Assessment
2022, Vol. 29(4) 826–841

DOI: 10.1177/1073191121993558
journals.sagepub.com/home/asm


Measurement Invariance of the Subjective Happiness Scale Across Countries, Gender, Age, and Time

Gaja Zager Kocjan[1], Paul E. Jose[2], Gregor Sočan[1,*], and Andreja Avsec[1,*]

Abstract

The purpose of this study was to examine measurement invariance of the Subjective Happiness Scale across countries, gender, and age groups and across time by multigroup confirmatory factor analysis. Altogether, 4,977 participants from nine European, American, and Australian countries were included in the study. Our results revealed that both configural and metric invariance held across countries, but scalar invariance was only partially confirmed with one item yielding varying intercepts in different countries. Measurement invariance was also confirmed across gender and age groups. Longitudinal measurement invariance was examined on a subsample of 478 English-speaking participants and was fully confirmed across five consecutive assessment points. Factor means were compared between groups and across time, and good convergent validity of the Subjective Happiness Scale was found in relation to a measure of temporal satisfaction with life. Overall, our results demonstrate that self-reported happiness was measured similarly in nine different countries, gender and age groups and over time, and provide a solid foundation for meaningful cross-group and cross-time comparisons in subjective happiness.

Keywords

Subjective Happiness Scale, happiness, well-being, measurement invariance, longitudinal measurement invariance, group comparison

Researchers in the area of positive psychology have at their disposal an abundance of measures tapping the construct of happiness. However, when measuring happiness or similar constructs in broad population surveys, researchers face the problem of length restrictions. Having short yet valid happiness measures for gathering population data is important because they enable comparisons across cultural, gender, and age groups which inform countries' social policies. In addition, in multiwave or cohort studies, data are gathered over time, thus allowing for the comparison of happiness levels over several consecutive time points. However, tests of measurement invariance of relevant measurement instruments have only recently been pursued (e.g., Bieda et al., 2017, 2019; Jang et al., 2017), and these results will provide the necessary foundation for valid comparisons between groups or over time.

Defining and Measuring Subjective Happiness

In the scientific literature, the term "happiness" is used as a synonym for the cognitive evaluation of one's life in general, for one's average hedonic affective state, or for both aspects united under the umbrella of "subjective well-being" (Carlquist et al., 2017). Since individuals are able to report directly on the extent to which they are happy or not, Lyubomirsky and Lepper (1999) introduced the Subjective Happiness Scale (SHS). The scale asks individuals to estimate their general feelings of happiness and reflects "a broader and more molar category of well-being" (p. 140), which is appropriate for a short generic measure of self-perceived happiness. The scale is composed of four items: the first two items ask respondents to characterize their happiness using absolute ratings and ratings relative to peers, respectively, whereas the other two items ask respondents to indicate the extent to which the descriptions of happy and unhappy individuals correspond to how they feel themselves. The SHS, with four items, allows for more rigorous assessment of its psychometric characteristics compared

[1]University of Ljubljana, Ljubljana, Slovenia
[2]Victoria University of Wellington, Wellington, New Zealand

*Shared last authors.

Corresponding Author:
Gaja Zager Kocjan, Faculty of Arts, University of Ljubljana, Aškerčeva 2, 1000 Ljubljana, Slovenia.
Email: gaja.zagerkocjan@ff.uni-lj.si

with various single-item measures that also tap into individuals' understanding of happiness (e.g., Global Happiness Item; Bradburn, 1969).

A related construct, with which we will correlate with subjective happiness in the present study to demonstrate convergent validity, is life satisfaction. Life satisfaction as a cognitive evaluation of one's own life is probably the most frequently used operationalization of the individual's happiness (Carlquist et al., 2017) and thus a construct of interest for the investigation of the convergent validity of the SHS. Although various single-item measures of life satisfaction exist (e.g., Cheung & Lucas, 2014), the five-item Satisfaction With Life Scale (SWLS; Diener et al., 1985) has been the most common choice in well-being studies (Oishi, 2018). The scale satisfies a wide range of metric characteristics (Pavot et al., 1991). A large cross-cultural study showed that configural and metric invariance held across 26 countries, while scalar invariance was partially confirmed with two items yielding varying intercepts in different countries (Jang et al., 2017). The correlations between the SHS and the SWLS usually fall between .50 and .60, (e.g., Extremera & Fernández-Berrocal, 2014; Jovanović, 2013; Lyubomirsky & Lepper, 1999; Spagnoli et al., 2012; Szabo, 2019), indicating good convergent validity of the scale, but at the same time some degree of nonoverlapping variance, as subjective happiness cannot be fully captured by the cognitive component of subjective well-being. On the other hand, the correlation of the latent variable, manifested by the measures of life satisfaction and positive/negative affect, with the latent factor of subjective happiness has been found to be as high as .90 (Chien et al., 2020), suggesting that happiness also includes an affective component of subjective well-being in addition to the cognitive component.

Similar to the SWLS, over the past 20 years, psychometric investigations of the SHS have been performed in a wide range of countries. In particular, the SHS has been validated in American and Russian (Lyubomirsky & Lepper, 1999), Austrian and Philippine (Swami et al., 2009), Arabic (Moghnie & Kazarian, 2012), Brazilian (Damásio et al., 2014), Chilean (Vera-Villarroel et al., 2011), Chinese (Chien et al., 2020; Nan et al. 2014), French (Kotsou & Leys, 2017), Greek (Karakasidou et al., 2016), Hungarian (Szabo, 2019), Japanese (Shimai et al., 2004), Lebanese (Moghnie & Kazarian, 2012), Italian (Iani et al., 2014), Malay (Swami, 2008), Mexican (Quezada et al., 2016), Portuguese (Spagnoli et al., 2012), Serbian (Jovanović, 2013), Spanish (Extremera & Fernández-Berrocal, 2014), and Turkish (Doğan & Totan, 2013) samples. Overall, these studies indicate good metric characteristics of the scale with clear unidimensional structure, and adequate internal consistency and test–retest reliability. Nevertheless, some cross-country differences have emerged in terms of the variance explained by the single factor (ranging from 83% in a Malaysian sample to 45.2% in a Lebanese sample), the size of the factor loadings (e.g., factor loadings for the fourth item varying between .62 in the Serbian sample and as low as .23 in a Turkish sample), and the degree of internal consistency (α values ranging from .93 in a Malaysian sample to .65/.70 in Turkish samples). Construct validity of the scale has been supported in relation to various related constructs, such as subjective well-being, meaning in life, anxiety and depression, personality traits, and others.

Group Differences in Subjective Happiness Levels

Cross-cultural researchers have used the SHS to compare happiness scores across ethnicities and cultures (e.g., Swami, 2008; Swami et al., 2009). While Swami (2008) found no ethnic differences between Malay and Chinese participants from Malaysia, Swami et al. (2009) reported higher scores on the SHS for participants from individualistic nations (British and Austrian samples) compared with those from collectivistic nations (Asian samples). Although the SHS has been translated into more than 20 different languages, direct comparisons of happiness levels between countries are difficult due to the heterogeneity of the samples. In addition, variables such as objective living conditions, lay cultural beliefs, emotional patterns, characteristics of the self-construal, self-presentational concerns, response style, memory, and judgmental bias also affect the prediction of a country's level of happiness (Suh & Choi, 2018). For example, an important predictor of happiness in a country is economic wealth. Studies have shown that citizens of wealthier countries report higher life satisfaction but not higher emotional well-being (Diener et al., 2010). On the other hand, Latin Americans are happier than their GDP would suggest, and researchers offer their strongly valued human relationships as an explanation for these results (Rojas, 2012). Another important cultural variable is the dimension of individualism—collectivism. While studies report higher levels of life satisfaction in individualistic versus collectivistic cultures (Diener et al., 1995), rapid socioeconomic changes in traditionally collectivist cultures (i.e., East Asian and presumably also East European countries) may also affect the level of happiness (Cheng et al., 2011). These particularities indicate the complexity of potential predictors of happiness in a given country.

Most of the studies listed above could not test differences in average SHS scores across age groups due to the limited age variation of the samples. The results of the studies that reported age differences are not consistent; while some studies did not report age differences in the SHS scores (Doğan & Totan, 2013; Lyubomirsky & Lepper 1999; Moghnie & Kazarian, 2012; Nan et al. 2014; Spagnoli et al., 2012; Swami, 2008), others reported a trend toward

greater happiness among older participants (e.g., Extremera & Fernández-Berrocal, 2014; Simons et al., 2018; Vera-Villarroel et al., 2011). The inverted U-curve, which prevails in studies on the cognitive component of subjective well-being (Beja, 2018), was not identified in these SHS studies with one exception (i.e., Iani et al., 2014). Although previous studies have found different age trends for different components of subjective well-being (Stone et al., 2010), happiness and other components of subjective well-being do not usually stagnate, but may even increase with age, which is contrary to laypersons' beliefs (Lansford, 2018). As individuals age, they redefine their concepts of happiness by placing greater emphasis on positive emotional states, making their ideals more realistic and easier to achieve, and consequently experiencing greater subjective well-being (McMahon & Estes, 2012). In addition, shifting selective focus from negative stimuli to positive ones as one ages has been observed (Carstensen & DeLiema, 2018), which may explain why stagnation of happiness with age is not usually observed despite dealing with disappointments and challenges in life.

