. All test of $\Delta\chi^2$ between models were not significant, and according to the ΔCFI cutoff of .01, both the SWLS-3 and the HILS-3 demonstrated *strict* measurement invariance. It is also noteworthy that both five-item scales also yielded *strict* invariance; although the HILS-3 showed better fit indices than the HILS-5, which only demonstrated an acceptable configural fit.

Table 4. Results from longitudinal measurement invariance analyses of the SWLS.

D.	Model	$\chi^2_{(df)}$	$\Delta\chi^2_{(df)}$	p	CFI	ΔCFI	RMSEA	AIC	BIC
1	SWLS-5								
	1: Configural	$64.3_{(29)}$	–	–	.995	–	.039	16254	16408
	2: Loadings	$67.0_{(33)}$	$3.24_{(4)}$	.52	.995	.000	.036	16249	16386
	3: Intercepts	$67.7_{(37)}$	$0.74_{(4)}$	.95	.995	.001	.031	16242	16362
	4: Residual variances	$77.2_{(42)}$	$7.17_{(5)}$	.21	.995	.001	.031	16241	16340
	5: Factor means	$79.5_{(43)}$	$2.44_{(1)}$	.12	.994	.000	.032	16241	16336
	SWLS-3								
	1: Configural	$17.8_{(5)}$	–	–	.996	–	.063	9479	9573
	2: Loadings	$19.7_{(7)}$	$2.002_{(2)}$	.37	.996	.000	.053	9477	9562
	3: Intercepts	$20.4_{(9)}$	$0.646_{(2)}$	.72	.996	.001	.044	9473	9550
	4: Residual variances	$21.9_{(12)}$	$0.975_{(3)}$	.81	.998	.001	.031	9469	9533
	5: Factor means	$24.3_{(13)}$	$2.470_{(1)}$	.12	.997	.001	.033	9469	9529
2	SWLS-5								
	1: Configural	$66.7_{(29)}$	–	–	.995	–	.043	13680	13830
	2: Loadings	$69.6_{(33)}$	$2.84_{(4)}$	.584	.995	.000	.039	13674	13808
	3: Intercepts	$70.7_{(37)}$	$1.16_{(4)}$	.884	.996	.001	.035	13668	13784
	4: Residual variances	$87.8_{(42)}$	$10.87_{(5)}$	.054	.994	.002	.039	13675	13771
	5: Factor means	$87.9_{(43)}$	$0.04_{(1)}$	.837	.994	.000	.038	13673	13764
	SWLS-3								
	1: Configural	$6.73_{(5)}$	–	–	1	–	.004	8043	8135
	2: Loadings	$8.12_{(7)}$	$1.44_{(2)}$	.49	1	0	.000	8041	8124
	3: Intercepts	$9.07_{(9)}$	$0.95_{(2)}$	.62	1	0	.000	8038	8113
	4: Residual variances	$17.25_{(12)}$	$4.58_{(3)}$	.21	1	0	.015	8040	8103
	5: Factor means	$17.26_{(13)}$	$0.01_{(1)}$	.92	1	0	.007	8038	8096
3	SWLS-3								
	1: Configural	$7.14_{(5)}$	–	–	1	–	.010	5296	5378
	2: Loadings	$9.85_{(7)}$	$3.08_{(2)}$	.214	.999	.001	.023	5295	5369
	3: Intercepts	$10.50_{(9)}$	$0.66_{(2)}$	.718	1	.001	.000	5292	5358
	4: Residual variances	$29.89_{(12)}$	$11.29_{(3)}$	.010*	.991	.009	.063	5305	5360
	5: Factor means	$29.90_{(13)}$	$0.01_{(1)}$	.922	.991	.001	.058	5303	5355

Notes. N = 535 in Dataset 1; N = 477 in Dataset 2; and N = 299 in Dataset 3.
Estimator = robust Maximum Likelihood (MLM)
D. = Dataset; SWLS = Satisfaction with life scale, where 3 and 5 refer to the number of items.

Table 5. Results from longitudinal measurement invariance analyses of the HILS.

