

The Measurement of Variability and Change in Life Satisfaction:
A Comparison of Single-Item and Multi-Item Instruments

Timo Gnams

Leibniz Institute for Educational Trajectories

Katja Buntins

University of Bamberg

Author Note

Timo Gnams, Leibniz Institute for Educational Trajectories, Germany; Katja Buntins, Psychology Department, University of Bamberg, Germany;

Correspondence concerning this article should be addressed to Timo Gnams, Leibniz Institute for Educational Trajectories, Wilhelmsplatz 3, 96047 Bamberg, Germany, Phone: +49 (0)951 863-3420, Email: timo.gnams@lifbi.de.

Summary

In many large-scale surveys life satisfaction is measured with a single item. Little is known about how sensitive single-item instruments react to changes across time or in the face of critical life events. The present study draws on a Dutch sample of $N = 4,034$ participants who were administered three different single items and an established multi-item instrument at four annual surveys within six years. An indicator-specific latent state-trait growth model was employed to estimate the agreement in change coefficients for the different instruments. The results indicated that single items estimated a similar developmental decline in life satisfaction across the life course as multi-item scales, but they were less sensitive to the loss of paid work. Interindividual differences in intraindividual change were better captured with multiple items. Moreover, single items reflected occasion-specific variance to a larger degree than longer scales. In addition, occasion-specific effects were larger among younger respondents as compared to older age groups. Therefore, single items appear to represent valid measures for the estimation of the average change in life satisfaction within a sample. However, they might exhibit a low power to detect moderating influences on the estimated change trajectories and capture less trait variance than multi-item scales.

Keywords: life satisfaction, latent state-trait theory, longitudinal, single item, validity

Introduction

Life satisfaction refers to an individual's global assessment of his or her quality of life. It represents the outcome of a cognitive, judgmental process in the form of a comparison of one's current life circumstances (e.g., with regard to social relationships, physical health, or working conditions) to an implicit standard that is deemed appropriate (Diener, Emmons, Larsen, & Griffin, 1985). In large-scale social and economic surveys life satisfaction is typically measured with a single item that asks respondents to evaluate their lives as a whole. Because single-item measurements are sometimes of a rather poor psychometric quality (e.g., Credé, Harms, Niehorster, & Gaye-Valentine, 2012; Diamantopoulos, Sarstedt, Fuchs, Wilczynski, & Kaiser, 2012), the present study compares two popular single-item measures to an established multi-item instrument for the measurement of continuous and episodic changes in life satisfaction. Moreover, latent state-trait (LST) analyses (Steyer, Mayer, Geiser, & Cole, 2015) were used to evaluate to what extent the different measures reflect stable trait-like characteristics or trait changes rather than temporary state-like fluctuations.

The State-Trait Debate on Life Satisfaction

The nature and dynamics of life satisfaction have been subject to some debate. On the one hand, changes in people's life circumstances (e.g., the loss of paid work or a beloved person) can have a pronounced impact on their satisfaction with life. For example, life satisfaction significantly increases after marriage (e.g., Diener & Diener McGavran, 2008), whereas even short unemployment spells have a detrimental effect on life satisfaction (e.g., Gnambs, Selenko, & Stiglbauer, 2015). This has been corroborated in longitudinal as well as meta-analytic studies (see Yap, Anusic, & Lucas, 2014, for a review). Thus, life satisfaction is strongly affected by temporary situational influences. On the other hand, life satisfaction has also a rather stable core: After some time most people adapt to the changes in their lives and, in the long run, their satisfaction gradually returns near its initial level—although it may never get fully restored (Lucas, Clark, Georgellis, & Diener, 2004). Moreover, life satisfaction also

exhibits systematic longitudinal associations with enduring personality traits (e.g., Soto, 2015). Together, these findings indicate that life satisfaction is a relatively stable construct that can be temporarily altered by external events and changes in one's life.

In order to determine long-term developmental trajectories or to study the influence of critical life events on life satisfaction, it is important to separate stable trait components from temporary situational influences. An adequate framework is provided by LST theory (Steyer et al., 2015) that partitions the observed variance of a measure into three components: (a) trait variance describing the stable component of life satisfaction, (b) state residual variance reflecting systematic influences of situational factors that are unique to a given measurement occasion, and (c) random measurement error. Various extensions of LST theory (e.g., Bishop, Geiser, & Cole, 2015; Eid & Hoffmann, 1998; Geiser, Keller, & Lockhart, 2013) also permit scrutinizing the state residual variance in more detail and distinguishing systematic short-term fluctuations (i.e., a form of transient measurement error) from long-term variations (i.e., developmental trait changes). Applications of LST theory to the construct of life satisfaction showed that trait-specific effects contribute somewhere between 30 to 80 percent of the observed variance in respective measures (e.g., Eid & Diener, 2004; Kaczmarek, Bujacz, & Eid, 2015; Lucas & Donnellan, 2007, 2012; Schimmack, Krause, Wagner, & Schupp, 2010). A recent meta-analysis (Anusic & Schimmack, 2016) concluded that about 52 percent of the reliable variance in life satisfaction is attributable to a stable trait, whereas a similar percentage (48 percent) reflected state residual variance. Moreover, studies that distinguished short-term situational influences from long-term changes across the life course demonstrated that the former accounted for about 10 to 20 percent of the observed variance in measures of life satisfaction whereas the latter accounted for about 30 percent (e.g., Lucas & Donnellan, 2007, 2012; Schimmack et al., 2010). Attempts aimed at describing the form of these long-term changes in more detail have not yet reached a definitive conclusion (see Ulloa, Møller, & Sousa-Poza, 2013, for a review). Whereas some authors identified a U-shaped change

through life with lowest levels of life satisfaction in midlife (Blanchflower & Oswald, 2008; Cheng, Powdthavee, & Oswald, 2015), others reported the exact opposite and found an inverted U-shaped trajectory (Easterlin, 2006). Yet others found a small linear decline in life satisfaction over time (Deaton, 2008; Frijters, & Beaton, 2012; Kassenboehmer & Haisken-DeNew, 2012). Thus, although there is no consensus yet as to the precise change pattern of life satisfaction across the life course, the available results indicate that life satisfaction exhibits enduring trait-like characteristics but also temporary state-like variations.

The Measurement of Life Satisfaction: Single versus Multiple Item Sets

Many large-scale social and economic studies such as the German Socio-Economic Panel or the European Social Survey routinely assess various aspects of subjective well-being. Life satisfaction is typically measured with a single item inquiring about the current or past satisfaction with people's lives (see Waldron, 2010). The use of single-item measurements in these surveys is rarely motivated by psychometric considerations, but is mostly based on economical and pragmatical reasons. Because the administration of each item is costly, single items are frequently used in order to collect a variety of different information without overly inflating the length of the questionnaire. Moreover, long surveys increase the likelihood that people answer without sufficient cognitive processing; they base their responses on simple heuristics rather than internal reflection and evaluation, that is, they tend to satisfice (e.g., Holbrook, Green, & Krosnick, 2003). Thus, single items reduce feelings of boredom or fatigue in respondents and consequently decrease the likelihood of non-response and careless responding.

Nonetheless, instruments including multiple items are supposed to hold several advantages. For example, they allow for the assessment of multiple facets of a construct and thus can make the construct more explicit and concrete. However, the use of many items that measure essentially the same thing may also confuse participants; consequently, they are lead to give consistent responses, or even to not take the questions seriously (Bergkvist & Rossiter,

2007). Another argument for the use of multiple item sets is that aggregated scores across multiple items typically lead to a larger between-person variance and thus make it easier to discriminate between individuals. However, this limitation of single items could potentially be remedied by using a sufficient number of response categories. The use of multiple items is further encouraged by the wide acceptance of classical test theory and its focus on measurement error (see Gnamb, 2015a, for a recent review). Multiple item scales enable more accurate assessments of hypothetical constructs because they tend to have higher reliabilities than single-item measures. Finally, some authors (e.g., Credé et al., 2012; Diamantopoulos et al., 2012) suggested that single-item measures show impaired construct validities as compared to multi-item instruments. However, this has not been substantiated by others (e.g., Bergkvist, 2015; Bergkvist & Rossiter, 2007; Gogol et al., 2014; Thalmeyer, Saucier, & Eigenhuis, 2011). In conclusion, the popularity of multi-item scales can be attributed to the belief that multiple items result in a higher reliability, validity and other psychometrically desirable features (for recent discussions on the benefits and pitfalls of short scales see Fisher, Matthews, & Gibbons, 2016, Rammstedt & Beierlein, 2015, or Ziegler, Kemper, & Kruyen, 2015).

With regard to the assessment of life satisfaction, a series of studies showed that single-item indicators behave in much the same way as multiple item scales. Both instruments exhibit validities that are highly comparable (e.g., Abdel-Khalek, 2006; Cheung & Lucas, 2014; Pavot & Diener, 1993; Sandvik, Diener, & Seidlitz, 1993). Thus, there is some evidence that single-item measures are not more biased than comparable multi-item instruments. Moreover, short-term test-retest correlations for single-item measures seldom fall below .60 (Kruger & Schkade, 2008; Schimmack et al., 2010) and, thus, exhibit pronounced stability. However, meta-analytic evidence indicates that stability correlations are significantly affected by the length of the measurement instrument with single items resulting in lower estimates (Anusic & Schimmack, 2016; Schimmack & Oishi, 2005). Stability

correlations derived from single items are about .20 lower than respective correlations using multi-item scales (Schimmack & Oishi, 2005). Given the higher reliabilities of multi-item scales, this might simply be a consequence of larger random error in single items.

Alternatively, single items might also be more strongly affected by situational influences or less sensitive to true trait changes. Global assessments of life satisfaction require respondents to consider different facets of their current lives as a whole (e.g., including their health, working conditions, and family life) and form an overall judgement reflecting a composite of these facets. However, these judgments might be disproportionately influenced by a single event that is easy to recall, for example, because of its temporal proximity or its emotional significance (e.g., a recent quarrel with one's fiancée). As a result, other important life domains might be insufficiently taken into account when responding to a single item. In contrast, multi-item scales might cover the breadth of life satisfaction more appropriately because respondents can differentially weight the various facets of their lives. So far, little is known about how sensitive different scale types react to changes across time or in the face of critical life events. Therefore, it is important to determine whether single items or multiple item sets are more strongly affected by situational influences or, rather, both capture the trait and trait change components of life satisfaction equally well.