Compared with age differences, gender differences in subjective happiness have been examined more frequently. The results are consistent in not showing any gender differences (e.g., Chien et al., 2020; Moghnie & Kazarian, 2012; Swami, 2008; Swami et al., 2009). This finding could be the result of adaptation to an environment, regardless of the objective differences in this environment for women and men. In addition, men and women may evaluate their lives in comparison to people of the same sex, resulting in a lack of differences in subjective well-being (Batz & Tay, 2018).

Cross-Group and Cross-Time Measurement Comparability

A valid comparison of the subjective happiness levels across groups or over time requires the assumption of measurement invariance to be met. For cross-group comparisons, the scale items should have the same meaning for participants across different groups, and for cross-time comparisons, participants should ascribe the same meaning to the same items at different points-in-time. A lack of measurement invariance implies that the scale is not psychometrically equivalent for members of different groups or that respondents' idea of the construct changes over time, and so any observed group or time differences may result from measurement bias rather than real group or time differences (Milfont & Fischer, 2010). Although happiness can be considered a general human phenomenon, diverse cultural practices, beliefs, and values might result in different understanding of this construct for people from various cultures (e.g., Delle Fave et al., 2016; Kitayama & Markus, 2000; Swami et al., 2009), and these influences might affect the reliability and validity of the chosen measurement instrument. Furthermore, an individual's idea of happiness can change over time (e.g., Delle Fave et al., 2016; Mogilner et al., 2011), thus resulting in biased longitudinal comparisons. Therefore, establishing measurement invariance of the SHS across countries and over time is a necessary condition for valid cross-country comparisons and meaningful conclusions about changes in the levels of happiness over time.

Given the widespread use of the SHS in happiness research around the world, surprisingly few studies have examined the measurement invariance of the scale across different groups of participants using prevalent methods such as multiple group confirmatory factor analysis (MG-CFA). A study in a large Italian community sample (Iani et al., 2014) tested for configural (assuming equal factor structure), metric (assuming equal factor loadings), and residual invariance (additionally assuming equal unique item variances) of the SHS across gender and age groups. While all three increasingly stringent forms of invariance were confirmed for men and women, residual invariance was only partially supported for younger (18-44 years old) and older (45-85 years old) adults, with varying unique variance of the Item 1 across the two age groups. In another study, the SHS exhibited full measurement invariance, including residual invariance, across both gender groups in a Chinese college student sample (Chien et al., 2020). And third, configural invariance of the SHS was investigated and confirmed also across five age groups (25-29, 30-34, 35-39, 40-44, and 45-50 years old) in a large representative Portuguese sample, but more stringent types of invariance were not tested in this sample (Spagnoli et al., 2012).

Measurement invariance of the SHS across countries and cultures and across time has been less studied. To our knowledge, the only study that examined measurement invariance across countries was performed by Bieda et al. (2017). The invariance of the scale was investigated in German, Russian, and Chinese university student samples. While the configural invariance model fitted the data well, the metric invariance model resulted in attenuated fit to the data. Hence, partial metric invariance was established by relaxing constraints for factor loadings for Items 3 and 4. Scalar invariance (assuming equal item intercepts across groups) did not hold even after relaxing constraints for two items' intercepts. Bieda et al. (2019) were also the first to investigate longitudinal measurement invariance of the SHS, focusing on a large sample of Chinese university students. They confirmed scalar measurement invariance of the SHS over four annual measurements. A moderate stability in happiness levels was observed in Chinese students during this period (the autoregressive effects ranged from .27 to .47), despite the presumed changeable "state" nature of happiness (Veenhoven, 1994).

The Present Study

Bieda et al.'s (2017, 2019) studies, although laudable in many ways, involved data collected from university student samples which limits the generalizability of the obtained findings. Although both studies included a large number of participants, only three countries (Germany, Russia, and China) were compared in the study on cross-cultural measurement invariance of the SHS (Bieda et al., 2017), and the longitudinal invariance of the SHS was only investigated in a Chinese sample (Bieda et al., 2019), leaving much room for further investigations of the measurement invariance of the SHS across countries and time. In addition, previous investigations of measurement invariance of the SHS across gender and age groups (e.g., Iani et al., 2014; Spagnoli et al., 2012) have not been extensive and have not always examined all common forms of measurement invariance.

Therefore, the first objective of the present study was to examine whether measurement invariance could be established for the SHS across nine different countries, using samples of participants of various ages. The second objective was to investigate longitudinal (temporal) measurement invariance of the SHS on a subsample of participants from four different English-speaking countries over five consecutive time points. The third objective was to extend previous findings regarding measurement invariance of the SHS by gender and age groups. Fourth, we compared factor means across countries, gender, and age groups and over time, respectively. We expected no differences in factor means between men and women and higher factor means in older participants compared with younger participants. Since myriad factors could influence the level of happiness in different countries, cross-country differences in happiness levels were investigated from an exploratory perspective. Our final objective was to examine convergent validity of the SHS in relation to a measure of life satisfaction (the Temporal Satisfaction With Life Scale [TSWLS], Pavot et al., 1998) across countries, gender, and age groups.

Method

Participants and Procedure

The data reported in this article were collected by the International Well-being Study (www.wellbeingstudy.com), a cross-national longitudinal study of positive psychology constructs assessed from adults. Persons of 16 years of age or older were invited to participate in the study. Volunteers completed the questionnaires in their native language on the survey website. The data were collected longitudinally in five consecutive assessment points (every 3 months) over a 1-year time period. The first cohort of participants' recruitment began in March 2009 and the last cohort was recruited in March 2012. Data from the first assessment point for all cohorts were collated over time of measurement and they were used in this study to examine measurement invariance of the SHS across countries, gender, and age groups. The online survey required answers for all questions, so there were no missing values in the database.

For the purpose of the present study, only countries with 200 or more participants were considered. Samples of such size are required for adequate power to detect a lack of scalar invariance (MacCallum et al., 1996) and recommended for measures with a small number of factor indicators (Marsh et al., 1998). Altogether, nine countries met this criterion which together yielded a total sample of 4,977 participants. Table 1 presents the number of participants from each of these countries, mean age of the participants, proportion of women, and the language of the assessment instruments used. In all countries, women represented the major part of the sample. The total sample comprised 4,075 (82%) female participants and 902 (18%) male participants. Although gender distribution varied significantly across countries, $\chi^2(8) = 24.391$, $p = .002$, country effect on gender distribution was practically negligible (Table 1). Countries differed in the mean age of the participants, $F(8, 4968) = 96.12$, $p < .001$, with the oldest participants coming from Australia and New Zealand and the youngest coming from Slovenia and Czech Republic.[1] To examine measurement invariance across age, participants were divided into four age groups based on the developmental periods defined in the literature (Sigelman & Rider, 2018): 2,048 (41%) participants were 17 to 30 years (emerging adulthood), 1,631 (33%) were 31 to 45 years (early adulthood), 1,052 (21%) were 46 to 60 years (middle adulthood), and 246 (5%) were 61 to 75 years (late adulthood). Participants above 75 years were rare, so they were excluded from the analysis.

Longitudinal invariance of the SHS was examined on a subsample of 478 participants who answered the questions five consecutive times over a 1-year period of time. To make the analyses homogeneous for language, only participants from English-speaking countries were included in these analyses. In total, 254 (53.1%) participants came from New Zealand, 124 (25.9%) from the United States, 64 (13.4%) from Australia, and 36 (7.5%) from the United Kingdom. By sex, 410 (85.8%) were female and 68 (14.2%) were male. The sample's mean age was 42.8 years (SD = 13.8 years).

Measures

The total assessment battery included 20 questionnaires that measured various aspects of well-being. Besides the SHS (Lyubomirsky & Lepper, 1999), we used the TSWLS (Pavot et al., 1998) to assess convergent validity of the SHS.

Table 1. The Number of Participants From Different Countries, Their Age, Proportion of Women, and the Language of the Assessment Instruments.