D.	Model	$\chi^2_{(df)}$	$\Delta\chi^2_{(df)}$	p	CFI	ΔCFI	RMSEA	AIC	BIC
1	HILS-5								
	1: Configural	$116_{(29)}$	–	–	.982	–	.067	14300	14455
	2: Loadings	$119_{(33)}$	3.29	.51	.983	.000	.062	14296	14433
	3: Intercepts	$120_{(37)}$	1.04	.90	.983	.001	.057	14289	14409
	4: Residual variances	$128_{(42)}$	4.72	.45	.983	.000	.054	14287	14386
	5: Factor means	$129_{(43)}$	0.27	.60	.984	.000	.053	14285	14380
	HILS-3								
	1: Configural	$1.92_{(5)}$	–	–	1	–	0	8532	8626
	2: Loadings	$2.21_{(7)}$	$0.304_{(2)}$	.86	1	0	0	8528	8614
	3: Intercepts	$2.34_{(9)}$	$0.131_{(2)}$	.94	1	0	0	8524	8601
	4: Residual variances	$7.30_{(12)}$	$2.958_{(3)}$	.40	1	0	0	8523	8587
	5: Factor means	$7.73_{(13)}$	$0.428_{(1)}$	.51	1	0	0	8521	8581
2	HILS-5								
	1: Configural	$140_{(29)}$	–	–	.984	–	.078	12603	12753
	2: Loadings	$146_{(33)}$	$4.72_{(4)}$	.32	.983	.000	.073	12601	12734
	3: Intercepts	$147_{(37)}$	$1.32_{(4)}$	.86	.984	.001	.068	12594	12711
	4: Residual variances	$152_{(42)}$	$1.74_{(5)}$	.88	.986	.002	.060	12589	12684
	5: Factor means	$152_{(43)}$	$0.06_{(1)}$	.81	.986	.000	.059	12587	12678
	HILS-3								
	1: Configural	$10.9_{(5)}$	–	–	.999	NA	.037	7517	7609
	2: Loadings	$12.3_{(7)}$	$1.170_{(2)}$	.56	.999	0	.027	7515	7598
	3: Intercepts	$13.6_{(9)}$	$1.294_{(2)}$	.52	1	0	.020	7512	7587
	4: Residual variances	$14.2_{(12)}$	$0.196_{(3)}$	.98	1	0	.000	7507	7569
	5: Factor means	$14.3_{(13)}$	$0.102_{(1)}$	.75	1	0	.000	7505	7563
3	1: Configural	$1.43_{(5)}$	–	–	1	–	.000	4836	4918
	2: Loadings	$2.63_{(7)}$	$1.31_{(2)}$	.520	1	0	.000	4834	4908
	3: Intercepts	$4.74_{(9)}$	$2.15_{(2)}$	.342	1	0	.000	4832	4898
	4: Residual variances	$13.90_{(12)}$	$6.37_{(3)}$	.095	1	0	.013	4835	4890
	5: Factor means	$13.93_{(13)}$	$0.03_{(1)}$	.857	1	0	.000	4833	4884

Notes. N = 535 in Dataset 1; N = 477 in Dataset 2; and N = 299 in Dataset 3.
Estimator = robust Maximum Likelihood (MLM)
D. = Dataset; HILS = Harmony in life scale, where 3 and 5 refer to the number of items.

Table 6. Test-retest reliability for the scales.

Dataset	Variables	r
1	SWLS-3 at T1 and T2	.83
	SWLS-5 at T1 and T2	.84
	HILS-3 at T1 and T2	.72
	HILS-5 at T1 and T2	.77
2	SWLS-3 at T1 and T2	.79
	SWLS-5 at T1 and T2	.82
	HILS-3 at T1 and T2	.70
	HILS-5 at T1 and T2	.71
3	SWLS-3 at T1 and T2	.81
	HILS-3 at T1 and T2	.74

Notes. For all correlations $p < .001$; N = 535 in Dataset 1; N = 477 in Dataset 2; and N = 299 in Dataset 3.
SWLS = Satisfaction with life scale; HILS = Harmony in life scale. The number after scales abbreviation refer to number of items.

Dataset 2

The configural models for both the SWLS-3 and the HILS-3 showed good fit; and again, the scales showed *strict* measurement invariance based on non-significant $\Delta\chi^2$ and ΔCFI below the threshold. Further, the five-item scales demonstrated *strict* invariance as well; although (again) the configural model of the HILS-5 were only acceptable.

Dataset 3

The configural models for both the SWLS-3 and the HILS-3 yielded good fit; where the scales, yet again, demonstrated *strict* measurement invariance based on ΔCFI below the threshold and non-significant $\Delta\chi^2$ (except for the SWLS-3, which based on $\Delta\chi^2$ demonstrated *strong* invariance as the model for residual variances were significant, $p < .010$).

Strong test-retest reliability

Overall, the test-retest of the scales were strong to very strong; although the three-item scales tend to demonstrate slightly (although probably not statistically) lower test-retest correlations than the five-item scales.

Dataset 1

Test-retest Pearson correlation for the three-item scales are strong to very strong. Test-retest correlation for the SWLS-3 is .01 units lower than for the SWLS-5 (i.e., SLWS-3: r= .83; SWLS-5: r = .84, Table 6); and for the HILS-3 it is .05 units lower than for the HILS-5 (i.e., HILS-3: r= .72; HILS-5: r = .77).