Overview of Research

The present study examined change processes in life satisfaction estimated from both single items and an established multi-item scale. In contrast to previous studies that primarily administered single items (e.g., Lucas & Donnellan, 2007, 2012; Schimmack et al., 2010), we adopted a multi-method strategy to evaluate the convergent validity of change processes as captured by different instruments. Given the higher reliability of multi-item scales (cf. Anusic & Schimmack, 2016; Schimmack & Oishi, 2005), single items might be more strongly affected by measurement error and, thus, underestimate true trait changes as compared to longer scales. We studied continuous change trajectories over a period of six years to

determine the convergent validities of the administered instruments for the measurement of developmental change. Subsequently, these analyses were replicated for the examination of episodic changes by focusing on the time before and after a critical life event (i.e., the transition from paid employment into non-employment). Furthermore, the different measures were evaluated within the LST paradigm to compare to what extent single items and multiple item scales reflect similar variance components (i.e., with regard to the trait and state residual variance in observed scores). Given the contradictory findings on the long-term stability of life satisfaction (cf. Ulloa et al., 2013), these analyses also examined potential age-related effects on these variance estimates. It has recently also been suggested that person characteristics including the age of the respondent might affect the degree of transient measurement error in observed scores (cf. Gnams, 2015b). Therefore, we studied age-related differences in the trait, state residual, and random error variances for the different measures.

Materials and Method

Sample

The participants were drawn from the Longitudinal Internet Studies for the Social Sciences (LISS) panel that covers Dutch-speaking individuals permanently residing in the Netherlands (cf. Blom et al., 2016). The panel is based on a true probability sample of households drawn from the population register. Its members are surveyed once a month using web-based interviews at their private homes. Households that could not otherwise participate are provided with a computer and Internet connection. The present study focuses on 4,034 respondents (54% female) that participated at least twice at four measurement occasions from 2008 to 2014 (at intervals of two years). The initial age of the participants ranged from 18 to 88 years ($M = 48.18$, $SD = 15.45$). About 30 percent had finished higher vocational education (e.g., college or university). Most people were currently married and in paid employment (see Table 1). The analyses of episodic change in life satisfaction due to a critical life event (i.e., non-employment) made use of a subsample of $N = 503$ respondents (51% female) with a

mean age of $M = 49.59$ years ($SD = 12.18$) that experienced a transition from paid work into non-employment (e.g., unemployment or retirement) within the six year period.

Materials and Procedure

We selected three single items that are often used interchangeably in well-being research: Life satisfaction (“How satisfied are you with the life you lead at the moment?”) was rated on an 11-point response scale from 0 “not at all satisfied” to 10 “completely satisfied” and using a version of Cantril’s (1965) ladder of life (“If you imagine a ‘ladder of life’, where the first step represents the worst possible life, and the tenth (top) step the best possible life, on what step would you place yourself?”) on an 11-point scale. Life happiness (“On the whole, how happy would you say you are?”) was measured on a response scale from 0 “totally unhappy” to 10 “totally happy”. Although life happiness is supposed to reflect a more affect-driven facet of well-being (Gallagher, Lopez, & Preacher, 2009), single-item measures of life satisfaction and happiness frequently show rather high correlations exceeding .70 (e.g., Cheung & Lucas, 2014). Finally, life satisfaction was also measured with an established multi-item instrument. The Satisfaction with Life Scale (SWLS; Diener et al., 1985) includes five items (see Appendix) to be rated on 7-point response scales from 1 “strongly disagree” to 7 “strongly agree”. The scale resulted in ω_h reliabilities (McDonald, 1999) of about .84 at the four measurement occasions. To be consistent with the response scales of the three single items, the scores of the SWLS were rescaled onto a range from 0 to 10.

The four measures were administered as part of two separate surveys in May and June of each year. The life happiness item was presented first followed by the life satisfaction item and the SWLS in the May survey, whereas the ladder of life item was included in the June survey. In both surveys, these instruments were administered at the beginning of the questionnaires after updating various socio-demographic information. The order of presentation was identical at all four measurement occasions. The May survey focused on

personality measurement in general and instructed respondents to give the first answer that came to mind without giving too much thought to the questions. In contrast, the survey administered in June pertained to the economic situation of the respondents. However, at the beginning of the questionnaire they were informed that the first questions would not concern their income but their own lives in a broad sense.

Attrition Analyses

About 78 percent of the sample participated at all four measurement occasions, whereas about 10 percent each either participated two or three times. To rule out a systematic bias due to nonresponse, we examined whether the frequency of participation (two to four times) was associated with the focal variables. We found that initial levels on the life satisfaction item, $F(2, 3,906) = 0.42, p = .65, \eta^2 = .00$, happiness item, $F(2, 3,893) = 0.09, p = .91, \eta^2 = .00$, ladder of life item, $F(2, 3,740) = 1.39, p = .25, \eta^2 = .00$, and satisfaction with life scale, $F(2, 3,926) = 0.06, p = .94, \eta^2 = .00$, were unrelated to the response frequency. Moreover, frequency of participation showed only negligible effects on various socio-demographic characteristics (see Table 1). Although younger respondents, unmarried singles, and students exhibited significantly ($p < .05$) higher unit non-response rates, the respective effect sizes were small and explained less than 1% of variance. Hence, it is unlikely that selective dropout introduced a systematic bias.

Statistical Analyses

Longitudinal changes in life satisfaction were examined using latent growth modeling (cf. McArdle & Grimm, 2010). Latent growth models estimate a growth trajectory, that is, an increase or a decrease of satisfaction over time, by modeling two latent factors: the latent intercept factor ξ_{int} represents individual differences in life satisfaction trait scores at the first measurement occasion whereas the latent slope factor ξ_{slp} reflects the trait change process across measurement occasions (see Figure 1). The latent growth models for the four measures were fitted simultaneously to the data. As a consequence, the correlations between the four

latent slope factors ξ_{slp} reflect the similarity of change trajectories for the different measures: Correlations approaching 1 would indicate convergent change processes for two measures. To separate true trait changes from situational influences that are unique to a given measurement occasion and random measurement error the latent growth models were extended to indicator-specific latent state-trait growth models (Bishop et al., 2015) that specified different intercept and slope factors for each measured variable (see Figure 1). Thus, we included a common latent state residual factor ζ_t for all four measures at each measurement occasion. The four state residual factors were neither correlated to each other nor to the latent growth factors. Formally, the indicator-specific latent state-trait growth model for the observed score Y_{it} of measure i at measurement occasion t is given as

$$Y_{it} = \tau_{it} + \varepsilon_{it} = \xi_{int_i} + \lambda_t \xi_{slp_i} + \gamma_i \zeta_t + \varepsilon_{it} . \quad (1)$$

The two latent growth factors ξ_{int_i} and ξ_{slp_i} reflect the stable trait and trait change components for a measure i , whereas the state residual factor ζ_t describes systematic situation effects unique to a specific measurement occasion t . Together, the three latent factors form the true score τ_{it} for measure i at time t . ε_{it} indicates the respective random measurement error. For each loading γ_i on the latent state residual factors measurement invariance was introduced by constraining the loadings to be equal across time. The loadings λ_t on each latent slope factor were fixed to *a priori* determined values that were a function of the respective measurement occasion and reflected the hypothesized change trajectory. For the continuous change model we modeled linear changes¹ across time and used $\lambda_t = t - 1$ resulting in factor loadings of 0, 1, 2, and 3 at the four measurement occasions (see Figure 1). In the episodic change model we

¹ Initially, we also modeled quadratic change processes. However, the respective latent factor means and all but one latent factor variance were not significant ($p > .05$). Moreover, model comparisons using the Bayesian Information Criterion (BIC; Schwarz, 1978) indicated a better fit for the linear model (BIC = 145,846) as compared to the quadratic model (BIC = 146,085). Therefore, we excluded the quadratic slope factor from our analyses.

expected a stronger decline in well-being after the critical life event (i.e., the transition into non-employment). Therefore, we modeled a non-linear change trajectory using $\lambda_t = (t - 1)^{1.6}$ that resulted in factor loadings of 0, 1, and 3 at the three measurement occasions. The transformation function for λ_t was empirically derived by comparing the information criteria for competing models that used different λ_t .

The latent state-trait methodology allows the dissection of the observed variance in each measure into several variance components: The consistency reflects the degree of stable person-specific effects and their changes across time (i.e., trait and trait change variance); the occasion-specificity indicates to what degree observed scores are influenced by temporary situational effects and person-situation interactions (i.e., state residual variance); finally, random measurement error reflects unsystematic influences on each score. Following Bishop and colleagues (2015) the observed variance $Var(Y_{it})$ in an indicator-specific latent state-trait growth model is given as

$$\begin{aligned} Var(Y_{it}) &= Var(\tau_{it}) + Var(\varepsilon_{it}) \\ &= Var(\xi_{int_t}) + \lambda_t^2 Var(\xi_{slp_t}) + 2\lambda_t Cov(\xi_{int_t}, \xi_{slp_t}) + \gamma_i^2 Var(\zeta_t) + Var(\varepsilon_{it}) \end{aligned} \quad (2)$$

Then, we calculated the consistency $Con(Y_{it})$ and occasion-specificity $OSpec(Y_{it})$ of the observed scores² as

$$Con(Y_{it}) = \frac{Var(\xi_{int_t}) + \lambda_t^2 Var(\xi_{slp_t}) + 2\lambda_t Cov(\xi_{int_t}, \xi_{slp_t})}{Var(Y_{it})} \quad \text{and} \quad (3)$$

$$OSpec(Y_{it}) = \frac{\gamma_i^2 Var(\zeta_t)}{Var(Y_{it})} \quad (4)$$

² In contrast to Bishop et al. (2015) who proposed the consistency and occasion-specificity of the true scores and used $Var(\tau_{it})$ in the denominators of (3) and (4), we calculated the consistency and occasion-specificity of the observed scores Y_{it} . This facilitates the comparison of different variance components (including measurement error) in the administered well-being measures (see Figure 3).

Consequently, the proportion of measurement error in the observed scores results as $1 - Con(Y_{it}) - OSpec(Y_{it})$. Thus, this variance partitioning enables comparisons of true score components in the four administered measures³.

The indicator-specific latent state-trait growth model was estimated using a full information maximum likelihood algorithm that has been shown to yield unbiased parameter estimates in covariance analyses when responses are missing at random (Enders & Bandalos, 2001; Newman, 2003). To account for departures from multivariate normality in our data, we adopted the Yuan-Bentler test statistic (Yuan & Bentler, 2000) and estimated heteroskedasticity-robust standard errors (cf. Hays & Cai, 2007). Model fit was evaluated in line with conventional standards (see Schweizer, 2010) using the comparative fit index (CFI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR). Model comparisons were based on the Bayesian Information Criterion (BIC; Schwarz, 1978) that indicates a better fit for models with a smaller BIC.