		Age			
	N	*M*	*SD*	Female %	Language
Australia	321	42.80	15.05	82.9	English
Colombia	196	33.72	10.95	76.0	Spanish
Czech Republic	224	27.75	10.39	81.7	Czech
England (UK)	362	32.90	12.90	78.2	English
Hungary	1,043	31.58	11.14	83.8	Hungarian
Mexico	297	35.63	14.03	75.1	Spanish
New Zealand	1,530	38.66	12.34	83.9	English
Slovenia	199	25.61	9.48	80.9	Slovene
The United States	805	32.90	13.94	81.1	English
Total	4,977				

Subjective Happiness Scale. The SHS (Lyubomirsky & Lepper, 1999) was developed to assess global subjective happiness and consists of four items, rated on a 7-point Likert-type scale. The first two items require the respondents to characterize their general happiness (1 = *not a very happy person* to 7 = *a very happy person*) and their happiness relative to peers (1 = *less happy* to 7 = *more happy*). The other two items present a short description of generally happy and unhappy people, and respondents are asked to indicate to what extent each characterization describes them (1 = *not at all* to 7 = *a great deal*). Lyubomirsky and Lepper (1999) have examined reliability and construct validity of the scale on samples of varying ages and occupations, twelve from the United States and two from Russia. Good internal consistency reliability has been found in all samples with α ranging from .79 to .94. Test–retest reliability coefficients obtained with longitudinal data from five U.S. samples ranged from .55 to .90. Principal component analyses indicated a one-component structure of the scale in each of the fourteen samples.

Temporal Satisfaction With Life Scale. The TSWLS (Pavot et al., 1998) measures the cognitive component of subjective well-being. It was developed to assess individual's global judgment of life satisfaction with the focus on three specific time frames—past, present, and future. The scale includes a total of 15 items (5 per each time frame). Respondents use a 7-point Likert-type scale (1 = *strongly disagree* to 7 = *strongly agree*) to indicate the extent to which they agree with a specific item. Authors report α reliabilities for the total score ranging from .91 to .93 and the test–retest reliability coefficients between .82 and .88. A principal component analysis using varimax rotation showed a three-factor solution determined by the three time frames. The scale was further validated by McIntosh (2001), who reported that past, present, and future life satisfaction were distinct, yet correlated, factors.

Statistical Analysis

Preliminary analysis included a review of means, standard deviations, skewness, and kurtosis of the items and the total score across the examined country, gender, and age groups as well as across time. In addition, the reliability of the total score was assessed for each group and measurement point using Cronbach's alpha coefficient.

M*plus* version 8.2 (Muthén & Muthén, 1998-2018) was used to conduct confirmatory factor analysis (CFA) and multi-group CFA (MG-CFA). First, the factor structure of the SHS was examined separately in each country, gender, and age group and at each time point. A theoretically presupposed single-factor model was fitted to the data. For identification purposes, the factor variance was set to 1 to allow all item loadings to be estimated. The robust maximum likelihood (MLR) estimator was used, because the departure from multivariate normality, as evaluated by the Mardia's test (Korkmaz et al., 2014), was significant in all groups and measurement points (see also, the measures of skewness and kurtosis in Table S1 in the online supplemental materials). Then, the MG-CFA was used to examine measurement invariance of the SHS across each of the groups and across time. The *configural invariance model* was tested first in which the same factor structure was posited for all groups/time points, but factor loadings and item intercepts were allowed to vary between groups/time points. Next, we tested the fit of the *metric invariance model* with factor loadings held equal across groups/time points, whereas intercepts were allowed to vary across groups/time points. Finally, *scalar invariance model* was assessed with both factor loadings and item intercepts constrained to be equal across all groups/time points. Strict invariance which additionally assumes equal residual variance of each item across groups/time points was not tested because scalar invariance is a sufficient condition for the comparisons of latent means across groups/time points (Meredith, 1993). In

longitudinal models, the latent factors were allowed to covary across all time points. Similarly, the corresponding item residuals were allowed to covary across measurement occasions.

To adjust for the unequal sample sizes that could lead to invalid conclusions due to the larger group overdetermining the fit function in the MG-CFA analysis, tests of measurement invariance across gender were complemented by a Monte Carlo simulation technique proposed by Yoon and Lai (2018). From the larger group of women, 1,000 multiple random samples were drawn, which corresponded to the size of the smaller group of men ($n = 902$). Tests of measurement invariance were performed 1,000 times, using the subsamples from the female group together with the sample of male participants. The mean value of each fit statistic was calculated over 1,000 replications.

Model fit was evaluated using both absolute fit indices and relative fit indices. Absolute fit indices included the Satorra–Bentler scaled chi-square test statistic ($SB\chi^2$), the root mean square error of approximation (RMSEA) with 90% confidence interval (CI), and the standardized root mean square residual (SRMR). RMSEA values lower than .06 and SRMR values lower than .08 are often considered to indicate good model fit (Hu & Bentler, 1999), while RMSEA values lower than .08 are considered to indicate reasonable model fit (Browne & Cudeck, 1992). Among incremental fit indices, the comparative fit index (CFI) was used. We considered values above .95 to indicate good model fit (Browne & Cudeck, 1992). Following recommendations by Chen (2007), we used $\Delta RMSEA \leq .015$, $\Delta CFI \leq -.010$, and $\Delta SRMR \leq .030$ for the evaluation of the fit of the successive models with increasingly stringent constraints. We interpreted the value of the $SB\chi^2$ with caution since with large samples, the chi-square likelihood ratio tests may give significant results even with practically negligible deviations from invariance (e.g., Cheung & Rensvold, 2002). Note that the subsampling approach that we used to adjust for the unequal size of the two gender groups does not allow the calculation of the chi-square difference test statistics, so in this case the fit of the models with increasingly stringent constraints was evaluated based on the changes in the approximate fit indices. If the scale would not achieve metric or scalar invariance, we intended to examine modification indices to identify noninvariant items and establish partial measurement invariance as proposed by Byrne et al. (1989).

After measurement invariance had been established across countries, gender, and age groups and over time, latent mean differences were compared between the members of these groups, respectively, and between the assessment points. Latent means were constrained to zero in one of the countries, gender, and age groups, respectively, and at one assessment point, and they were freely estimated in each of the remaining groups/time points. The freely estimated latent means in each target group are direct estimates of the difference from the reference group/time point, expressed in units of standard deviation. Wald tests were used to examine the significance of the latent mean differences in comparison to the reference group/time point. All other pairs of groups/time points were compared with one another using the MODEL CONSTRAINT command in M*plus*. Due to the exploratory nature of the comparisons, the Holm-Bonferroni correction (Holm, 1979) was used to counterbalance the problem of multiple comparisons between country groups and assessment points.

To examine convergent validity of the SHS, Pearson's correlations were computed between the SHS score and the scores from the TSWLS (past, present, and future life satisfaction), separately for each country, gender, and age group. The correlations were then transformed into Fisher's *z*-values and Cohen's *q* effect sizes were used to compare differences in the magnitude of correlations between pairs of countries, genders, and age groups. Values of .10, .30, and .50 were interpreted as small, medium, and large effect sizes, respectively (Cohen, 1988).

Results

Descriptive statistics of the SHS items and the total score across countries, gender, and age groups and over time are presented in Table S1 in the online supplemental materials. Across all groups/time points, the distributions of items and the total score were slightly skewed to the left, however most of the skewness values were reasonably low. Kurtosis values were more variable across groups/time points, but also rarely exceeded an absolute value of 1 for the four items and they were at most .45 for the total score. Generally, the corrected item-total correlations were the highest for the first item and the lowest for the last, fourth item. Alpha reliability coefficients exceeded .80 in almost all groups and at all assessment points, the two exceptions being the Mexican and Colombian samples (Table S2, available in the online supplemental materials).

Measurement Invariance of the SHS

Before examining measurement invariance of the SHS across countries, the hypothesized single-factor model was tested for each country separately. Although RMSEA values fell below .08 in all but one country, thus indicating reasonable model fit, RMSEA CIs were generally large and included values above .10 for five of the nine countries. However, RMSEA tends to be positively biased (i.e., it tends to falsely indicate a poor fitting model) and RMSEA CIs tend to be wider in smaller models with lower degrees of freedom (Kenny et al., 2015). Therefore, based on the CFI and SRMR values, the fit of the single-factor model was considered good in all countries (Table 2). Factor

Table 2. Confirmatory Factor Analysis Fit Statistics for the Single-Factor SHS Model by Country.