Dataset 2

The test-retest correlation for the three-item scales were strong in Dataset 2. Test-retest correlation for the SWLS-3 is .03 lower than for the SWLS-5 (i.e., SLWS-3: r = .79; SWLS-5: r = .82); and for the HILS-3 it is .01 lower than for the HILS-5 (HILS-3: r = .70; HILS-5: r = .71).

Dataset 3

Importantly, the SWLS-3 yields very strong test-retest reliability, and the HILS-3 yields strong test-retest reliability when being answered with the fourth and fifth items removed.

Measurement invariance across gender

Measurement invariance analyses were used to assess psychometric equivalence of the SWLS and the HILS across gender (i.e., between females and males). Both three-item scales demonstrated *strict* invariance across gender (see Table 7 for SWLS, and Table 8 for HILS) for all three datasets.

Dataset 1

The configural fit for the SWLS-3 and the HILS-3 were just identified. Both scales demonstrated *strict* measurement invariance based on both non-significant $\Delta\chi^2$ and ΔCFI below the threshold. It is noteworthy that the SWLS-5 yielded an acceptable to good configural fit, and the HILS-5 yielded an unacceptable to good configural fit. The SWLS-5 demonstrated non-significant $\Delta\chi^2$ and ΔCFI less than .01, whilst the HILS-5 demonstrated non-significant $\Delta\chi^2$ for all models but model 3 (and model 5, which is not the focus here) and ΔCFI less than .01 for all model comparisons.

Dataset 2

Again, the configural model was just identified, and the SWLS-3 and the HILS-3 showed *strict* measurement invariance in regard to both non-significant $\Delta\chi^2$ and ΔCFI below the threshold for all model comparisons. In this dataset the measurement invariance of SWLS-5 can be considered strict based on ΔCFI less than .01. However, it might be concerning that the $\Delta\chi^2$ for both model 2 and 3 are significant ($p = .005$, and $p < .001$, respectively). The configural model for the HILS-5 is not acceptable (see RMSEA) to good (see CFI), with strict invariance based on non-significant $\Delta\chi^2$ and ΔCFI less than .01 for all model comparisons.

Dataset 3

Replicating the results from both Dataset 1 and 2, both the SWLS-3 and the HILS-3 had a configural model that was just identified and demonstrated *strict* measurement invariance. This was in regard to both non-significant $\Delta\chi^2$ and ΔCFI below the threshold for all model comparisons.

Comparing the validity between the three- and five-item scales

The three- and five-item scales, yield similar correlation coefficients with other related psychological constructs of mental health (i.e., subjective happiness and psychological well-being), psychological problems (i.e., depression, anxiety and stress) and social desirability. This is demonstrated in Dataset 1 (see Table 9) and 2 (see Table 10).

Table 7. Results from Measurement Invariance Analyses of the SWLS Across Gender.

D.	Model	$\chi^2_{(df)}$	$\Delta\chi^2_{(df)}$	p	CFI	ΔCFI	RMSEA	AIC	BIC
1		**SWLS-5**							
	1: Configural	$32.4_{(10)}$	–		.992	–	.066	12945	13085
	2: Loadings	$42.3_{(14)}$	$9.35_{(4)}$	.052	.989	.003	.064	12947	13068
	3: Intercepts	$48.5_{(18)}$	$6.46_{(4)}$	.167	.988	.001	.060	12945	13048
	4: Residual variances	$55.0_{(23)}$	$4.10_{(5)}$	.535	.988	.001	.051	12942	13021
	5: Factor means	$58.9_{(24)}$	$3.78_{(1)}$	.051	.987	.001	.053	12944	13018
		SWLS-3							
	1: Configural	$0.00_{(0)}$	–	–	–	–	.000	7648	7732
	2: Loadings	$4.40_{(2)}$	$3.55_{(2)}$	.169	.998	–	.049	7648	7723
	3: Intercepts	$4.40_{(4)}$	$0.00_{(2)}$	1.00	1	.002	.000	7644	7710
	4: Residual variances	$7.19_{(7)}$	$1.61_{(3)}$	.657	1	.000	.000	7641	7693
	5: Factor means	$12.46_{(8)}$	$5.15_{(1)}$	.023*	.999	.001	.024	7645	7691
2		**SWLS-5**							
	1: Configural	$31.3_{(10)}$	–		.996	–	.057	13386	13529
	2: Loadings	$45.5_{(14)}$	$14.65_{(4)}$	.005**	.994	.003	.064	13392	13516
	3: Intercepts	$74.4_{(18)}$	$31.80_{(4)}$	<.001***	.987	.006	.080	13413	13518
	4: Residual variances	$82.7_{(23)}$	$3.88_{(5)}$	.566	.988	.001	.069	13411	13492
	5: Factor means	$83.2_{(24)}$	$0.54_{(1)}$	.465	.988	.000	.067	13410	13486
		SWLS-3							
	1: Configural	$0.00_{(0)}$	–	–	–	–	.000	7980	8066
	2: Loadings	$1.87_{(2)}$	$1.76_{(2)}$	.414	1	–	.000	7978	8054
	3: Intercepts	$4.22_{(4)}$	$2.36_{(2)}$	.307	1	0	.008	7977	8043
	4: Residual variances	$12.02_{(7)}$	$2.90_{(3)}$	.407	1	0	.000	7979	8031
	5: Factor means	$12.04_{(8)}$	$0.02_{(1)}$	.899	1	0	.000	7977	8024
3		**SWLS-3**							
	1: Configural	$0.00_{(0)}$	–	–	NA	–	0	3312	3381
	2: Loadings	$0.67_{(2)}$	$0.69_{(2)}$	.709	1	–	0	3308	3370
	3: Intercepts	$1.24_{(4)}$	$0.57_{(2)}$	.754	1	0	0	3305	3359
	4: Residual variances	$9.61_{(7)}$	$3.81_{(3)}$	.283	1	0	0	3307	3349
	5: Factor means	$10.97_{(8)}$	$1.39_{(1)}$	.239	1	0	0	3307	3345