Software and Data Availability

The indicator-specific latent state-trait growth models were estimated with the *lavaan* software version 0.5-20 (Rosseel, 2012) in *R* version 3.3.0. The raw data can be retrieved from <http://lissdata.nl>.

Results

Analyses of continuous change

The average stability correlations across two years were $\bar{r}_t = .59$, $.61$, and $.56$ for the life satisfaction, happiness, and ladder of life items (see Table 2). In contrast, the respective stability correlations for the multi-item satisfaction with life scale amounted to $\bar{r}_t = .69$. Thus,

³ At each measurement occasion the four measures were administered as part of two separate surveys in May and June. Although the time lag of one month is negligible for the examination of trait changes (see Gnambs, 2014), the degree of occasion-specific variance will likely be underestimated for the ladder of life item.

the multi-item scale exhibited stronger rank-order stability than comparable single items.

Moreover, all scales showed mean-level changes across the four measurement occasions that resulted in a gradual decline in well-being over time (see left panel of Figure 2).

Parameter estimates. The indicator-specific latent state-trait growth model resulted in a good fit to the data, $\chi^2(85) = 147.05$, CFI = .998, RMSEA = .013, RMSEA 90% CI = [.010, .016], SRMR = .028, BIC = 145,782. Relaxing the invariance constraints for the loadings on the latent state residual factors did not improve the model fit significantly, $\chi^2(76) = 143.74$, CFI = .997, RMSEA = .015, RMSEA 90% CI = [.012, .018], SRMR = .027, BIC = 145,846, $\Delta\chi^2(9) = 5.92$, $p = .75$. Thus, the measurement invariance assumption was supported. However, the model contained an improper parameter estimate (i.e., “Heywood case”) in the form of a correlation greater than 1. Because the respective correlation was not significantly ($z = 0.28$, $p = .61$) different from 1, it can likely be explained by sampling error (McDonald, 2004).

The parameter estimates of the indicator-specific latent state-trait growth model are summarized in Table 3. The latent slope factors for the four measures showed negative means that were significantly different from zero ($p < .05$) indicating that, on average, well-being declined over the four measurement occasions. Placing equality constraints on the latent means of the four slope factors resulted in a significant decline in fit, $\Delta\chi^2(3) = 8.25$, $p = .04$. Whereas the three measures of life satisfaction had comparable means ($\Delta\chi^2(2) = 2.98$, $p = .22$), the life happiness item exhibited a smaller decline in well-being over time. This indicates that each life satisfaction measure reflected a similar average decline in well-being, whereas the life happiness item exhibited a different change process. The variances of the latent slope factors that were significantly different from zero ($p < .05$) revealed substantial interindividual differences in the change processes across individuals for each measure. Moreover, the four measures resulted in significantly different slope factor variances ($\Delta\chi^2(3) = 49.32$, $p < .001$). The variance of the multi-item scale was nearly twice as large as the

variance of the life satisfaction and life happiness items and about four times as large as the respective variance of the ladder of life item. Thus, the multi-item scale reflected differences in change processes between respondents more strongly than comparable single items.

To examine to what degree the estimated change trajectories converged across measures we inspected the correlations between the four latent slope factors (see Table 4). The slope factor correlation for the life satisfaction and happiness items was not significantly different from 1 (i.e., the 95% confidence interval ranged from .96 to 1.00) and thus estimated similar change processes. Moreover, the confidence intervals for the slope factor correlations between the multi-item satisfaction with life scale and the two single items on life satisfaction included 1 (see Table 4) and thus reflected comparable rank-order differences in the estimated change trajectories. Only life happiness showed a smaller slope factor correlation with the multi-item scale, $r = .77$, 95% CI [.74, .91].

Variance components. The adopted latent state-trait methodology allowed a decomposition of the observed variance of each measure into trait and trait change variance (i.e., consistency), occasion-specific variance, and random measurement error. The average proportions of these three variance components in the observed scores of the four measures are summarized in Figure 3 (left panel). The single-item scores captured considerably more occasion-specific variance than the multi-item scale⁴: About a quarter of the observed variance in life satisfaction and happiness items was due to situational influences, whereas the respective proportion fell at about 7 percent for the multi-item scale. Moreover, the trait score component in observed scores was about 10 percent higher in the satisfaction with life scale as compared to the three single-item measures.

⁴ The occasion-specificity of the ladder of life item is likely to be underestimated because it was administered in a separate survey.

To determine the influence of the respondents' age on the variance decomposition, we reestimated the consistency and occasion-specificity coefficients in six age cohorts: from 18 to 29 years ($N = 559$), 30 to 39 years ($N = 679$), 40 to 49 years ($N = 832$), 50 to 59 years ($N = 908$), 60 to 69 years ($N = 745$), and 70 to 88 years ($N = 311$). The respective multi-group model included partial measurement invariance constraints on the latent state residual factors across groups. In the two highest age cohorts the loadings of the multi-item scale were significantly smaller than in the other cohorts. Therefore, the respective loadings were constrained across age cohorts 1 to 4 and across age cohorts 5 and 6. The means and variances of all latent factors were allowed to vary across the age groups. This model did not exhibit a loss in fit as compared to an unconstrained model, $\chi^2(524) = 629.12$, CFI = .996, RMSEA = .017, RMSEA 90% CI = [.012, .022], SRMR = .039, BIC = 147,215, $\Delta\chi^2(14) = 21.66$, $p = .09$. The means of the latent slope factors for the life satisfaction item ($M(\xi_{slp}) = -0.06$) the ladder of life item ($M(\xi_{slp}) = -0.05$) and the multi-item scale ($M(\xi_{slp}) = -0.05$) showed remarkable consistency across the age groups. The BIC favored a model that constrained the latent slopes for each measure to be equal across the age groups as compared to an unconstrained model. In contrast, the decline in life happiness was only comparable across the three younger age groups ($M(\xi_{slp}) = -0.09$); life happiness did not change among the three older age groups ($M(\xi_{slp}) = 0.00$). Then, the variance components were estimated in each age cohort (see right panel of Figure 3). A regression of the thus derived variance components on the respective age cohorts showed that the average proportions of trait and trait change variance in the observed scores significantly ($p < .05$) increased with the age cohort, whereas the proportions of occasion-specific variance significantly decreased (see Table 5). For example, for the life satisfaction item the consistency increased from about 48% among respondents in their twenties to about 61% among respondents in their seventies; at the same time the occasion-specificity decreased from about 37% to about 19% (see right panel of

Figure 3). Similar differences in these variance components were observed for the multi-item scale: The consistency increased by about 12 percentage points and the occasion-specificity decreased by about 9 percentage points across the six-year period. Despite these differences in consistency and occasion-specificity for the six age cohorts, the overall reliabilities of the administered measures were unaffected by age ($p > .05$).

Analyses of Episodic Change

Episodic changes in life satisfaction due to a critical life event (i.e., non-employment) were examined by rearranging the responses of a subsample of 503 respondents that experienced a change in employment status into three measurement units: at the first and second measurement before the critical event all respondents were in paid employment, whereas at the third measurement after the event they were out of the labor force (e.g., unemployed, pensioners, homemakers). Again, the average stability correlations across two years (see Table 6) were larger for the multi-item satisfaction with life scale ($\bar{r}_{tt} = .69$) than for the single items ($\bar{r}_{tt} = .59$ to $.63$). Moreover, the multi-item scale exhibited a stronger decline in well-being due to non-employment than the three single item measures (see right panel of Figure 2).

Parameter estimates. The indicator-specific latent state-trait growth model, $\chi^2(28) = 44.43$, CFI = .994, RMSEA = .034, RMSEA 90% CI = [.014, .051], SRMR = .037, BIC = 13,617, revealed negative means for all four latent slope factors (see Table 3) that were significantly different from zero ($p < .05$). In addition, the four latent slope factor means were significantly different from each other ($\Delta\chi^2(3) = 17.25$, $p < .001$), that is, the four measures exhibited a different average decline in well-being due to non-employment. The estimated decline in life satisfaction was between 1.5 to 2.3 times larger for the multi-item scale as compared to the two single life satisfaction items. Moreover, the multi-item satisfaction with life scale was the only instrument resulting in a latent slope factor variance that was significantly different from zero ($p < .05$); for the three single items the respective variances

were not significant ($p > .05$). The latent slope factor variance for the multi-item scale was over three times as large as the respective variance of the single life satisfaction item. Thus, it would be difficult to identify moderating influences on the change trajectories with the single-item measures as compared to the multi-item scale, which reflected interindividual differences in change more strongly. However, constraining the variances of the latent slope factors to be equal across the four measures revealed only a marginally significant loss of fit, $\Delta\chi^2(3) = 7.29, p = .06$.

Variance components. The variance decomposition in the indicator-specific latent state-trait growth model (see Figure 4) showed that the single-item scores captured pronouncedly more occasion-specific variance than the multi-item scale⁴. For the life satisfaction and happiness items about 30% and 23% of the observed variances were due to situational influences, whereas the respective proportion fell at about 11 percent for the multi-item scale. In contrast, the trait score component in observed scores was about 10 percent larger in the satisfaction with life scale as compared to the three single-item measures. Taken together, the analyses of episodic change closely mirrored the results for the continuous change models presented in the previous sections.

Discussion

Many large-scale social and economic studies include an increasing number of psychological constructs in their surveys. However, due to economic reasons, these are frequently measured with short scales or even single items (see Rammstedt & Beierlein, 2015). Because of concerns regarding a potentially rather poor psychometric quality of these instruments (e.g., Credé et al., 2012; Diamantopoulos et al., 2012), the present study examined three single-item measures of life satisfaction and life happiness to determine to what extent they are sensitive to developmental and situational changes. The presented findings allow the following conclusions: First, single items identified similar mean-level and rank-order changes in life satisfaction across the life course as compared to an established

multi-item instrument. However, single items were less sensitive in determining the impact of critical life events; longer scales estimated a decline in life satisfaction due to non-employment that was about twice as large as the respective effect estimated from single items. Second, the degree of interindividual differences in intraindividual change tended to be underestimated by single items; multiple item sets estimated about two to three times larger variances of the change parameters. As a consequence, single items are likely to exhibit a limited power to detect moderating influences on the change trajectories. Third, for all administered instruments over 60 percent of the observed score variance reflected a trait component (albeit the respective proportion was slightly larger for the multi-item scale). Moreover, single items also captured a substantial amount of occasion-specific variance. Thus, they are likely to confound trait and situation effects in cross-sectional studies. Fourth, the variance decomposition of the observed well-being scores was age-dependent. In older age groups life satisfaction scores reflected trait and trait change variance to a larger degree than in younger age groups. In contrast, situation-specific influences decreased with the mean age of the respondents. However, overall reliability estimates were unaffected by the age of the respondents. Finally, the life happiness item that operationalizes a more affect-laden facet of subjective well-being (cf. Gallagher et al., 2009) exhibited discriminant change trajectories as compared to the three life satisfaction instruments. Life happiness showed smaller mean-level and rank-order change trajectories over time as well as a result of non-employment. Only the two single items measuring life satisfaction and life happiness exhibited a virtually indistinguishable rank-order change.