Country	$SB\chi^2_{(df)}$	*p*	RMSEA	90% CI	CFI	SRMR
Australia	$0.623_{(2)}$	.733	.000	[.000, .078]	1.000	.006
Colombia	$0.960_{(2)}$	.619	.000	[.000, .114]	1.000	.015
Czech Republic	$5.108_{(2)}$	.078	.083	[.000, .176]	.989	.025
England (UK)	$4.689_{(2)}$	.096	.061	[.000, .135]	.994	.022
Hungary	$1.717_{(2)}$	.424	.000	[.000, .059]	1.000	.004
Mexico	$1.677_{(2)}$	.432	.000	[.000, .109]	1.000	.011
New Zealand	$11.212_{(2)}$	.004	.055	[.027, .088]	.996	.010
Slovenia	$4.239_{(2)}$	.120	.075	[.000, .176]	.993	.018
The United States	$5.226_{(2)}$	.078	.045	[.000, .094]	.998	.010

Note. SHS = Subjective Happiness Scale; $SB\chi^2$ = Satorra–Bentler scaled chi-square test statistic; *df* = degrees of freedom; RMSEA = root mean square error of approximation; CI = confidence interval; CFI = comparative fit index; SRMR = standardized root mean square residual.

Table 3. The Comparison of Configural, Metric, Scalar, and Partial Scalar Invariance Models by Country.

	Goodness-of-fit						Model comparison					
	$SB\chi^2_{(df)}$	*p*	RMSEA	90% CI	CFI	SRMR	Ref. model	$\Delta SB\chi^2_{(df)}$	*p*	ΔRMSEA	ΔCFI	ΔSRMR
M1: Configural	$35.266_{(18)}$	.009	.042	[.020, .062]	.997	.012						
M2: Metric	$87.445_{(42)}$	.000	.044	[.031, .057]	.993	.048	M1	$52.425_{(24)}$	.001	.002	−.004	.036
M3: Scalar	$355.423_{(66)}$	.000	.089	[.080, .098]	.955	.056	M2	$284.009_{(24)}$	<.001	.045	−.038	.008
M4: Partial scalar	$141.693_{(58)}$	.000	.051	[.040, .062]	.987	.049	M2	$56.511_{(16)}$	<.001	.007	−.006	.001

Note. Ref. model = reference model; $\Delta SB\chi^2$, ΔCFI, and ΔRMSEA = change in fit indices between contiguous nested models; $SB\chi^2$ = Satorra–Bentler scaled chi-square test statistic; *df* = degrees of freedom; RMSEA = root mean square error of approximation; CI = confidence interval; CFI = comparative fit index; SRMR = standardized root mean square residual.

loadings of all items in each country are presented in Table S2 in online supplemental materials. Loadings were high for the first three items and somewhat lower for the fourth item, but exceeded .40 in all countries except Colombia. McDonald's ω reliability coefficients were also high, exceeding .80 in all samples but two (i.e., the Mexican and Colombian samples), where they were above .70.

After the good-fitting baseline models had been established for each country separately, they were combined in a multiple-group model and tested for configural invariance. The configural invariance model provided good fit to the data (Table 3), suggesting that the factor structure was equivalent across all groups tested. The metric invariance model, which postulates invariant factor loadings across groups, fitted the data well, and the deterioration in model fit compared with the configural model fell within the change criteria recommended for the RMSEA and CFI values, although the change in SRMR fell slightly above the recommended cutoff value (Chen, 2007). The $SB\chi^2$ was significant, but this was likely due to a large sample size in this study. The scalar invariance model, which assumes equal item intercepts across all groups, however failed to fit the data well and showed decrement in model fit compared with the less restrictive metric invariance model, indicated by ΔRMSEA and ΔCFI values. A review of the modification indices suggested that relaxing the intercept of the Item 2 should increase the model fit (Byrne et al., 1989). Accordingly, a partial scalar invariance model in which the intercept of Item 2 was relaxed fitted the data well and the deterioration in model fit compared with the metric invariance model fell in the acceptable range.

We further examined measurement invariance of the SHS across gender and age groups using the aggregated data from the nine countries (Table 4). Again, the baseline models were first established by examining the fit of the single-factor model in each gender and age group, respectively. For both gender groups and for all age groups, this model provided good fit to the data. The RMSEA value was higher for younger groups with larger sample sizes, but remained in the reasonable range. Factor loadings for the four items exceeded .40 in all gender and age groups, with the highest loadings for Item 1 and the lowest loadings for Item 4. McDonald's ω reliability coefficients were above .80 in all groups (Table S2, available in online supplemental materials).

The configural invariance model for gender showed good fit to the data (Table 5). The more restrictive metric invariance model yielded slightly better fit as compared with the configural invariance model. Similarly, the scalar invariance model showed a slight improvement in model fit compared with the metric invariance model according to ΔRMSEA and ΔSRMR values, whereas the fit was equal

Table 4. Confirmatory Factor Analysis Fit Statistics for the Single-Factor SHS Model by Gender and Age.

	$SB\chi^2_{(df)}$	*p*	RMSEA	90% CI	CFI	SRMR
Gender						
Female	$15.771_{(2)}$	.000	.041	[.024, .061]	.997	.009
Male	$7.602_{(2)}$	.022	.056	[.018, .100]	.995	.013
Age						
17-30 Years	$17.149_{(2)}$	.000	.061	[.036, .089]	.995	.014
31-45 Years	$6.400_{(2)}$	.041	.037	[.007, .070]	.998	.008
46-60 Years	$4.097_{(2)}$	.129	.032	[.000, .076]	.999	.007
61-75 Years	$0.436_{(2)}$	.866	.000	[.000, .079]	1.000	.004

Note. SHS = Subjective Happiness Scale $SB\chi^2$ = Satorra–Bentler scaled chi-square test statistic; *df* = degrees of freedom; RMSEA = root mean square error of approximation; CI = confidence interval; CFI = comparative fit index; SRMR = standardized root mean square residual.

Table 5. The Comparison of Configural, Metric, and Scalar Invariance Models by Gender and Age.

	Goodness-of-fit						Model comparison					
	$SB\chi^2_{(df)}$	*p*	RMSEA	90% CI	CFI	SRMR	Ref. model	$\Delta SB\chi^2_{(df)}$	*p*	ΔRMSEA	ΔCFI	ΔSRMR
Gender												
M1g: Configural	$24.230_{(4)}$	.000	.045	[.029, .063]	.997	.010						
M2g: Metric	$29.859_{(7)}$	.000	.036	[.023, .050]	.996	.018	M1g	$5.739_{(3)}$	.125	−.009	−.001	.008
M3g: Scalar	$35.852_{(10)}$	.000	.032	[.021, .044]	.996	.017	M2g	$5.324_{(3)}$	.150	−.004	.000	−.001
Gender: Monte Carlo simulation												
M1gMC: Configural	$12.562_{(4)}$	—	.047	—	.996	.012						
M2gMC: Metric	$17.237_{(7)}$	—	.039	—	.996	.026	M1gMC	—	—	−.008	.000	.014
M3gMC: Scalar	$22.041_{(10)}$	—	.035	—	.995	.024	M2gMC	—	—	−.004	−.001	−.002
Age												
M1a: Configural	$29.702_{(8)}$	.000	.047	[.029, .065]	.997	.011						
M2a: Metric	$51.331_{(17)}$	.000	.040	[.028, .053]	.995	.031	M1a	$21.487_{(9)}$	.011	−.007	−.002	.020
M3a: Scalar	$96.905_{(26)}$	.000	.047	[.037, .057]	.989	.033	M2a	$46.444_{(9)}$	<.001	.007	−.006	.002

Note. $SB\chi^2$ = Satorra–Bentler scaled chi-square test statistic; *df* = degrees of freedom; RMSEA = root mean square error of approximation; CI = confidence interval; CFI = comparative fit index; SRMR = standardized root mean square residual.

according to ΔCFI value. The results of the measurement invariance testing for gender using Yoon and Lai's method (2018) for unbalanced samples were very similar to those obtained with unequal sample sizes of the two gender groups. The configural invariance model also provided good fit to the data for the four age groups (Table 5). Again, metric invariance testing resulted in slightly improved model fit compared with the configural invariance model, at least as according to the ΔRMSEA value. The scalar invariance model also fitted the data well. The deterioration in model fit compared with the metric invariance model was small and within the acceptable range.

Longitudinal measurement invariance of the SHS was examined on a subsample of 478 participants from English-speaking countries and over five consecutive assessment points with 3-month intervals. First, the fit of the single-factor model was examined for each assessment point separately. All models showed good fit to the data, with somewhat higher RMSEA values and CIs for the second and the fourth assessment point (Table 6). Also, as was previously noted, factor loadings were high for the first three items and somewhat lower for the fourth item, but in all cases exceeded .50. McDonald's ω reliability coefficients were high, approaching .90 (Table S2, available in online supplemental materials).

The results of the longitudinal measurement invariance testing are presented in Table 7. Configural, metric, and scalar invariance models fitted the data well, with no significant deterioration, but rather a mild improvement in the fit for the more restrictive models compared with the less restrictive models as indicated by ΔRMSEA and ΔCFI values.