Notes. N = 787 in Dataset 1; N = 860 in Dataset 2; and N = 342 in Dataset 3.
Estimator = robust Maximum Likelihood (MLM)
D. = Dataset; SWLS = Satisfaction with life scale, where 3 and 5 refer to the number of items.

Table 8. Results from measurement invariance analyses of the HILS across gender.

D.	Model	$\chi^2_{(df)}$	$\Delta\chi^2_{(df)}$	p	CFI	ΔCFI	RMSEA	AIC	BIC
1		**HILS-5**							
	1: Configural	$64.3_{(10)}$	–	–	.979	–	.109	11194	11334
	2: Loadings	$71.7_{(14)}$	$4.97_{(4)}$	.291	.979	.001	.093	11194	11315
	3: Intercepts	$81.7_{(18)}$	$10.81_{(4)}$	.029*	.976	.003	.088	11196	11298
	4: Residual variances	$86.4_{(23)}$	$2.05_{(5)}$	.842	.979	.003	.073	11190	11270
	5: Factor means	$93.0_{(24)}$	$6.31_{(1)}$	.012*	.976	.002	.075	11195	11270
		HILS-3							
	1: Configural	$0.00_{(0)}$	–	–	–	–	.000	6759	6843
	2: Loadings	$4.37_{(2)}$	$3.29_{(2)}$	.193	.999	–	.047	6759	6834
	3: Intercepts	$9.82_{(4)}$	$5.65_{(2)}$	.059	.997	.002	.058	6761	6826
	4: Residual variances	$12.42_{(7)}$	$1.02_{(3)}$	.797	1	.003	.007	6757	6809
	5: Factor means	$19.95_{(8)}$	$7.33_{(1)}$	.007**	.996	.004	.046	6763	6809
2		**HILS-5**							
	1: Configural	$63.4_{(10)}$	–	–	.990	–	.099	11718	11861
	2: Loadings	$65.1_{(14)}$	$1.2_{(4)}$	.877	.991	.001	.080	11712	11836
	3: Intercepts	$66.8_{(18)}$	$1.8_{(4)}$	.778	.992	.000	.068	11706	11810
	4: Residual variances	$102.1_{(23)}$	$9.4_{(5)}$	.094	.988	.004	.073	11731	11812
	5: Factor means	$102.2_{(24)}$	$0.0_{(1)}$	.865	.988	.000	.071	11729	11805
		HILS-3							
	1: Configural	$0.00_{(0)}$	–	–	–	–	.000	7034	7119
	2: Loadings	$0.50_{(2)}$	$0.42_{(2)}$	.811	1	–	.000	7030	7106
	3: Intercepts	$1.39_{(4)}$	$0.89_{(2)}$	.640	1	.000	.000	7027	7094
	4: Residual variances	$23.51_{(7)}$	$5.02_{(3)}$	.170	.998	.002	.044	7043	7095
	5: Factor means	$23.55_{(8)}$	$0.05_{(1)}$	.829	.998	.000	.038	7041	7089
3		**HILS-3**							
	1: Configural	$0.00_{(0)}$	–	–	1	–	0	2976	3045
	2: Loadings	$0.11_{(2)}$	$0.13_{(2)}$	.936	1	0	0	2972	3034
	3: Intercepts	$0.94_{(4)}$	$0.82_{(2)}$	.664	1	0	0	2969	3023
	4: Residual variances	$5.81_{(7)}$	$2.83_{(3)}$	.418	1	0	0	2968	3010
	5: Factor means	$9.11_{(8)}$	$3.52_{(1)}$	.060	1	0	0	2969	3008

Notes. N = 787 in Dataset 1; N = 860 in Dataset 2; and N = 342 in Dataset 3.
Estimator = robust Maximum Likelihood (MLM)
D. = Dataset; HILS = Harmony in life scale, where 3 and 5 refer to the number of items.