Implications for Well-Being Research

Single item measurements are a necessary compromise in many large-scale panel studies. On the one hand, they are an unobtrusive, cost-effective way for the assessment of a large number of people. On the other hand, the presented analyses indicate that they are likely to increase Type 2 errors in longitudinal research. The absolute change in life satisfaction can

be considerably underestimated using single items as compared to multi-item scales. Thus, single item studies with low to moderate power might erroneously conclude that life satisfaction does not change, whereas comparable studies using longer instruments are likely to identify significant change trajectories. This weakness is even more severe for the analyses of interindividual differences in change. Single items are less likely to identify significant moderating effects on the examined change process or to document concurrent change trajectories between two or more constructs (cf. Soto, 2015). Thus, the reliance on single item measures decreases the likelihood of identifying substantial effects in longitudinal research.

Furthermore, in our study single items were subject to stronger situational influences. For the multi-item scale about seven to ten percent of the observed score variance was attributed to the specific measurement occasion. Comparable proportions were reported in previous research on enduring personality traits such as extraversion, the need for affect, or competitive anxiety (Appel, Gnambs, & Maio, 2012; Deinzer et al., 1995; Ziegler, Ehrlenspiel, & Brand, 2009). In contrast, the proportion of occasion-specific variance in single items was about three times as large; about a quarter of the observed variance in life satisfaction was due to situational effects. Given that even in explicit state measures occasion-specific effects account for only about half of the observed variance (Gnambs, Appel, Schreiner, Richter, & Isberner, 2014), the state-trait debate on life satisfaction (cf. Anusic & Schimmack, 2016; Schimmack & Oishi, 2005) might partly confound questions on the nature of life satisfaction with measurement issues. The consequences of these differences for longitudinal research are less severe because change trajectories are unlikely to be affected by differences in occasion-specificity. However, cross-sectional research using single items is more likely to confound trait and state components of life satisfaction.

Finally, the trait and state residual variances in observed life satisfaction scores were subject to pronounced age-specific variations. Responses from older individuals were less affected by transient situational influences than responses from younger people; at the same

time, they reflected the trait component of life satisfaction to a larger degree. In part, this might explain the discrepant findings on the form of long-term change in life satisfaction (cf. Ulloa et al., 2013). Many of these studies adopted cross-sectional designs and analyzed differences between age groups to infer change trajectories across the life course. However, if situational conditions exerted non-uniform effects on observed life satisfaction scores in the different age groups, the reported change analyses might have confounded trait differences with situation effects.

The underlying mechanisms governing these age-related differences in the variance components of life satisfaction scores have not yet been fully explored. For example, they might be a result of changes in person environments that become more stable over the life course (cf. Roberts & DelVecchio, 2000). In youth, individuals are frequently confronted with changing environmental demands, for example, with regard to new romantic relationships, new jobs, or new living conditions. However, as people mature their environments gradually become less variable and more stable (i.e., they live in enduring partnerships and have steady jobs). At the same time, people accumulate experiences. With age, they encounter fewer novel experiences and have less trouble in coping with environmental demands (Tyler & Schuller, 1991). Consequently, environmental changes associated with the age of the respondents could systematically affect the proportion of occasion-specific variance. Alternatively, cognitive changes associated with age might serve as another explanation. Responding to survey items draws on several cognitive resources such as executive functioning or memory (cf. Tourangeau & Bradburn, 2010), all of which are affected by developmental changes (Tucker-Drob & Briley, 2014). As a result, age-related differences in cognitive functioning can affect how survey items are understood and answered. For example, older respondents are less influenced by context information; consequently their answers more strongly reflect the item content (Knäuper, Schwarz, Park, & Fritsch, 2007; Diersch & Walther, 2016). In line with these findings, it could be assumed that the proportion of trait variance increases with the age

of the respondents. Although the actual mechanism underlying the observed age-differences is still in the dark, our findings showed that cross-sectional and longitudinal cohort analyses of life satisfaction typically cannot distinguish true cohort differences from differences in the response process due to cognitive or environmental changes (cf. Schwarz, 2003).

Limitations and Directions for Future Research

Some limitations should be noted that might compromise the generalizability of our findings. First, the single and multi-item instruments were administered to all respondents within the same survey in identical sequence. Thus, the lengthy questionnaire might have counteracted some advantages of single items and resulted, for example, in more uniform, less elaborated answers (Galesic & Bosnjak, 2009) or a higher tendency to satisfice (see Van Vaerenbergh & Thomas, 2013). Moreover, order effects due to different positions of the administered well-being scales within the survey might have distorted our analyses to some degree. Particularly, the assessment of subjective well-being has been shown to be sensitive to the item content presented earlier in a survey (e.g., Schwarz, Strack, & Mai, 1991). Finally, the simultaneous presentation of different scales measuring essentially the same construct might have resulted in participants to respond in a consistent way and, as a result, might have led to an overestimation of the convergent validities for the single and multiple item scales (cf. Knowles & Byers, 1996). Therefore, future research is encouraged to investigate single and multiple item sets using separate questionnaires, for example, in a longitudinal or an experimental between-subjects design. Second, the examined sample was rather heterogeneous in terms of sociodemographic characteristics. Little is known about varying validities of change trajectories in different subgroups. Moreover, person characteristics might also have distorted the reported change trajectories. For example, the exact form of developmental change in life satisfaction is still subject to a controversial debate (cf. Ulloa et al., 2013). Some authors identified a linear decline in life satisfaction over time (Frijters, & Beaton, 2012) or curvilinear change trajectories (Cheng, Powdthavee, & Oswald, 2015;

Easterlin, 2006), whereas others reported little change (Kassenboehmer & Haisken-DeNew, 2012) or even a slight increase in life satisfaction over the life course (Sutin et al., 2013). Similarly, critical life events can have a differential impact on well-being in different subgroups of respondents (see Paul & Moser, 2009, for respective effects with regard to unemployment). Therefore, future studies could adopt cohort or subgroup analyses to identify boundary conditions with regard to the reported results. Third, seasonal effects might have influenced self-reported well-being scores. For example, in our surveys the instruments were always administered during summer. Previous research revealed that life satisfaction increases in warm, sunny weathers as compared to winterly weathers (Connolly 2012; Schkade & Kahnemann, 1998). Similarly, people's employment status may vary over the cycle of a year due to seasonal employment, which is quite common in many European countries. Therefore, it is recommended to emphasize situational conditions in future research on life satisfaction. Moreover, since the current study investigated only a specific type of critical life event (i.e., non-employment), it seems important to replicate the results with other critical life events (e.g., death of a loved person or birth of a child). Finally, our study spanned a rather short time. Developmental changes were examined over a six-year period. A more informative picture of changes in life satisfaction over the life course would be given by longer periods that span several decades from youth to old age. These designs would also allow for the examination of potential non-linear relationship and could contribute to the prevalent discussion on the precise form of change in life satisfaction (see Cheng et al., 2015; Frijters, & Beaton, 2012; Kassenboehmer & Haisken-DeNew, 2012; Sutin et al., 2013). Similarly, three measurement occasions are rather uninformative to identify non-linear effects due to a critical life event. In the present study, we had to fix the respective parameters to *a priori* derived values rather than estimating the change trajectory from the data. Increasing the number of measurement occasions would also allow for the identification of different stages

before and after the critical life event that might exhibit different forms of change (see Anusic, Yap, & Lucas, 2013).

Conclusion

The measurement of continuous change in people's life satisfaction with single items identifies similar change trajectories as the use of multi-item scales. In contrast, episodic change due to a critical life event (i.e., non-employment) is more strongly reflected by multi-item scales. Moreover, estimates of interindividual differences in intraindividual change are notably smaller using single items. Thus, single items are likely to yield low power for identifying moderating influences on the change trajectories.

Acknowledgments

In this paper we make use of data of the LISS (Longitudinal Internet Studies for the Social Sciences) panel administered by CentERdata (Tilburg University, The Netherlands) through its MESS project funded by the Netherlands Organization for Scientific Research.

References

- Abdel-Khalek, A. M. (2006). Measuring happiness with a single-item scale. *Social Behavior and Personality, 34*, 139-150. doi:10.2224/sbp.2006.34.2.139
- Anusic, I., & Schimmack, U. (2016). Stability and change of personality traits, self-esteem, and well-being: Introducing the meta-analytic stability and change model of retest correlations. *Journal of Personality and Social Psychology, 110*, 766-781. doi:10.1037/pspp0000066
- Anusic, I., Yap, S. C., & Lucas, R. E. (2014). Testing set-point theory in a Swiss national sample: Reaction and adaptation to major life events. *Social Indicators Research, 119*, 1265-1288. doi:10.1007/s11205-013-0541-2
- Appel, M., Gnambs, T., & Maio, G. (2012). A short measure of the need for affect. *Journal of Personality Assessment, 94*, 418-426. doi:10.1080/00223891.2012.666921-
- Bergkvist, L. (2015). Appropriate use of single-item measures is here to stay. *Marketing Letters, 26*, 245-255. doi:10.1007/s11002-014-9325-y
- Bergkvist, L., & Rossiter, J. R. (2007). The predictive validity of multiple-item versus single-item measures of the same constructs. *Journal of Marketing Research, 44*, 175-184. doi:10.1509/jmkr.44.2.175
- Bishop, J., Geiser, C., & Cole, D. A. (2015). Modeling latent growth with multiple indicators: A comparison of three approaches. *Psychological Methods, 20*, 43-62. doi:10.1037/met0000018
- Blanchflower, D. G., & Oswald, A. J. (2008). Is well-being U-shaped over the life cycle?. *Social Science & Medicine, 66*, 1733-1749. doi:10.1016/j.socscimed.2008.01.030
- Blom, A. G., Bosnjak, M., Cornilleau, A., Cousteaux, A. S., Das, M., Douhou, S., & Krieger, U. (2016). A comparison of four probability-based online and mixed-mode panels in Europe. *Social Science Computer Review, 34*, 8-25. doi:10.1177/0894439315574825