Cross-Group and Cross-Time Comparisons of Subjective Happiness Levels

Establishing measurement invariance across countries, gender, and age groups and across time allowed us to examine differences in the latent means between the members of these groups and between the assessment points. Given a

Table 6. Confirmatory Factor Analysis Fit Statistics for the Single-Factor SHS Model by Time.

	$SB\chi^2_{(df)}$	*p*	RMSEA	90% CI	CFI	SRMR
Time 1	$1.426_{(2)}$	.490	.000	[.000, .082]	1.000	.007
Time 2	$7.216_{(2)}$	.027	.074	[.021, .135]	.994	.012
Time 3	$1.074_{(2)}$	.585	.000	[.000, .076]	1.000	.004
Time 4	$1.498_{(2)}$	.473	.000	[.000, .083]	1.000	.005
Time 5	$5.884_{(2)}$	.053	.064	[.000, .126]	.996	.009

Note. SHS = Subjective Happiness Scale; $SB\chi^2$ = Satorra–Bentler scaled chi-square test statistic; *df* = degrees of freedom; RMSEA = root mean square error of approximation; CI = confidence interval; CFI = comparative fit index; SRMR = standardized root mean square residual.

Table 7. The Comparison of Configural, Metric, and Scalar Invariance Models by Time.

	Goodness-of-fit						Model comparison					
	$SB\chi^2_{(df)}$	*p*	RMSEA	90% CI	CFI	SRMR	Ref. model	$\Delta SB\chi^2_{(df)}$	*p*	ΔRMSEA	ΔCFI	ΔSRMR
M1: Configural	$159.205_{(120)}$	.010	.026	[.013, .036]	.995	.022						
M2: Metric	$172.854_{(132)}$	.010	.025	[.013, .035]	.995	.028	M1	$13.651_{(12)}$	.324	−.001	.000	.006
M3: Scalar	$180.236_{(144)}$	.022	.023	[.009, .033]	.996	.029	M2	$6.659_{(12)}$	.879	−.002	.001	.001

Note. $SB\chi^2$ = Satorra–Bentler scaled chi-square test statistic; *df* = degrees of freedom; RMSEA = root mean square error of approximation; CI = confidence interval; CFI = comparative fit index; SRMR = standardized root mean square residual.

significant average within-country correlation between age and SHS score ($r = .13$, $p < .001$), differences in latent means between countries were examined controlling for age. Specifically, age was entered as a covariate in a measurement invariance model, based on the previously established partial scalar invariance model. Since the regression coefficients for age predicting subjective happiness were largely similar across most of the countries (Australia, Colombia, Mexico, New Zealand, England, and the United States), we constrained them to equality across these countries. In the other three countries (Czech Republic, Hungary, and Slovenia), the regression weights were very low, so we constrained them to zero. It is notable that the first group includes English-speaking and/or American countries, whereas the other group is only constituted by East European countries. These constraints resulted in an equally well fitting model, $SB\chi^2(93) = 251.565$, $p < .001$, CFI = .980, RMSEA = .056, CFI = .980, SRMR = .048, compared with the model with freely estimated regression coefficients in all countries, $SB\chi^2(85) = 245.382$, $p < .001$, CFI = .979, RMSEA = .058, CFI = .979, SRMR = .047. The latent mean was set to zero in the Australian sample and freely estimated in the rest of the samples. The resulting mean values are presented in Table 8 arranged by size. The highest subjective happiness mean values were obtained for Czech Republic, Mexico, and Slovenia, and the lowest mean values were obtained for England, New Zealand, and Australia.

Latent means were also compared across gender and age groups, respectively (Table 8). The latent mean was constrained to zero for the female group and freely estimated for the male group. The two means did not differ significantly from one another. Similarly, the latent mean was set to zero for the youngest group (17-30 years) and freely estimated for the older age groups. The latent mean of each consecutive age group was significantly higher compared with the latent mean of the previous age group.

Finally, latent means were compared over five assessment points within a 1-year period. The first assessment point was chosen as a reference time point, and its latent mean was set to zero. None of the differences between the assessment points remained significant after the Holm–Bonferroni correction (Table 8).

Convergent Validity of the SHS Against TSWLS

To assess the convergent validity of the SHS, the present study examined the relationship between the SHS and TSWLS across countries, gender, and age groups. All correlations were positive and significant (Table 9). In most of the groups, the SHS had the largest correlation with present life satisfaction as was expected since the SHS measures present feelings of happiness, as opposed to past or future happiness. Cohen's *q* effect sizes (Tables S3-S5, available in the online supplemental materials) showed either no or small effects for the differences in the size of the correlations with present life satisfaction within pairs of countries, with the exception of medium effects for the differences in the correlations between the Slovenian sample and the Colombian and Czech samples, whereby the Slovenian sample showed the highest correlation with the present life satisfaction. The effect sizes were below .10 for the comparison between gender and age groups, respectively,

Table 8. Latent Mean Values by Country, Gender, Age, and Time Groups, and the Comparison of the Latent Mean Values.

	Latent mean	Groups with significantly lower mean
Country		
England (UK)	−.178	
New Zealand	−.042	
Australia	.000	
The United States	.072	England (UK)
Colombia	.202	England (UK), New Zealand
Hungary	.291	England (UK), New Zealand, Australia, The United States
Slovenia	.397	England (UK), New Zealand, Australia, The United States
Mexico	.420	England (UK), New Zealand, Australia, The United States
Czech Republic	.547	England (UK), New Zealand, Australia, The United States, Colombia, Hungary
Gender		
Male	−.071	
Female	.000	
Age		
17-30 Years	.000	
31-45 Years	.121	17-30
46-60 Years	.223	17-30, 31-45
61-75 Years	.598	17-30, 31-45, 46-60
Time		
Time 1	.000	
Time 2	.025	
Time 3	.003	
Time 4	.065	
Time 5	.054	

Table 9. Pearson Correlations Between SHS (Subjective Happiness) and TSWLS (Satisfaction With Life—Past, Present, Future) by Country, Gender, and Age Groups.

	TSWLS—past	TSWLS—present	TSWLS—future
Country			
Australia	.43	.62	.49
Colombia	.26	.53	.29
Czech Republic	.50	.48	.25
England (UK)	.52	.63	.56
Hungary	.55	.58	.37
Mexico	.45	.64	.35
New Zealand	.46	.57	.48
Slovenia	.52	.71	.51
The United States	.42	.58	.54
Gender			
Female	.45	.59	.46
Male	.51	.61	.47
Age			
17-30 Years	.51	.59	.44
31-45 Years	.47	.57	.46
46-60 Years	.45	.62	.55
61-75 Years	.42	.59	.49

Note. All correlations are significant at $p < .01$. TSWLS = Temporal Satisfaction With Life Scale.

suggesting that this covariance did not vary markedly among these different groups.

The correlations between the SHS and past life satisfaction were medium to large in size, with the exception of the Colombian sample where this correlation was small. Cohen's *q* effect sizes for the differences in the size of correlations with past life satisfaction between pairs of countries were mostly small (i.e., below .10). Somewhat higher, although still small (i.e., up to .20), were the effect sizes for the differences in correlations between the Colombian sample and other country samples; the Colombian sample yielded the lowest correlation with past life satisfaction. The effect sizes for the differences in correlations between gender and age groups, respectively, were again negligible.

The correlations with future life satisfaction were more variable. For all gender and age groups and most of the country groups, these correlations were medium to large in magnitude, but they were small in the Czech and Colombian samples. Again, Cohen's *q* effect sizes were small to medium for differences in the size of correlations with future life satisfaction between pairs of countries, and they were below .10 for differences in correlations between gender and age groups, respectively.

Discussion

Our main goal in the present study was to rigorously examine the psychometric properties of the SHS by investigating measurement invariance of the scale across countries, gender and age groups, and over time. The testing of measurement invariance allowed us to determine whether understanding of the measure varied across different groups of participants and over multiple assessment points. Additionally, our goal was to compare happiness levels across countries, gender, and age groups and over time, respectively, and to examine convergent validity of the SHS in relation to a measure of life satisfaction.

Measurement Invariance of the SHS

As a prestep to measurement invariance examination, the hypothesized single-factor model of the SHS was tested and largely confirmed for each group/time point separately. The unidimensional structure of the SHS has already been established in previous studies (e.g., Doğan & Totan, 2013; Extremera & Fernández-Berrocal, 2014; Kotsou & Leys, 2017; Nan et al., 2014) where metric characteristics of the SHS were usually examined in a single country or language. These studies, similar to our findings, reported somewhat lower factor loadings for Item 4 (the lowest was found in the Turkish samples; Doğan & Totan, 2013) that has been attributed to the reversed wording of this item (Lindwall et al., 2012).