Table 9. Pearson's r correlation comparisons between three- and five-item scales in dataset 1.

Variables	SWLS-3	SWLS-5	Δr$_{SWLS}$	HILS-3	HILS-5	Δr$_{HILS}$
Happiness (SHS)	.66***	.65***	.01	.69***	.72***	−.03
Psychological Well-Being (SPWB)	.40***	.39***	.01	.48***	.53***	−.05
Autonomy	−.05	−.03	−.02	.03	.05	−.02
Environmental mastery	.59***	.58***	.01	.64***	.69***	−.05
Personal growth	.10*	.08	.02	.17***	.20***	−.03
Positive relations with others	.29***	.28***	.01	.34***	.38***	−.04
Purpose in life	−.02	−.04	.02	.03	.05	−.02
Self-acceptance	.66***	.66***	.00	.68***	.72***	−.04
Depression (DASS-21)	−.28***	−.24***	−.04	−.34***	−.39***	.05
Anxiety (DASS-21)	−.03	−.01	−.02	−.11*	−.13**	.02
Stress (DASS-21)	−.14***	−.10*	−.04	−.23***	−.26***	.03
Social Desirability	.30***	.29***	.01	.33***	.35***	−.02

Notes. N = 535; *p < .05; **p < .01; ***p < .001.
SWLS = Satisfaction with Life Scale; HILS = Harmony in Life Scale; SHS = Subjective Happiness Scale; SPWB = Scales of Psychological Well-Being; DASS-21 = Depression Anxiety and Stress Scales with 21 items; Social Desirability = Marlow-Crowne Social Desirability Scale-Form A.

Table 10. Pearson's r correlation comparisons between three- and five-item scales in dataset 2.

Variables	SWLS-3	SWLS-5	Δr$_{SWLS}$	HILS-3	HILS-5	Δr$_{HILS}$
Depression (PHQ-9)	−.65***	−.64***	−.01	−.66***	−.67***	.01
Anxiety (GAD-7)	−.61***	−.60***	−.01	−.66***	−.67***	.01
Social Desirability	.16***	.19***	−.03	.18***	.19***	−.01

Notes. N = 477; ***p < .001.
SWLS = Satisfaction with Life Scale; HILS = Harmony in Life Scale; PHQ-9 = Patient Health Questionnaire-9; GAD-7 = Generalized Anxiety Disorder-7; Social Desirability = Marlow-Crowne Social Desirability Scale-Form A

Dataset 1

The differences of correlation coefficients between the SWLS-3 and the SWLS-5 are very small, ranging from −.04 to .02 (Median = .01, Mean = −.003, SD = .02). The differences are also very small between the HILS-3 and the HILS-5, as the differences range from −.05 to .05 (Median = −.03, Mean = −.02; SD = .03).

Dataset 2

The differences of correlations between the SWLS-3 and the SWLS-5 are again very small, ranging from −.03 to −.01 (Median = −.01; Mean = −.02; SD = .01). Similarly, for the HILS-3 and the HILS-5 the difference in correlation coefficients are very small, ranging from −.01 to .01 (Median = .01; Mean = .003; SD = .01).

Discussion

Overall, the three-item scales of SWL and HIL yield strong psychometric properties. Often the three-item, as compared with the five-item scales, produced psychometric improvements, although these were small and they probably have little practical importance. In addition, the three-item scales form a two-factor solution of cognitive well-being with good fit, that tend to include better fit indices than the five-item scales.

First, the SWLS-3 and the HILS-3 yield high internal consistency as well as very strong item total correlations in accordance to H$_1$. In fact, the Cronbach's alphas were slightly higher (.01 − .02) for the three-item scales in comparison to the five-item scales; whilst McDonald's omega total and item total correlations were similar.

Second, the SWLS-3 and the HILS-3 demonstrate a good fit in a two-factor solution, which is in accordance to H$_2$ (although, in Dataset 1 the fit was just about unacceptable based on the RMSEA criteria, but good as indicated by the

CFI criteria). It is further important to note that the two-factor fit is better than a one-factor fit throughout all three datasets. In addition, the fit indices tend to be better for the three-item scales than the five-item scales (the only exception is in Dataset 1 where the RMSEA is somewhat better for the five-item scales, but this is not true for CFI).