- Cantril, H. (1965). *The Patterns of Human Concern*. New Brunswick, NJ: Rutgers University Press.
- Cheng, T. C., Powdthavee, N., & Oswald, A. J. (2015). Longitudinal evidence for a midlife nadir in human well-being: Results from four data sets. *Economic Journal*. Advance online publication. doi:10.1111/eoj.12256
- 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*, 2809-2818. doi:10.1007/s11136-014-0726-4
- Connolly, M. (2013). Some like it mild and not too wet: The influence of weather on subjective well-being. *Journal of Happiness Studies, 14*, 457-473. doi:10.1007/s10902-012-9338-2
- Credé, M., Harms, P., Niehorster, S., & Gaye-Valentine, A. (2012). An evaluation of the consequences of using short measures of the Big Five personality traits. *Journal of Personality and Social Psychology, 102*, 874-888. doi:10.1037/a0027403-
- Deaton, A. (2008). Income, health, and well-being around the world: Evidence from the Gallup World Poll. *Journal of Economic Perspectives, 22*, 53-72. doi:10.1257/jep.22.2.53
- Deinzer, R., Steyer, R., Eid, M., Notz, P., Schwenkmezger, P., Ostendorf, F., & Neubauer, A. (1995). Situational effects in trait assessment: The FPI, NEOFFI, and EPI questionnaires. *European Journal of Personality, 9*, 1-23. doi:10.1002/per.2410090102
- Diamantopoulos, A., Sarstedt, M., Fuchs, C., Wilczynski, P., & Kaiser, S. (2012). Guidelines for choosing between multi-item and single-item scales for construct measurement: a predictive validity perspective. *Journal of the Academy of Marketing Science, 40*, 434-449. doi:10.1007/s11747-011-0300-3

- Diener, M. L., & Diener McGavran, M. B. (2008). What makes people happy? A developmental approach to the literature on family relationships and well-being. In M. Eid & R. J. Larsen (Eds.), *The science of subjective well-being* (pp. 347-375). New York, NY: Guilford Press.
- Diener, E. D., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment, 49*, 71-75.
doi:10.1207/s15327752jpa4901_13
- Diersch, N., & Walther, E. (2016). The impact of question format, context, and content on survey answers in early and late adolescence. *Journal of Official Statistics, 32*, 307-328.
- Easterlin, R. (2006). Life cycle happiness and its sources: Intersections of psychology, economics, and demography. *Journal of Economic Psychology, 27*, 463-482.
doi:10.1016/j.joep.2006.05.002
- Eid, M., & Diener, E. (2004). Global judgments of subjective well-being: Situational variability and long-term stability. *Social Indicators Research, 65*, 245-277.
doi:10.1023/B:SOCI.0000003801.89195.bc
- Eid, M., & Hoffmann, L. (1998). Measuring variability and change with an item response model for polytomous variables. *Journal of Educational and Behavioral Statistics, 23*, 193-215. doi:10.3102/10769986023003193
- Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling, 8*, 430-457. doi:10.1207/S15328007SEM0803_5
- Fisher, G. G., Matthews, R. A., & Gibbons, A. M. (2016). Developing and investigating the use of single-item measures in organizational research. *Journal of Occupational Health Psychology, 21*, 3-23. doi:10.1037/a0039139

- Frijters, P. & Beaton, T. (2012). The mystery of the U-shaped relationship between happiness and age. *Journal of Economic Behaviour and Organisation*, 82, 525-542.
doi:10.1016/j.jebo.2012.03.008
- Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opinion Quarterly*, 73, 349-360.
doi:10.1093/poq/nfp031
- Gallagher, M. W., Lopez, S. J., & Preacher, K. J. (2009). The hierarchical structure of well-being. *Journal of Personality*, 77, 1025-1050. doi:10.1111/j.1467-6494.2009.00573.x
- Geiser, C., Keller, B., & Lockhart, G. (2013). First versus second order latent growth curve models: Some insights from latent state-trait theory. *Structural Equation Modeling*, 20, 479-503. doi:10.1080/10705511.2013.797832
- Gnambs, T. (2014). A meta-analysis of dependability coefficients (test-retest reliabilities) for measures of the Big Five. *Journal of Research in Personality*, 52, 20-28.
doi:10.1016/j.jrp.2014.06.003
- Gnambs, T. (2015a). Facets of measurement error for scores of the Big Five: Three reliability generalizations. *Personality and Individual Differences*, 84, 84-89.
doi:10.1016/j.paid.2014.08.019
- Gnambs, T. (2015b). Sociodemographic effects on the test-retest reliability of the Big Five Inventory. *European Journal of Psychological Assessment*. Advance online publication. doi:10.1027/1015-5759/a000259
- Gnambs, T., Appel, M., Schreiner, C., Richter, T., & Isberner, M.-B. (2014). Experiencing narrative worlds: A latent state-trait analysis. *Personality and Individual Differences*, 69, 187-192. doi:10.1016/j.paid.2014.05.034
- Gnambs, T., Stiglbauer, B., & Selenko, E. (2015). Psychological effects of (non)employment: A cross-national comparison of the United States and Japan. *Scandinavian Journal of Psychology*, 56, 659-669. doi:10.1111/sjop.12240

- Gogol, K., Brunner, M., Goetz, T., Martin, R., Ugen, S., Keller, U., Fischbach, A., & Preckel, F. (2014). "My questionnaire is too long!" The assessments of motivational-affective constructs with three-item and single-item measures. *Contemporary Educational Psychology, 39*, 188-205. doi:10.1016/j.cedpsych.2014.04.002
- Hayes, A. F., & Cai, L. (2007). Using heteroskedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation. *Behavior Research Methods, 39*, 709-722. doi:10.3758/BF03192961
- Holbrook, A. L., Green, M. C., & Krosnick, J. A. (2003). Telephone versus face-to-face interviewing of national probability samples with long questionnaires: Comparisons of respondent satisficing and social desirability response bias. *Public Opinion Quarterly, 67*, 79-125. doi:10.1086/346010
- Kaczmarek, L. D., Bujacz, A., & Eid, M. (2015). Comparative latent state-trait analysis of satisfaction with life measures: The Steen Happiness Index and the Satisfaction with Life Scale. *Journal of Happiness Studies, 16*, 443-453. doi:10.1007/s10902-014-9517-4
- Kassenboehmer, S. C., & Haisken-DeNew, J. P. (2012). Heresy or enlightenment? The well-being age U-shape effect is flat. *Economic Letters, 117*, 235-238. doi:10.1016/j.econlet.2012.05.013
- Knäuper, B., Schwarz, N., Park, D., & Fritsch, A. (2007). The perils of interpreting age differences in attitude reports: Question order effects decrease with age. *Journal of Official Statistics, 23*, 515-528.
- Knowles, E. S., & Byers, B. (1996). Reliability shifts in measurement reactivity: Driven by content engagement or self-engagement? *Journal of Personality and Social Psychology, 70*, 1080-1090. doi:10.1037/0022-3514.70.5.1080
- Krueger, A. B., & Schkade, D. A. (2008). The reliability of subjective well-being measures. *Journal of Public Economics, 92*, 1833-1845. doi:10.1016/j.jpubeco.2007.12.015

- Lucas, R. E., Clark, A. E., Georgellis, Y., & Diener, E. (2004). Unemployment alters the set point for life satisfaction. *Psychological Science, 15*, 8-13. doi:10.1111/j.0963-7214.2004.01501002.x
- Lucas, R. E., & Donnellan, M. B. (2007). How stable is happiness? Using the STARTS model to estimate the stability of life satisfaction. *Journal of Research in Personality, 41*, 1091-1098. doi:10.1016/j.jrp.2006.11.005
- Lucas, R. E., & Donnellan, M. B. (2012). Estimating the reliability of single-item life satisfaction measures: Results from four national panel studies. *Social Indicators Research, 105*, 323-331. doi:10.1007/s11205-011-9783-z
- McArdle, J. J., & Grimm, K. J. (2010). Five steps in latent curve and latent change score modeling with longitudinal data. In K. van Monfort, J. H. L. Our, & A. Satorra (Eds.), *Longitudinal research with latent variables* (pp. 245-273). Berlin, Germany: Springer.
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Erlbaum.
- McDonald, R. P. (2004). Respecifying improper structures. *Structural Equation Modeling, 11*, 194-209. doi:10.1207/s15328007sem1102_3
- Newman, D. A. (2003). Longitudinal modeling with randomly and systematically missing data: A simulation of ad hoc, maximum likelihood, and multiple imputation techniques. *Organizational Research Methods, 6*, 328-362. doi:10.1177/1094428103254673
- Pavot, W., & Diener, E. (1993). The affective and cognitive context of self-reported measures of subjective well-being. *Social Indicators Research, 28*, 1-20. doi:10.1007/BF01086714
- Rammstedt, B., & Beierlein, C. (2015). Can't we make it any shorter?. *Journal of Individual Differences, 35*, 212-220. doi:10.1027/1614-0001/a000141

- Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: a quantitative review of longitudinal studies. *Psychological Bulletin*, *126*, 3-25. doi:10.1037/0033-2909.126.1.3
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*, 1-36. doi:10.18637/jss.v048.i02
- Sandvik, E., Diener, E., & Seidlitz, L. (1993). Subjective well-being: The convergence and stability of self-report and non-self-report measures. *Journal of Personality*, *61*, 317-342. doi:10.1111/j.1467-6494.1993.tb00283.x
- Schkade, D. A., & Kahneman, D. (1998). Does living in California make people happy? A focusing illusion in judgments of life satisfaction. *Psychological Science*, *9*, 340-346. doi:10.1111/1467-9280.00066
- Schimmack, U., Krause, P., Wagner, G. G., & Schupp, J. (2010). Stability and change of well being: An experimentally enhanced latent state-trait-error analysis. *Social Indicators Research*, *95*, 19-31. doi:10.1007/s11205-009-9443-8
- Schimmack, U., & Oishi, S. (2005). The influence of chronically and temporarily accessible information on life satisfaction judgments. *Journal of Personality and Social Psychology*, *89*, 395-406. doi:10.1037/0022-3514.89.3.395.
- Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461-464, doi:10.1214/aos/1176344136
- Schwarz, N. (2003). Self-reports in consumer research: The challenge of comparing cohorts and cultures. *Journal of Consumer Research*, *29*, 588-594. doi:10.1086/346253
- Schwarz, N., Strack, F., & Mai, H. P. (1991). Assimilation and contrast effects in part-whole question sequences: A conversational logic analysis. *Public Opinion Quarterly*, *55*, 3-23. doi:10.1086/269239