Measurement invariance of the SHS was first examined across countries. The results of the three stages of measurement invariance testing revealed that configural and metric invariances held, but we could only achieve partial scalar invariance because Item 2 yielded varying intercepts across countries. Item 2 asks respondents to compare their happiness with that of their peers, thus requiring a comparison with a specific reference group (as opposed to the other three items), which could be the reason for different understandings of this item. In addition, the meaning of the term "peers" might vary across languages. For example, in Slovene, the term *peers* ("vrstniki") is usually used to refer to children or adolescents of a similar age, but it is rarely used for adults. "Peers" of adults in different countries might also vary in quality and quantity (e.g., Heine et al., 2002), posing further risk to equal understandings of this item across counties. According to the literature, at least two invariant items are required for a valid factor mean comparison (Steenkamp & Baumgartner, 1998). Therefore, the chief implication of our results is that SHS latent means can be meaningfully compared between countries. Our results, which indicate partial scalar invariance of the SHS across nine countries, complement findings from the only previous study on cross-country measurement invariance of the SHS (Bieda et al., 2017), in which only partial metric invariance was achieved in samples of university students from three countries.

Measurement invariance of the SHS was further examined and fully confirmed across gender and age groups, respectively, suggesting that individuals of both genders and different ages understood the SHS items in very similar ways. Our results are consistent with previous studies showing measurement invariance of the SHS across gender in an Italian (Iani et al., 2014) and a Chinese (Chien et al., 2020) sample, thus indicating very similar structure or/and meaning of the SHS items for males and females. Measurement invariance across age groups was examined in two previous studies. Spagnoli et al. (2012) reported configural invariance across five age groups in a Portuguese adult sample, however, other types of invariance were not tested in their study. Iani et al. (2014) examined invariance of the SHS factor loadings and error variances for two Italian age groups (18-44 years and 45-85 years), and found partial invariance after freeing up the error variance of Item 1. Our results suggest that the SHS evidences full measurement invariance across different age groups in adulthood, and they also strengthen our confidence in the cross-country measurement invariance results, considering the differences in the mean ages of the participants from the nine countries of our sample.

Our results also demonstrated full measurement invariance of the SHS over five consecutive assessment points, suggesting that the happiness construct was measured similarly by the SHS over time, at least over a 1-year period.

Apparent changes in SHS scores at different occasions can thus be attributed to actual changes in the levels of subjective happiness. These findings are consistent with the only previous study on the longitudinal measurement invariance of the SHS, which showed that the scale was invariant across four measurement points with 1-year intervals (Bieda et al., 2019). In Bieda et al.'s study, a relatively large sample of university students from a single cultural environment was used (i.e., Chinese). Therefore, our results extend previous findings on the longitudinal measurement invariance of the SHS to English-speaking samples from Australian, American, and European cultural contexts and to participants of various ages.

Cross-Group and Cross-Time Comparisons of Subjective Happiness Levels

Cross-country comparison of SHS latent mean values revealed lower subjective happiness levels in English-speaking countries (England, New Zealand, Australia, and the United States) compared with Hungary, Slovenia, Mexico, and Czech Republic. Although previous studies have shown that a country's economic conditions are related to the life satisfaction of its citizens, this relationship has not been observed for emotional well-being (Diener et al., 2010). It is therefore notable that in our study, happiness, which includes not only a cognitive component (life satisfaction) but also an emotional component (positive/negative affect), does not show the same pattern as one would expect based on previous cross-country comparisons of life satisfaction. In particular, the Latin American countries did not follow the proposed relationship between GDP and happiness in our study (Mexico and Colombia) as well as in a previous study (Rojas, 2012), as they have high happiness scores despite lower GDP. In our study, participants from East European countries (Czech Republic, Slovenia, Hungary) reported the highest happiness values among the nine countries. Although these countries are traditionally collectivist, they have experienced rapid socioeconomic changes in individualistic directions (Cheng et al., 2011), which may be the reason for their higher happiness scores compared with traditionally individualistic cultures such as the United Kingdom, New Zealand, Australia, and the United States.

SHS latent means were similar for male and female participants. Previous studies also reported nonsignificant gender differences in SHS scores (e.g., Iani et al., 2014; Lyubomirsky & Lepper, 1999; Spagnoli et al., 2012; Vera-Villarroel et al., 2011). A lack of gender differences could be the result of men and women adapting to their respective environments, even though these environments may be objectively different. Few and small gender differences could also be the result of evaluative standards, when individuals compare themselves with other people of the same sex (Batz & Tay, 2018).

Regarding age differences in subjective happiness levels, previous studies reported either no age trends (Doğan & Totan, 2013; Lyubomirsky & Lepper, 1999; Nan et al. 2014; Spagnoli et al., 2012; Swami, 2008; Moghnie & Kazarian, 2012) or a positive relationship between SHS scores and age (Extremera & Fernández-Berrocal, 2014; Simons et al., 2018; Vera-Villarroel et al., 2011). Our results are consistent with the theory of the positivity effect, which states that compared with younger adults, older adults attend to and remember positive information better than negative information (Carstensen & DeLiema, 2018), their goals become more realistic and easier to achieve, and consequently they may experience greater subjective well-being (McMahon & Estes, 2012).

We found no differences in subjective happiness levels over five assessment points within a 1-year period. These findings are consistent with Bieda et al. (2019), who reported relative stability of SHS scores in Chinese university students over a 4-year time span. Our study, conducted over one year, does not elucidate changes over the lifespan, but our results on the measurement invariance of the SHS across time and age groups suggest that SHS can be meaningfully used to examine stability/malleability in happiness levels over the life course.

Convergent Validity of the SHS against TSWLS

Convergent validity of the SHS was examined against a measure of life satisfaction, the TSWLS. The correlations between the SHS and the TSWLS were similar to those established in previous studies with the SWLS (e.g., Extremera & Fernández-Berrocal, 2014; Jovanović, 2013; Lyubomirsky & Lepper, 1999; Spagnoli et al., 2012; Szabo, 2019). They show good convergent validity of the SHS across different groups, with little variation in the magnitude of correlations between groups. Although life satisfaction as a cognitive assessment of one's own life is probably the most commonly used operationalization of an individual's happiness (Carlquist et al., 2017), these correlations suggest that life satisfaction as a cognitive component of subjective well-being cannot encompass the full meaning of happiness, in which the emotional component is also important.

Limitations and Future Directions

Our study includes several limitations. First, participants were obtained through convenience sampling and not representatively drawn from the population. One obvious manifestation was the asymmetrical ratio of male and female participants, which limits the representativeness of our results and their interpretation. Samples from different countries were not equal in their sizes and some samples were relatively small. When testing the measurement invariance, we only adjusted for the uneven distribution of participants across two gender groups. We acknowledge that unbalanced sample sizes might lead to inaccurate results of factorial invariance even across countries and age

groups. However, Yoon and Lai (2018) tested their method for two groups only, emphasizing that the effects of unbalanced group sizes on factorial invariance may be more complex in studies with more than two groups than in studies with only two groups, and that this remains an open issue to be addressed in future studies. To obtain more robust conclusions, future studies should focus on representative and similarly large samples from different nations.

Although the results indicate that the SHS scores can be meaningfully compared across countries, only European, American, and Australian countries were included in the present study, thus measurement invariance of the SHS on samples from Asia and Africa will need to be examined in future studies. In addition, longitudinal invariance was investigated only on a subsample of English-speaking participants and over short time intervals of three months, suggesting the need for further replications of our findings in different language contexts and over larger time intervals (e.g., Bieda et al., 2019).

Finally, we note that although the approximate fit indices (RMSEA, CFI, and SRMR) generally indicated a good fit of the single-factor model, the χ^2 test statistics were statistically significant in many cases. This outcome can be partly attributed to the large sample size and the high statistical power of our data set, so that even small deviations from perfect unidimensionality could be detected. Many authors (e.g., McCrae et al., 1996; Muthén & Asparouhov, 2012) have already noted that the factor analysis model is overly restrictive for the item-analysis level. We nevertheless decided to opt for the confirmatory factor analysis framework because of its versatility and because most fit indices invariably indicated a reasonably good fit of the configural model.