Third, in accordance to H$_3$, the HILS-3 and the SWLS-3 yield *strong* longitudinal measurement invariance in all three datasets; in fact, both scales consistently demonstrated *strict* measurement invariance based on the CFI difference threshold in all three datasets. Hence, the results support invariant factor structure (i.e., see Model 1), invariant factor loadings (i.e., i.e., see Model 2), invariant item intercepts (i.e., see Model 3), and invariant residual variance (i.e., see Model 4). This demonstrates that the meaning of the constructs as measured by the SWLS-3 and the HILS-3 are similar across the repeated assessment occasions. So, the results support meaningful comparisons of the means across different measurement times. It is also of interest that the HILS-5 only demonstrated an acceptable [based on RMSEA] rather than a good configural fit; hence, from a longitudinal measurement invariance perspective it might, in fact, be more appropriate to use the HILS-3 rather than the HILS-5.

Fourth, the SWLS-3 yields strong to very strong test-retest reliability, and the HILS-3 demonstrates strong test-retest reliability, which is in agreement with H$_4$. Although, the three-item scales show smaller test-retest correlations when compared to the five-item scales, this difference can be considered small (i.e., rs are .01 to .05 units smaller). However, the removed items thus appear to be somewhat more stable over time than the included items. For example, this might be because one of the removed items in the SWLS concerns one's past (item 5: *If I could live my life over, I would change almost nothing*), and the perception of one's past might not change as quickly as one's perception of SWB level. The reasons for the lower test-retest

correlation of the HILS may be because the removed items tap in to *fitting in* and *accepting various conditions,* might have stronger stability than the core items that more directly tap into *harmony* and *balance*. Hence, the slightly lower test-retest correlations of the three-item scales might actually reflect more true changes in the targeted constructs; although these conjectures require more research.

Fifth, the HILS-3 and the SWLS-3 yield *strong* (and even *strict*) measurement invariance across gender in all three datasets, which is in accordance to H_5. Thus, there are support for measurement invariance on all four levels. Importantly, this enables the comparison of means between females and males. Further, it is interesting to note that the three-item scales did not demonstrate potential problems that is indicated by the five-item scales. For example, the SWLS-5 yields significant differences for model 2 and 3 in Dataset 2; and the HILS-5 yields significant difference between model 3 in Dataset 1. It may also be concerning that the configural model of the HILS-5 demonstrates unacceptable fit based on the RMSEA (although the CFI indicates good fit). So, from a measurement invariance perspective it might be a better choice to use the abbreviated three-item scales rather than the longer five-item versions.

In addition to the pre-registered hypotheses it is also noteworthy that the correlations between the abbreviated and original scales are very strong (r = .95 – r = .98); and the correlations between the three-item scales and the five-item scales are very similar (r difference of .01). In terms of validity, the three- and five-item scales yield very similar correlation coefficients to other well-being measures (including subjective happiness and the dimensions/subscales of psychological well-being), assessments of mental health problems (including depression, anxiety and stress) as well as social desirability.

Furthermore, it is important to note that presenting the SWLS-3 and the HILS-3 together on the same page with just one shared instruction does not increase the correlation as compared to when they are delivered on separate pages with their own instruction. This is important as presenting the items together could have made respondents more likely to interpret them similarly.

Limitations and future research

Although the samples are diverse including participants from the US, the UK, India and some other countries; all samples are collected online. Hence, future research could benefit from examining measurement invariance across nations and in samples not only collected online. Further, the SWLS-3 and the HILS-3 are only tested on their own in a very short survey in this study, where future studies will be able to show how they more specifically relate to other constructs. However, considering the very strong correlations between respective three- and five-item scales, there is very little room for differences between the scales and their correlations to other constructs. In addition, in Dataset 3, at T1 participants that *did not* complete the T2 survey significantly differed from those who *did* complete it in terms of age, gender and HILS-3 score.

However, notable the effect sizes were small, and the overall response rate at T2 were very high; that is, 87% of the participants that partook at T1 completed the survey at T2.

Moreover, the internal reliability of the Scales of Psychological Well-Being demonstrated low internal consistency as measured with Cronbach's alpha and McDonald's omega. Hence, these analyses should be interpreted with caution; however, the correlations of the three- and five-items scales did not differ considerably for these subscales. Lastly, a potential limitation of these scales might be the absence of reversed scored items. There are, however, an ongoing debate about the potential benefits of reversed items (e.g., see Suárez-Alvarez et al., 2018; Van Sonderen, Sanderman, & Coyne, 2013; Weijters, Baumgartner, & Schillewaert, 2013); especially for short scales where boredom and inattention is less likely than for longer scales.