- Schweizer, K. (2010). Some guidelines concerning the modeling of traits and abilities in test construction. *European Journal of Psychological Assessment, 26*, 1-2.
doi:10.1027/1015-5759/a000001
- Soto, C. J. (2015). Is happiness good for your personality? Concurrent and prospective relations of the Big Five with subjective well-being. *Journal of Personality, 83*, 45-55.
doi:10.1111/jopy.12081
- Steyer, R., Mayer, A., Geiser, C., & Cole, D. A. (2015). A theory of states and traits-revised. *Annual Review of Clinical Psychology, 11*, 71-98. doi:10.1146/annurev-clinpsy-032813-153719
- Sutin, A. R., Terracciano, A., Milaneschi, Y., An, Y., Ferrucci, L., & Zonderman, A. B. (2013). The effect of birth cohort on well-being: The legacy of economic hard times. *Psychological Science, 24*, 379-385. doi:10.1177/0956797612459658.
- Thalmeyer, A. G., Saucier, G., & Eigenhuis, A. (2011). Comparative validity of brief and medium-length Big Five and Big Six personality questionnaires. *Psychological Assessment, 23*, 995-1009. doi:10.1037/a0024165
- Tourangeau, R., & Bradburn, N. M. (2010). The psychology of survey response. In P. V. Marsden & J. D. Wright (Eds.), *Handbook of survey research* (pp. 315-346). Bingley, United Kingdom: Emerald.
- Tucker-Drob, E. M., & Briley, D. A. (2014). Continuity of genetic and environmental influences on cognition across the life span: A meta-analysis of longitudinal twin and adoption studies. *Psychological Bulletin, 140*, 949-979. doi:10.1037/a0035893
- Tyler, T. R., & Schuller, R. A. (1991). Aging and attitude change. *Journal of Personality and Social Psychology, 61*, 689-697. doi:10.1037/0022-3514.61.5.689
- Ulloa, B. F. L., Møller, V., & Sousa-Poza, A. (2013). How does subjective well-being evolve with age? A literature review. *Journal of Population Ageing, 6*, 227-246.
doi:10.1007/s12062-013-9085-0

- Van Vaerenbergh, Y., & Thomas, T. D. (2013). Response styles in survey research: A literature review of antecedents, consequences, and remedies. *International Journal of Public Opinion Research, 25*, 195-217. doi:10.1093/ijpor/eds021
- Waldron, S. (2010). *Measuring subjective well-being in the UK*. Newport, United Kingdom: Office for National Statistics.
- Yap, S. C., Anusic, I., & Lucas, R. E. (2014). Does happiness change? Evidence from longitudinal studies. In K. M. Sheldon, & R. E. Lucas (Eds.), *Stability of happiness: Theories and evidence on whether happiness can change* (pp. 127-145). London, England: Academic Press. doi:10.1016/B978-0-12-411478-4.00007-2
- Yuan, K., & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological Methodology, 30*, 167-202. doi:10.1111/0081-1750.00078
- Ziegler, M., Ehrlenspiel, F., & Brand, R. (2009). Latent state-trait theory: An application in sport psychology. *Psychology of Sport and Exercise, 10*, 344-349. doi:10.1016/j.psychsport.2008.12.004
- Ziegler, M., Kemper, C. J., & Krueger, P. (2015). Short scales - Five misunderstandings and ways to overcome them. *Journal of Individual Differences, 35*, 185-189. doi:10.1027/1614-0001/a000148

Table 1.

Sample Characteristics and Tests for Sample Attrition

Socio-demographics		Descriptive statistics	Tests for sample attrition ^a
Sex:	Women	54%	$\chi^2(2) = 1.77, p = .41,$ Cramer's $V = .02$
	Men	46%	
Age:		$M = 48.18$ ($SD = 15.45$)	$F(2, 4,031) = 7.16,$ $p < .001, \eta^2 = .00$
Education:	Less than secondary school	36%	$\chi^2(6) = 9.28, p = .16,$ Cramer's $V = .03$
	Secondary school (US: high school)	34%	
	Vocational education (US: college)	23%	
	University	7%	
Civil status:	Married	62%	$\chi^2(4) = 28.36, p < .001,$ Cramer's $V = .06$
	Separated / divorced / widowed	12%	
	Single	25%	
Occupation:	Paid employment / self-employed	58%	$\chi^2(8) = 20.76, p = .01,$ Cramer's $V = .05$
	Job-seeker	1%	
	Student	7%	
	Homemaker	10%	
	Pensioner	16%	

Note. $N = 4,034$. Sample characteristics at the first measurement occasion in 2008.

^a Comparisons between frequency of participation (2 vs. 3 vs. 4 times)

Table 2.

Summary Statistics for Continuous Change Model

	First measurement				Second measurement				Third measurement				Fourth measurement				<i>M</i>	<i>SD</i>	ω_h
	LS	HP	LL	SL	LS	HP	LL	SL	LS	HP	LL	SL	LS	HP	LL	SL			
LS		1.02	0.87	0.77	1.45	0.96	0.83	0.71	0.79	0.70	0.66	0.55	1.65	1.32	1.18	1.08	7.58	1.36	-
HP	.83		1.01	0.88	0.90	1.37	0.94	0.84	0.68	0.87	0.79	0.67	1.24	1.68	1.34	1.20	7.53	1.30	-
LL	.51	.51		0.96	0.77	0.93	1.37	0.91	0.60	0.73	0.92	0.68	1.07	1.28	1.71	1.34	7.47	1.31	-
SL	.70	.67	.55		0.69	0.81	0.87	1.36	0.60	0.71	0.82	0.89	0.99	1.17	1.34	1.79	7.43	1.32	.84
LS	.58	.54	.47	.56		0.90	0.79	0.69	0.72	0.63	0.59	0.50	1.48	1.18	1.07	0.99	7.65	1.29	-
HP	.58	.58	.48	.55	.87		0.91	0.80	0.64	0.80	0.73	0.60	1.14	1.51	1.21	1.11	7.61	1.22	-
LL	.42	.41	.55	.46	.56	.55		0.87	0.57	0.67	0.86	0.62	1.00	1.18	1.53	1.20	7.55	1.22	-
SL	.55	.52	.49	.66	.73	.70	.56		0.56	0.65	0.75	0.79	0.94	1.08	1.21	1.60	7.54	1.21	.84
LS	.50	.47	.41	.48	.61	.60	.47	.56		0.75	0.72	0.67	1.07	0.98	0.91	0.92	7.35	1.14	-
HP	.51	.51	.42	.48	.60	.62	.46	.55	.86		0.82	0.81	0.96	1.21	1.10	1.06	7.29	1.21	-
LL	.40	.38	.52	.45	.50	.50	.56	.53	.58	.57		0.87	0.94	1.13	1.32	1.24	7.27	1.22	-
SL	.48	.46	.45	.59	.57	.55	.50	.70	.72	.69	.59		0.82	0.98	1.09	1.32	7.23	1.27	.84
LS	.43	.41	.40	.44	.54	.52	.45	.51	.58	.57	.52	.57		2.02	1.85	1.74	6.90	1.72	-
HP	.44	.45	.41	.45	.56	.56	.45	.52	.60	.62	.51	.56	.85		2.25	2.06	6.83	1.77	-
LL	.32	.31	.47	.37	.43	.41	.54	.45	.42	.41	.57	.48	.54	.52		2.34	6.83	1.82	-
SL	.44	.42	.44	.55	.52	.51	.48	.64	.58	.55	.56	.71	.74	.72	.57		6.75	1.84	.85

Note. Correlations are shown below the diagonal, and covariances are shown above the diagonal. Convergent validities are in bold. All correlations are significant at $p < .05$. Missing values were handled using pairwise deletion. ω_h = Omega hierarchical reliability (cf. McDonald, 1999); LS = Life satisfaction item, HP = Life happiness item, LL = Ladder of life item, SL = Satisfaction with life scale.

Table 3.

Parameter Estimates of the Indicator-Specific Growth Models

	Continuous Change Model		Episodic Change Model	
	<i>M</i> / λ (<i>SE</i>)	<i>VAR</i> (<i>SE</i>)	<i>M</i> / λ (<i>SE</i>)	<i>VAR</i> (<i>SE</i>)
<i>Latent intercept factor means and variances</i>				
Life satisfaction item	7.58* (0.02)	1.13* (0.06)	7.45* (0.06)	0.80* (0.16)
Ladder of life item	7.34* (0.02)	0.75* (0.04)	7.22* (0.05)	0.71* (0.10)
Life happiness item	7.64* (0.02)	1.02* (0.05)	7.50* (0.06)	0.77* (0.13)
Satisfaction with Life Scale	6.90* (0.03)	2.10* (0.08)	6.76* (0.09)	2.17* (0.31)
<i>Latent slope factor means and variances</i>				
Life satisfaction item	-0.06* (0.01)	0.07* (0.01)	-0.10* (0.02)	0.05 (0.08)
Ladder of life item	-0.05* (0.01)	0.03* (0.01)	-0.07* (0.02)	0.00 (0.00)
Life happiness item	-0.05* (0.01)	0.06* (0.01)	-0.05* (0.02)	0.07 (0.06)
Satisfaction with Life Scale	-0.06* (0.01)	0.12* (0.01)	-0.16* (0.03)	0.18* (0.08)
<i>Latent state residual factor means and variances</i>				
Measurement occasion T1	0.00 [†]	0.54* (0.05)	0.00 [†]	0.66* (0.28)
Measurement occasion T2	0.00 [†]	0.48* (0.04)	0.00 [†]	0.64* (0.17)
Measurement occasion T3	0.00 [†]	0.50* (0.04)	0.00 [†]	0.10 (0.49)
Measurement occasion T4	0.00 [†]	0.39* (0.05)	-	-
<i>Latent state residual factor loadings</i>				
Life satisfaction item	1.00 [‡]		1.00 [‡]	
Ladder of life item	0.27* (0.02)		0.35* (0.08)	
Life happiness item	0.86* (0.02)		0.90* (0.11)	
Satisfaction with Life Scale	0.71* (0.03)		0.81* (0.14)	

Note. λ = Unstandardized factor loading (with measurement invariance across measurement occasions), *SE* = Standard error. For the episodic change model the respondents were in paid employment at the first and second measurement occasion and out of the labor force (e.g., unemployed, pensioner, homemakers) at the third measurement occasion.

* $p < .05$; [†] By definition fixed parameter according to LST theory; [‡] Fixed for identification.