Implications and Conclusion

In terms of practice and assessment implications, our results suggest first that, since full metric invariance with equal item loadings was established, structural relationships between subjective happiness and other relevant constructs can be reliably and validly compared between countries, gender and age groups, and over time. Second, the full scalar invariance finding enables researchers to confidently compare results between gender and age groups both concurrently as well as longitudinally, thus also allowing observations of developmental trends over time. Cross-country comparisons of subjective happiness are also possible, albeit with minor caution due to the small degree of intercept noninvariance between countries. Therefore, cross-cultural researchers should either compare latent means between countries or consider excluding the noninvariant Item 2 from direct mean comparisons, although later practice could pose a certain risk to content coverage and thus the construct validity of the measurement. The finding on the cross-cultural invariance of SHS is likely to be a useful contribution to the growing interest in cross-country comparisons of happiness levels, where most often single-item measures of happiness are used. Single-item measurement does not allow for the psychometric assessment of measurement characteristics, except for the retest reliability. Our analyses also showed that in all groups studied and over time, Item 4 was the weakest of the four SHS indicators of subjective happiness, although it was still an adequate indicator. Since Item 4 is the only reverse-scored item in the SHS, the lowest loadings for this item could result from individual differences in acquiescent response bias combined with poor reading of the items. Finally, for practitioners using SHS in individual diagnostics, there is the important implication that while the same norms could be used for both genders, separate norms should be developed and used for people from different countries and of different ages.

In conclusion, the present study's findings further our understanding of the measurement invariance of the SHS. The scale proved to exhibit good invariant factor structure in nine countries (the one exception was noninvariance for a single item's intercept) as well as in both gender groups, in different age groups, and over time. Constituted with four items, the SHS represents a significant improvement over one-item happiness measures that are often used in large population surveys.

Acknowledgments

Appreciation is expressed to Aaron Jarden and the IWS research team (Kennedy McLachlan, Paul Jose, Alex Mackenzie, Ormond Simpson, and Todd Kashdan).

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: This work was supported from the Slovenian Research Agency (research core funding No. P5-0110 and P5-0062).

ORCID iD

Gaja Zager Kocjan https://orcid.org/0000-0002-1934-2831

Supplemental Material

Supplemental material for this article is available online.

Note

1. A review of countries by age (http://worldpopulationreview.com/) reveals that these differences do not stem from actual age differences between the respective countries but can rather be attributed to convenience sampling.

References

Batz, C., & Tay, L. (2018). Gender differences in subjective well-being. In E. Diener, S. Oishi, & L. Tay (Eds.), *Handbook of well-being*. DEF Publishers.

Beja, E. L. (2018). The U-shaped relationship between happiness and age: Evidence using World Values survey data. *Quality & Quantity*, *52*(4), 1817-1829. https://doi.org/10.1007/s11135-017-0570-z

Bieda, A., Hirschfeld, G., Schönfeld, P., Brailovskaia, J., Lin, M., & Margraf, J. (2019). Happiness, life satisfaction and positive mental health: Investigating reciprocal effects over four years in a Chinese student sample. *Journal of Research in Personality*, *78*(February), 198-209. https://doi.org/10.1016/j.jrp.2018.11.012

Bieda, A., Hirschfeld, G., Schönfeld, P., Brailovskaia, J., Zhang, X. C., & Margraf, J. (2017). Universal happiness? Cross-cultural measurement invariance of scales assessing positive mental health. *Psychological Assessment*, *29*(4), 408-421. https://doi.org/10.1037/pas0000353

Bradburn, N. M. (1969). *The structure of psychological well-being*. Alpine. https://doi.org/10.1037/t10756-000

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research*, *21*(2), 230-258. https://doi.org/10.1177/0049124192021002005

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*(3), 456-466. https://doi.org/10.1037/0033-2909.105.3.456

Carlquist, E., Ulleberg, P., Delle Fave, A., Nafstad, H. E., & Blakar, R. M. (2017). Everyday understandings of happiness, good life, and satisfaction: Three different facets of well-being. *Applied Research in Quality of Life*, *12*(2), 481-505. https://doi.org/10.1007/s11482-016-9472-9

Carstensen, L. L., & DeLiema, M. (2018). The positivity effect: A negativity bias in youth fades with age. *Current Opinion in Behavioral Sciences*, *19*(February), 7-12. https://doi.org/10.1016/j.cobeha.2017.07.009

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464-504. https://doi.org/10.1080/10705510701301834

Cheng, C., Jose, P. E., Sheldon, K. M., Singelis, T. M., Cheung, M. W. L., Tiliouine, H., Alao, A. A., Chio, J. H. M., Lui, J. Y. M., Chun, W. Y., de Zavala, A. G., Hakuzimana, A., Hertel, J., Liu, J.-T., Onyewadume, M., & Sims, C. (2011). Sociocultural differences in self-construal and subjective well-being: A test of four cultural models. *Journal of Cross-Cultural Psychology*, *42*(5), 832-855. https://doi.org/10.1177/0022022110381117

Cheung, F., & Lucas, R. E. (2014). Assessing the validity of single-item life satisfaction measures: Results from three large samples. *Quality of Life Research*, *23*(10), 2809-2818. https://doi.org/10.1007/s11136-014-0726-4

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233-255. https://doi.org/10.1207/S15328007SEM0902_5

Chien, C.-L., Chen, P.-L., Chu, P.-J., Wu, H.-Y., Chen, Y.-C., & Hsu, S.-C. (2020). The Chinese version of the Subjective Happiness Scale: Validation and convergence with multidimensional measures. *Journal of Psychoeducational Assessment*, *38*(2), 222-235. https://doi.org/10.1177/0734282919837403

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum.

Damásio, B. F., Zanon, C., & Koller, S. H. (2014). Validation and psychometric properties of the Brazilian version of the Subjective Happiness Scale. *Universitas Psychologica*, *13*(1), 17-24. https://doi.org/10.11144/Javeriana.UPSY13-1.vppb

Delle Fave, A., Brdar, I., Wissing, M. P., Araujo, U., Castro Solano, A., Freire, T., Del Rocío Hernández-Pozo, M., Jose, P., Martos, T., Nafstad, H. E., Nakamura, J., Singh, K., & Soosai-Nathan, L. (2016). Lay definitions of happiness across nations: The primacy of inner harmony and relational connectedness. *Frontiers in Psychology*, *7*, Article 30. https://doi.org/10.3389/fpsyg.2016.00030

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction with Life Scale. *Journal of Personality Assessment*, *49*(1), 71-75. https://doi.org/10.1207/s15327752jpa4901_13

Diener, E., Ng, W., Harter, J., & Arora, R. (2010). Wealth and happiness across the world: Material prosperity predicts life evaluation, whereas psychosocial prosperity predicts positive feeling. *Journal of Personality and Social Psychology*, *99*(1), 52-61. https://doi.org/10.1037/a0018066

Diener, E., Suh, E. M., Smith, H., & Shao, L. (1995). National differences in reported subjective well-being: Why do they occur? *Social Indicators Research*, *34*(1), 7-32. https://doi.org/10.1007/BF01078966

Doğan, T., & Totan, T. (2013). Psychometric properties of Turkish version of the Subjective Happiness Scale. *Journal of Happiness and Well-Being*, *1*(1), 21-28. https://toad.halileksi.net/sites/default/files/pdf/the-subjective-happiness-scale-toad_0.pdf

Extremera, N., & Fernández-Berrocal, P. (2014). The Subjective Happiness Scale: Translation and preliminary psychometric evaluation of a Spanish version. *Social Indicators Research*, *119*(1), 473-481. https://doi.org/10.1007/s11205-013-0497-2

Heine, S. J., Lehman, D. R., Peng, K., & Greenholtz, J. (2002). What's wrong with cross-cultural comparisons of subjective Likert scales? The reference-group effect. *Journal of Personality and Social Psychology*, *82*(6), 903-918. https://doi.org/10.1037/0022-3514.82.6.903

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*(2), 65-70. https://www.ime.usp.br/~abe/lista/pdf4R8xPVzCnX.pdf

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1-55. https://doi.org/10.1080/10705519909540118

Iani, L., Lauriola, M., Layous, K., & Sirigatti, S. (2014). Happiness in Italy: Translation, factorial structure and norming of the Subjective Happiness Scale in a large community sample. *Social Indicators Research*, *118*(3), 953-967. https://doi.org/10.1007/s11205-013-0468-7

Jang, S., Kim, E. S., Cao, C., Allen, T. D., Cooper, C. L., Lapierre, L. M., O'Driscoll, M. P., Sanchez, J. I., Spector, P. E.,

Poelmans, S. A. Y., Abarca, N., Alexandrova, M., Antoniou, A.-S., Beham, B., Brough, P., Carikci, I., Ferreiro, P., Fraile, G., Geurts, S., . . .Woo, J.-M. (2017). Measurement invariance of the Satisfaction with Life Scale across 26 countries. *Journal of Cross-Cultural Psychology*, *48*(4), 560-576. https://doi.org/10.1177/0022022117697844

Jovanović, V. (2013). Psychometric evaluation of a Serbian version of the Subjective Happiness Scale. *Social Indicators Research*, *119*(2), 1095-1104. https://doi.org/10.1007/s11205-013-0522-5

Karakasidou, E., Pezirkianidis, C., Stalikas, A., & Galanakis, M. (2016). Standardization of the Subjective Happiness Scale (SHS) in a Greek Sample. *Psychology*, *7*(14), 1753-1765. https://doi.org/10.4236/psych.2016.714164

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research*, *44*(3), 486-507. https://doi.org/10.1177/0049124114543236

Kitayama, S., & Markus, H. R. (2000). The pursuit of happiness and the realization of sympathy: Cultural patterns of self, social relations, and well-being. In E. Diener & E. M. Suh (Eds.), *Culture and subjective well-being* (pp. 113-161). MIT Press.