Conclusions

To summarize, results from three different datasets show that the three-item scales of SWL and HIL demonstrate (very) high internal consistency and very strong item total correlations, where a two-factor model of cognitive well-being yield a good fit (which is better than a one-factor model). The three-item scales also demonstrate strong to very strong test-retest reliability. In addition, the scales yield strict longitudinal measurement invariance as well as strict measurement invariance across gender, which importantly enables meaningful comparisons between means across both time and gender. Lastly, the three- and five-item scales demonstrate comparable validity by yielding very similar correlation coefficients to constructs of well-being, mental health problems and social desirability.

In conclusion, the SWLS-3 and the HILS-3 demonstrate good psychometric properties and can efficiently be presented together with shared instructions. In fact, although the five-item scales demonstrated good psychometric properties, the three-item scales appear to overall yield better or competitive properties, particularly in forming a two-factor model with good fit, yielding longitudinal measurement invariance as well as measurement invariance across gender. Hence, using the SWLS-3 and the HILS-3 might be particularly useful in situations where it is important to shorten the surveys and limit the demands put on respondents. Furthermore, there are no loss of strong psychometrics, and perhaps even some small improvements in the three-item scales as compared with the five-item scales. Considering the successful abbreviation of both these scales, future research may consider shortening other commonly used scales to reap similar benefits.

Open Scholarship

References

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*(4), 349–354. doi:10.1037/h0047358

Delle Fave, A., Brdar, I., Freire, T., Vella-Brodrick, D., & Wissing, M. P. (2011). The Eudaimonic and hedonic components of happiness: qualitative and quantitative findings. *Social Indicators Research, 100*(2), 185–207. doi:10.1007/s11205-010-9632-5

Delle Fave, A., Brdar, I., Wissing, M. P., Araujo, U., Castro Solano, A., Freire, T., … Soosai-Nathan, L. (2016). Lay definitions of happiness across nations: The primacy of inner harmony and relational connectedness. *Frontiers in Psychology, 7.* doi:10.3389/fpsyg.2016.00030

Diener, E. (1984). Subjective well-being. *Psychological Bulletin, 95*(3), 542–575. doi:10.1037//0033-2909.95.3.542

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment, 49*(1), 71–75. Retrieved from //A1985AEV0700014

Emerson, S. D., Guhn, M., & Gadermann, A. M. (2017). Measurement invariance of the satisfaction with life scale: Reviewing three decades of research. *Quality of Life Research, 26*(9), 2251–2264. doi:10.1007/s11136-017-1552-2

Esnaola, I., Benito, M., Antonio-Agirre, I., Axpe, I., & Lorenzo, M. (2019). Longitudinal measurement invariance of the satisfaction with life scale in adolescence. *Quality of Life Research, 28*(10), 2831–2837. doi:10.1007/s11136-019-02224-7

Harrell, F. E. Jr. with contributions from Charles Dupont and many others. (2019). Hmisc: Harrell Miscellaneous. R package version 4.2-0. https://CRAN.R-project.org/package=Hmisc.

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. Jr. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality, 37*(6), 504–528. doi:10.1016/S0092-6566(03)00046-1

Julien, B., François, B., & Joseph, L. (2018). questionr: Functions to Make Surveys Processing Easier. R package version 0.7.0. Retrieved from https://CRAN.R-project.org/package=questionr.

Kjell, O. N. E. (2011). Sustainable well-being: a potential synergy between sustainability and well-being research. *Review of General Psychology, 15*(3), 255–266. doi:10.1037/a0024603

Kjell, O. N. E. (2018). *Conceptualizing and measuring well-being using statistical semantics and numerical rating scales.* Lund, Sweden: Lund University.

Kjell, O. N. E., Daukantaitė, D., Hefferon, K., & Sikström, S. (2016). The harmony in life scale complements the satisfaction with life scale: Expanding the conceptualization of the cognitive component of subjective well-being. *Social Indicators Research, 126*(2), 893–919. doi:10.1007/s11205-015-0903-z

Kjell, O. N. E., Kjell, K., Garcia, D., & Sikström, S. (2018). Semantic measures: Using natural language processing to measure, differentiate, and describe psychological constructs. *Psychological Methods, 24*(1), 92–115. doi:10.1037/met0000191

Kroenke, K., & Spitzer, R. L. (2002). The PHQ-9: A new depression diagnostic and severity measure. *Psychiatr Ann, 32*(9), 1–7.