Table 3. (continued)

	Continuous Change Model		Episodic Change Model	
	$M / \lambda (SE)$	$VAR (SE)$	$M / \lambda (SE)$	$VAR (SE)$
<i>Latent error factor means and variances</i>				
Life satisfaction item T1	0.00 [†]	0.25* (0.03)	0.00 [†]	0.22* (0.09)
Life satisfaction item T2	0.00 [†]	0.16* (0.02)	0.00 [†]	0.25* (0.06)
Life satisfaction item T3	0.00 [†]	0.18* (0.02)	0.00 [†]	0.45* (0.12)
Life satisfaction item T4	0.00 [†]	0.26* (0.04)	0.00 [†]	-
Ladder of life item T1	0.00 [†]	0.55* (0.04)	0.00 [†]	0.43* (0.09)
Ladder of life item T2	0.00 [†]	0.60* (0.03)	0.00 [†]	0.55* (0.09)
Ladder of life item T3	0.00 [†]	0.55* (0.03)	0.00 [†]	0.80* (0.15)
Ladder of life item T4	0.00 [†]	0.65* (0.06)	0.00 [†]	-
Life happiness item T1	0.00 [†]	0.29* (0.02)	0.00 [†]	0.24* (0.08)
Life happiness item T2	0.00 [†]	0.19* (0.02)	0.00 [†]	0.26* (0.06)
Life happiness item T3	0.00 [†]	0.19* (0.01)	0.00 [†]	0.11 (0.11)
Life happiness item T4	0.00 [†]	0.19* (0.02)	0.00 [†]	-
Satisfaction with Life Scale T1	0.00 [†]	0.73* (0.05)	0.00 [†]	0.74* (0.17)
Satisfaction with Life Scale T2	0.00 [†]	0.71* (0.03)	0.00 [†]	0.82* (0.10)
Satisfaction with Life Scale T3	0.00 [†]	0.74* (0.05)	0.00 [†]	0.66* (0.32)
Satisfaction with Life Scale T4	0.00 [†]	0.61* (0.04)	0.00 [†]	-

Table 4.

Correlations between Latent Intercept and Slope Factors for Continuous Change Model

	<i>Intercept Factors</i>				<i>Slope Factors</i>			
	1.	2.	3.	4.	1.	2.	3.	4.
1. Life satisfaction item	-	[.73, .81]	[.96, .98]	[.86, .91]	-	[.55, .86]	[.96, 1.00]	[.86, 1.00]
2. Ladder of life item	.78	-	[.72, .81]	[.78, .85]	.68	-	[.51, .95]	[.83, 1.00]
3. Life happiness item	.97	.77	-	[.81, .86]	1.01 ^a	.70	-	[.74, .91]
4. Satisfaction with Life Scale	.88	.82	.84	-	.90	.95	.77	-

Note. The 95% bias-corrected confidence intervals above the diagonal are based on 1,000 bootstrap samples. All correlations are significantly ($p < .05$) different from 0.00. ^a Correlation is not significantly ($z = 0.28, p = .61$) different from 1.00.

Table 5.

Regressions of Variance Components of the Continuous Change Model on Age Cohorts and Type of Measure

	Consistency		Occasion-Specificity		Reliability	
	Model 1	Model 2	Model 1 ^c	Model 2 ^c	Model 1 ^c	Model 2 ^c
Intercept	0.58* (0.01)	0.64* (0.02)	0.26* (0.02)	0.12* (0.02)	0.84* (0.01)	0.76* (0.02)
1. Age cohort ^a	0.02* (0.00)	0.03* (0.01)	-0.03* (0.01)	-0.02* (0.01)	0.00 (0.00)	0.01 (0.01)
2. Measure: Life satisfaction item ^b		-0.11* (0.03)		0.23* (0.02)		0.12* (0.02)
3. Measure: Ladder of life item ^b		-0.09* (0.03)				
4. Measure: Life happiness item ^b		-0.06* (0.03)		0.20* (0.02)		0.13* (0.02)
5. Interaction 1. x 2.		0.01 (0.01)		-0.01 (0.01)		-0.01 (0.01)
6. Interaction 1. x 3.		-0.02 (0.01)				
7. Interaction 1. x 4.		-0.01 (0.01)		-0.02* (0.01)		-0.02* (0.01)
Adjusted R^2	.17	.50	.15	.84	.00	.50

Note. Weighted regressions of the variance components derived from the indicator-specific latent state-trait growth models in six age cohorts (see main text). Weights were proportional to the sample sizes of the six age cohorts. The table presents the unstandardized regression coefficients with standard errors in parentheses.

^a Coded 0 to 5. ^b Dummy-coded using the multi-item scale as reference category. ^c Excludes ladder of life item (see footnotes 3 and 4).

* $p < .05$.

Table 6.

Summary Statistics for Episodic Change Model

	First measurement				Second measurement				Third measurement				<i>M</i>	<i>SD</i>
	LS	HP	LL	SL	LS	HP	LL	SL	LS	HP	LL	SL		
LS		1.30	0.59	1.33	0.78	0.69	0.60	1.12	0.97	0.77	0.71	0.98	7.49	1.25
HP	.85		0.55	1.28	0.76	0.71	0.58	1.06	0.87	0.70	0.62	0.95	7.54	1.22
LL	.47	.47		0.91	0.61	0.48	0.62	0.92	0.70	0.64	0.84	0.88	7.36	1.04
SL	.64	.62	.53		1.12	1.01	0.92	2.06	1.23	0.93	0.92	1.75	6.84	1.68
LS	.46	.45	.43	.48		1.41	0.99	2.01	1.30	1.18	0.95	1.55	7.38	1.39
HP	.49	.49	.39	.51	.83		0.76	1.74	1.11	1.06	0.75	1.36	7.47	1.26
LL	.44	.44	.55	.47	.61	.52		1.34	0.96	0.83	0.94	1.20	7.13	1.21
SL	.48	.45	.47	.64	.74	.72	.58		1.75	1.56	1.33	2.57	6.67	1.95
LS	.53	.49	.48	.51	.62	.59	.53	.59		1.87	1.48	2.60	7.14	1.59
HP	.52	.48	.51	.47	.66	.65	.54	.61	.86		1.27	2.05	7.35	1.38
LL	.43	.38	.61	.42	.51	.44	.58	.50	.66	.66		2.00	7.01	1.42
SL	.41	.41	.44	.54	.55	.53	.50	.65	.78	.71	.66		6.28	2.11

Note. Correlations are shown below the diagonal, and covariances are shown above the diagonal. Convergent validities are in bold. All correlations are significant at $p < .05$. Missing values were handled using pairwise deletion. LS = Life satisfaction item, HP = Life happiness item, LL = Ladder of life item, SL = Satisfaction with life scale. All respondents were in paid employment at the first and second measurement occasion and out of the labor force (e.g., unemployed, pensioner, homemakers) at the third measurement occasion.

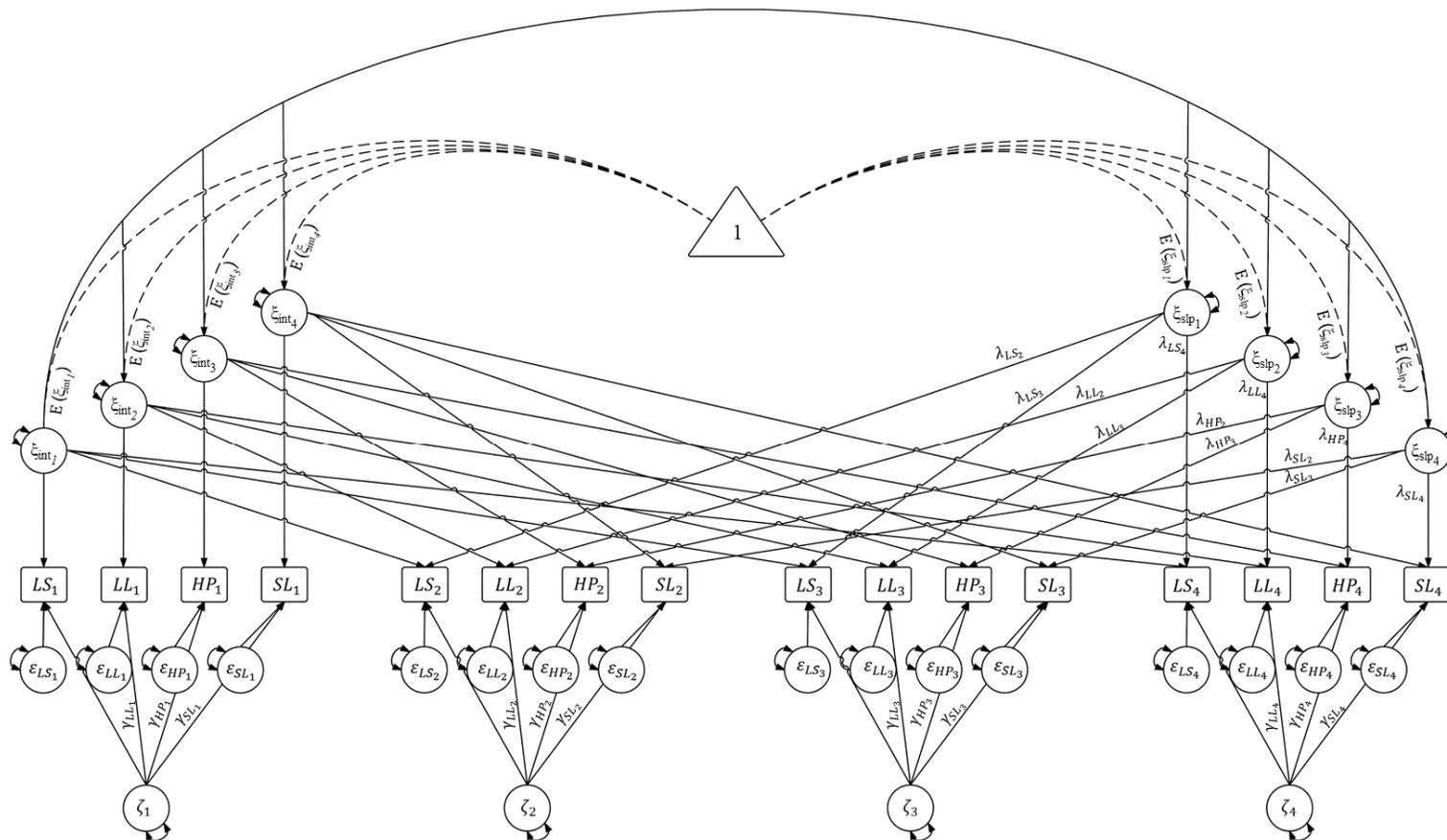


Figure 1. Path diagram of the indicator-specific latent state-trait growth model for single items on life satisfaction (LS), ladder of life (LL), and happiness (HP), and the multi-item Subjective Well-Being Scale (SL). Paths without labels (except for covariances) were fixed to 1.