Korkmaz, S., Goksuluk, D., & Zararsiz, G. (2014). MVN: An R package for assessing multivariate normality. *R Journal*, *6*(2), 151-162. https://journal.r-project.org/archive/2014/RJ-2014-031/RJ-2014-031.pdf

Kotsou, I., & Leys, C. (2017). Echelle de bonheur subjectif (SHS): Propriétés psychométriques de la version française de l'échelle (SHS-F) et ses relations avec le bien-être psychologique, l'affect et la dépression [Subjective Happiness Scale (SHS): Psychometric properties of the French version of the scale (SHS-F) and its relationship to psychological wellbeing, affect and depression]. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, *49*(1), 1-6. https://doi.org/10.1037/cbs0000060

Lansford, J. E. (2018). A lifespan perspective on subjective well-being. In E. Diener, S. Oishi, & L. Tay (Eds.), Handbook of well-being. DEF Publishers.

Lindwall, M., Barkoukis, V., Grano, C., Lucidi, F., Raudsepp, L., Liukkonen, J., & Thøgersen-Ntoumani, C. (2012). Method effects: The problem with negatively versus positively keyed items. *Journal of Personality Assessment*, *94*(2), 196-204. https://doi.org/10.1080/00223891.2011.645936

Lyubomirsky, S., & Lepper, H. S. (1999). A measure of subjective happiness: Preliminary reliability and construct validation. *Social Indicators Research*, *46*(2), 137-155. https://doi.org/10.1023/A:1006824100041

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*(2), 130-149. https://doi.org/10.1037/1082-989X.1.2.130

Marsh, H. W., Hau, K.-T., Balla, J. R., & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research*, *33*(2), 181-220. https://doi.org/10.1207/s15327906mbr3302_1

McCrae, R. R., Zonderman, A. B., Costa, P. T., Jr., Bond, M. H., & Paunonen, S. V. (1996). Evaluating replicability of factors in the Revised NEO Personality Inventory: Confirmatory factor analysis versus Procrustes rotation. *Journal of Personality and Social Psychology*, *70*(3), 552-566. https://doi.org/10.1037/0022-3514.70.3.552

McIntosh, C. N. (2001). Report on the construct validity of the temporal Satisfaction with Life Scale. *Social Indicators Research*, *54*(1), 37-56. https://doi.org/10.1023/A:1007264829700

McMahon, E., & Estes, D. (2012). Age-related differences in lay conceptions of well-being and experienced wellbeing. *Journal of Happiness Studies*, *13*(1), 79-101. https://doi.org/10.1007/s10902-011-9251-0

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*(4), 525-543. https://doi.org/10.1007/BF02294825

Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of Psychological Research*, *3*(1), 112-131. https://doi.org/10.21500/20112084.857

Moghnie, L., & Kazarian, S. S. (2012). Subjective happiness of Lebanese college youth in Lebanon: Factorial structure and invariance of the Arabic subjective happiness scale. *Social Indicators Research*, *109*(2), 203-210. https://doi.org/10.1007/s11205-011-9895-5

Mogilner, C., Kamvar, S. D., & Aaker, J. (2011). The shifting meaning of happiness. *Social Psychological and Personality Science*, *2*(4), 395-402. https://doi.org/10.1177/1948550610393987

Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods*, *17*(3), 313-335. https://doi.org/10.1037/a0026802

Muthén, L. K., & Muthén, B. O. (1998-2018). *Mplus user's guide* (8th ed.). Muthén & Muthén.

Nan, H., Ni, M. Y., Lee, P. H., Tam, W. W. S., Lam, T. H., Leung, G. M., & McDowell, I. (2014). Psychometric evaluation of the Chinese version of the Subjective Happiness Scale: Evidence from the Hong Kong FAMILY Cohort. *International Journal of Behavioral Medicine*, *21*(4), 646-652. https://doi.org/10.1007/s12529-014-9389-3

Oishi, S. (2018). Culture and subjective well-being: Conceptual and measurement issues. In E. Diener, S. Oishi, & L. Tay (Eds.), *Handbook of well-being*. DEF Publishers.

Pavot, W. G., Diener, E., Colvin, C. R., & Sandvik, E. (1991). Further validation of the Satisfaction with Life Scale: Evidence for the cross-method convergence of well-being measures. *Journal of Personality Assessment*, *57*(1), 149-161. https://doi.org/10.1207/s15327752jpa5701_17

Pavot, W. G., Diener, E., & Suh, E. (1998). The Temporal Satisfaction with Life Scale. *Journal of Personality Assessment*, *70*(2), 340-354. https://doi.org/10.1207/s15327752jpa7002_11

Quezada, L., Landero, R., & González, M. T. (2016). A validity and reliability study of the Subjective Happiness Scale in Mexico. *Journal of Happiness and Well-Being*, *4*(1), 90-100. https://www.journalofhappiness.net/frontend/articles/pdf/v04i01/8.pdf

Rojas, M. (2012). Happiness in Mexico: The importance of human relations. In H. Selin & G. Davey (Eds.), *Happiness across cultures: Views of happiness and quality of life in non-Western cultures* (*Vol. 6*, pp. 241-251). Springer Science + Business Media. https://doi.org/10.1007/978-94-007-2700-7_17

Shimai, S., Otake, K., Utsuki, N., Ikemi, A., & Lyubomirsky, S. (2004). Development of a Japanese version of the Subjective Happiness Scale (SHS), and examination of its validity and reliability. *Nippon Koshu Eisei Zasshi*, *51*(10), 845-853.

Sigelman, C. K., & Rider, E. A. (2018). *Life-span human development*. Cengage Learning.

Simons, M., Peeters, S., Janssens, M., Lataster, J., & Jacobs, N. (2018). Does age make a difference? Age as moderator in the association between time perspective and happiness. *Journal of Happiness Studies*, *19*(1), 57-67. https://doi.org/10.1007/s10902-016-9806-1

Spagnoli, P., Caetano, A., & Silva, A. (2012). Psychometric properties of a Portuguese version of the Subjective Happiness Scale. *Social Indicators Research*, *105*(1), 137-143. https://doi.org/10.1007/s11205-010-9769-2

Steenkamp, J.-B., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, *25*(1), 78-90. https://doi.org/10.1086/209528

Stone, A. A., Schwartz, J. E., Broderick, J. E., & Deaton, A. (2010). A snapshot of the age distribution of psychological well-being in the United States. *Proceedings of the National Academy of Sciences*, *107*(22), 9985-9990. https://doi.org/10.1073/pnas.1003744107

Suh, E. M., & Choi, S. (2018). Predictors of subjective well-being across cultures. In E. Diener, S. Oishi, & L. Tay (Eds.), *Handbook of well-being*. DEF Publishers.

Swami, V. (2008). Translation and validation of the Malay Subjective Happiness Scale. *Social Indicators Research*, *88*(2), 347-353. https://doi.org/10.1007/s11205-007-9195-2

Swami, V., Stieger, S., Voracek, M., Dressler, S., Eisma, L., & Furnham, A. (2009). Psychometric evaluation of the Tagalog and German Subjective Happiness Scales and a cross-cultural comparison. *Social Indicators Research*, *93*(2), 393-406. https://doi.org/10.1007/s11205-008-9331-7

Szabo, A. (2019). Validity of the Hungarian version of the Subjective Happiness Scale (SHS-HU). *Mentálhigiéné és Pszichoszomatika*, *20*(2), 180-201. https://doi.org/10.1556/0406.20.2019.010

Veenhoven, R. (1994). Is happiness a trait? *Social Indicators Research*, *32*(2), 101-160. https://doi.org/10.1007/BF01078732

Vera-Villarroel, P., Celis-Atenas, K., & Córdova-Rubio, N. (2011). Evaluation of happiness: Psychometric analysis of the Subjective Happiness Scale in Chilean population. *Terapia Psicologica*, *29*(1), 127-133. https://doi.org/10.4067/S0718-48082011000100013

Yoon, M., & Lai, M. H. (2018). Testing factorial invariance with unbalanced samples. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(2), 201-213. https://doi.org/10.1080/10705511.2017.1387859