Li, C. (2008). The philosophy of harmony in classical Confucianism. *Philosophy Compass, 3*(3), 423–435. doi:10.1111/j.1747-9991.2008.00141.x

Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the beck depression and anxiety inventories. *Behaviour Research and Therapy, 33*(3), 335–343. doi:10.1016/0005-7967(94)00075-U

Lyubomirsky, S., & Lepper, H. S. (1999). A measure of subjective happiness: Preliminary reliability and construct validation. *Social Indicators Research, 46*(2), 137–155.

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology, 45*(4), 867–872. doi:10.1016/j.jesp.2009.03.009

Pavot, W., & Diener, E. (2009). Review of the satisfaction with life scale. In E. Diener (Ed.), *Assessing well-being: The collected works of Ed Diener* (pp. 101–117). Dordrecht: Springer Netherlands.

Revelle, W. (2018). *psych: Procedures for personality and psychological research.* Evanston, IL: Northwestern University. Retrieved from https://CRAN.Rproject.org/package=psych Version = 1.8.10.

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. doi:10.1002/1097-4679(198201)38:1 < 119::aid-jclp2270380118 > 3.0.co;2-i

Robins, R. W., Hendin, H. M., & Trzesniewski, K. H. (2001). Measuring global self-esteem: Construct validation of a single-item measure and the Rosenberg Self-Esteem Scale. *Personality and Social Psychology Bulletin, 27*(2), 151–161. doi:10.1177/0146167201272002

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36. doi:10.18637/jss.v048.i02

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. Retrieved from //A1989CJ75000014

Ryff, C. D., & Keyes, C. L. M. (1995). The structure of psychological well-being revisited. *Journal of Personality and Social Psychology, 69*(4), 719–727. doi:10.1037/0022-3514.69.4.719

Sandy, C. J., Gosling, S. D., Schwartz, S. H., & Koelkebeck, T. (2017). The development and validation of brief and ultrabrief measures of values. *Journal of Personality Assessment, 99*(5), 545–555. doi:10.1080/00223891.2016.1231115

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research, 99*(6), 323–337. doi:10.3200/JOER.99.6.323-338

semTools Contributors. (2016). semTools: Useful tools for structural equation modeling. R package version 0.4–14. Retrieved from https://CRAN.R-project.org/package=semTools.

Shevlin, M., Brunsden, V., & Miles, J. (1998). Satisfaction with life scale: Analysis of factorial invariance, mean structures and reliability. *Personality and Individual Differences, 25*(5), 911–916. doi:10.1016/S0191-8869(98)00088-9

Shin, D. C., & Johnson, D. M. (1978). Avowed happiness as an overall assessment of the quality of life. *Social Indicators Research, 5*(1–4), 475–492. doi:10.1007/BF00352944

Sinclair, S. J., Siefert, C. J., Slavin-Mulford, J. M., Stein, M. B., Renna, M., & Blais, M. A. (2012). Psychometric evaluation and normative data for the Depression, Anxiety, and Stress Scales-21 (DASS-21) in a nonclinical sample of U.S. adults. *Evaluation & the Health Professions, 35*(3), 259–279. doi:10.1177/0163278711424282

Spitzer, R. L., Kroenke, K., Williams, J. W. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The gad-7. *Archives of Internal Medicine, 166*(10), 1092–1097. doi:10.1001/archinte.166.10.1092

Suárez-Alvarez, J., Pedrosa, I., L., Fernández, L. M., García-Cueto, E., Cuesta, M., & Muñiz, J. (2018). *Using reversed items in Likert scales: A questionable practice, 30*(2), 149–158. doi:10.7334/psicothema2018.33

Torchiano, M. (2019). _effsize: Efficient Effect Size Computation_. doi:10.5281/zenodo.1480624, R package version 0.7.6. Retrieved from https://CRAN.R-project.org/package=effsize

Van Sonderen, E., Sanderman, R., & Coyne, J. C. (2013). Ineffectiveness of reverse wording of questionnaire items: Let's learn from cows in the rain. *PloS One., 8*(7), e68967. doi:10.1371/journal.pone.0068967

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4rth ed.). New York: Springer. ISBN 0-387-95457-0

Vittersø, J., Biswas-Diener, R., & Diener, E. (2005). The divergent meanings of life satisfaction: Item response modeling of the satisfaction with life scale in Greenland and Norway. *Social Indicators Research, 74*(2), 327–348. doi:10.1007/s11205-004-4644-7

Weijters, B., Baumgartner, H., & Schillewaert, N. (2013). Reversed item bias: An integrative model. *Psychological Methods, 18*(3), 320–334. doi:10.1037/a0032121

Ziegler, M., Kemper, C. J., & Kruyen, P. (2014). Short scales–Five misunderstandings and ways to overcome them. *Journal of Individual Differences, 35*(4), 185–189. doi:10.1027/1614-0001/a000148