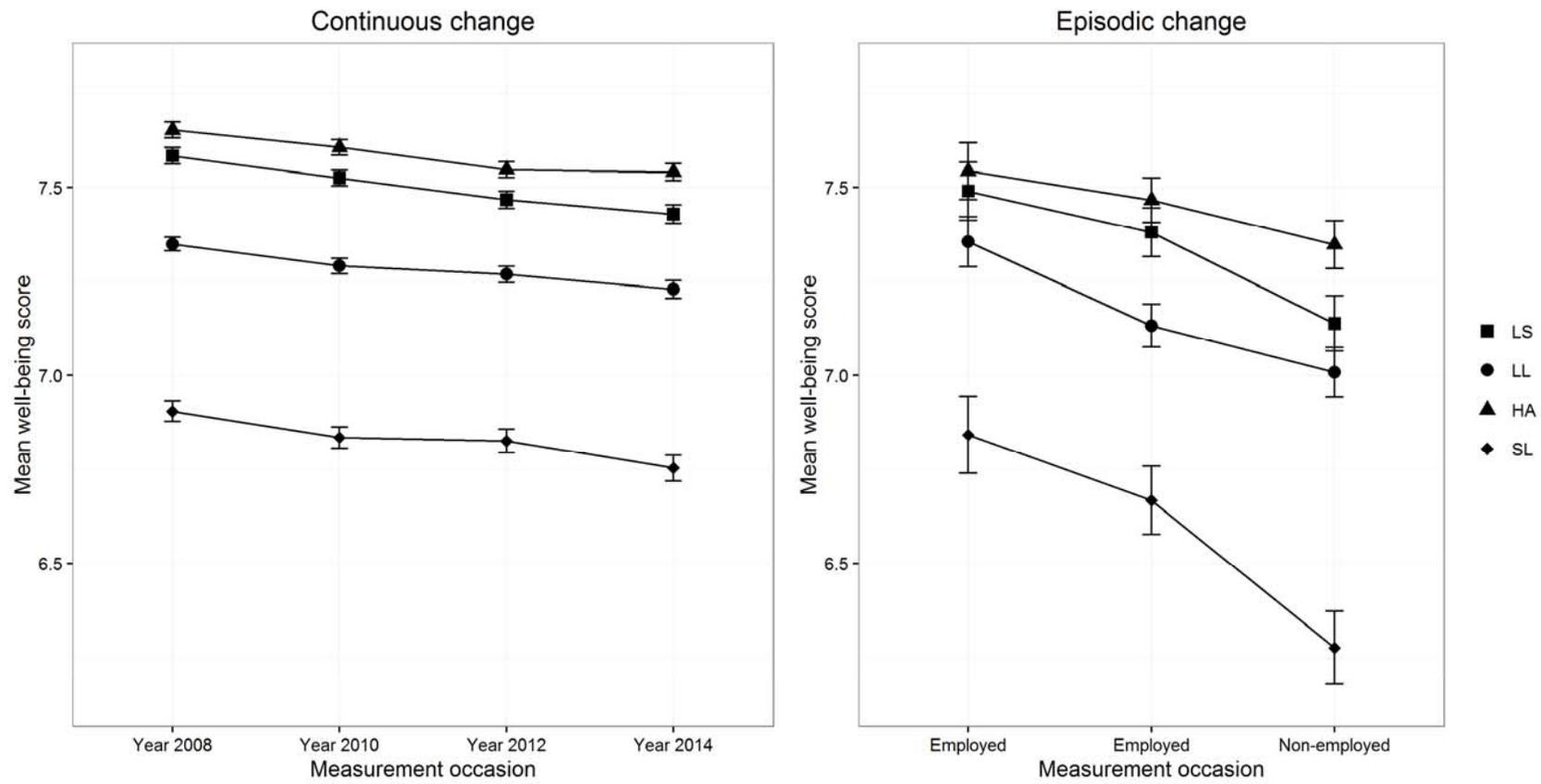


Figure 2. Observed mean scores including standard errors for single items on life satisfaction (LS), ladder of life (LL), and happiness (HP), and the multi-item Subjective Well-Being Scale (SL).

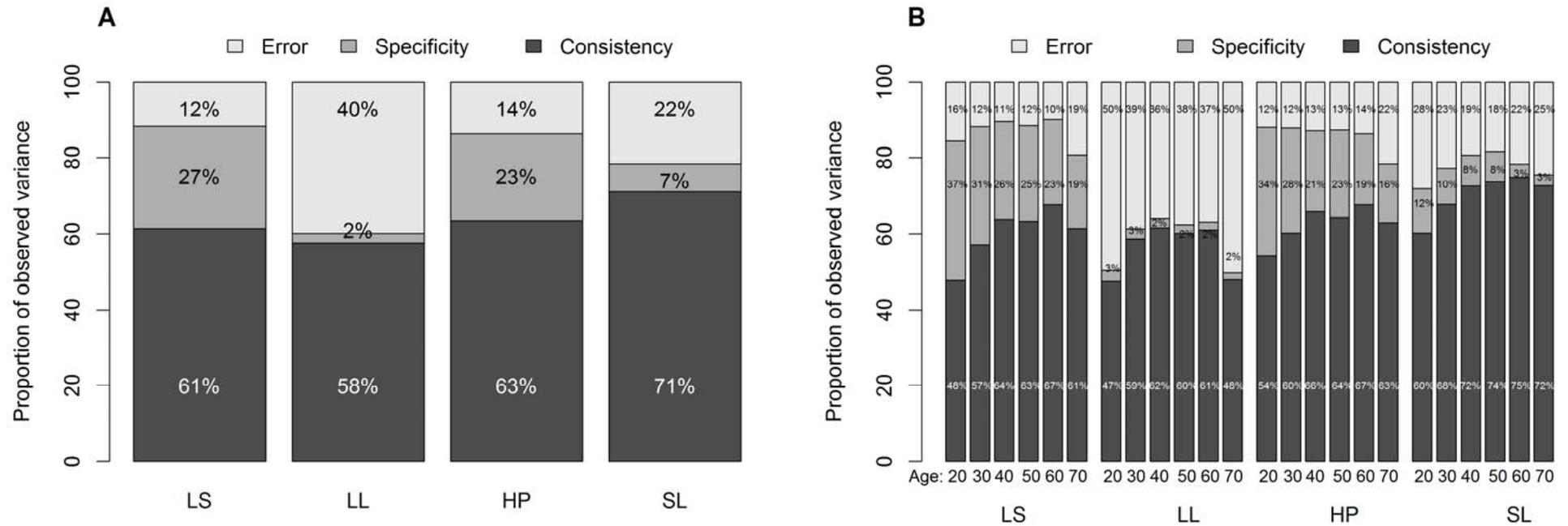


Figure 3. Consistency, occasion-specificity, and measurement error in observed score variance for single items on life satisfaction (LS), ladder of life (LL), and happiness (HP), and the multi-item Subjective Well-Being Scale (SL) in the continuous change model. A: Overall sample, B: By age cohort.

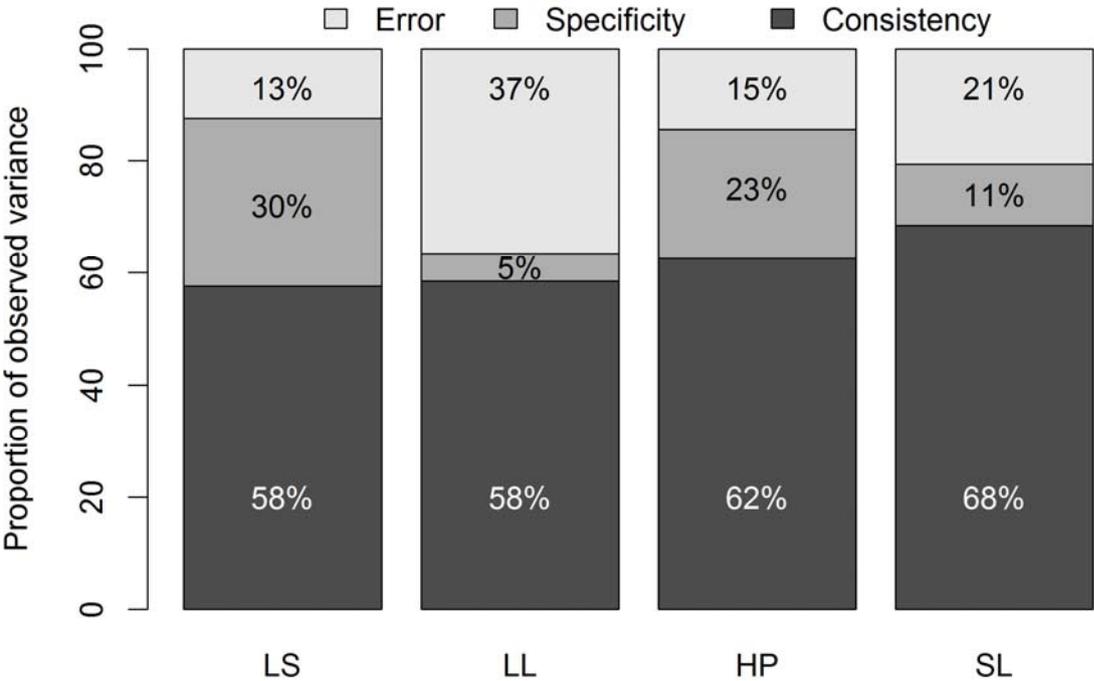


Figure 4. Consistency, occasion-specificity, and measurement error in observed score variance for single items on life satisfaction (LS), ladder of life (LL), and happiness (HP), and the multi-item Subjective Well-Being Scale (SL) in the episodic change model

Appendix: Satisfaction with Life Scale (Diener et al., 1985)

Below are five statements with which you may agree or disagree. Using the 1-7 scale below, indicate your agreement with each item by placing the appropriate number on the line preceding that item. Please be open and honest in your responding.

1. In most ways my life is close to my ideal.
2. The conditions of my life are excellent.
3. I am satisfied with my life.
4. So far I have gotten the important things I want in life.
5. If I could live my life over, I would change almost nothing.

Response scale: 1 = strongly disagree, 2 = disagree, 3 = slightly disagree, 4 = neither agree, nor disagree, 5 = slightly agree, 6 = agree, 7 = strongly agree