Stability and Change of Well Being: An Experimentally Enhanced Latent State-Trait-Error Analysis

Ulrich Schimmack · Peter Krause · Gert G. Wagner · Jürgen Schupp

Abstract   This study uses longitudinal panel data and short-term retest data from the same respondents in the German Socio-economic Panel to estimate the contribution of state and trait variance to the reliable variance in judgments of life satisfaction and domain satisfaction. The key finding is that state and trait variance contribute approximately equally to the reliable variance in well being measures. Most of the occasion specific variance is random measurement error, although occasion-specific variation in state variance makes a reliable contribution for some measures. Moreover, the study shows high similarity in life satisfaction and average domain satisfaction for the stable trait component ($r = .97$), indicating that these two measures are influenced by the same stable dispositions. In contrast, state variance of the two measures is distinct, although still highly correlated ($r = .77$). Error variances of the two measures are only weakly correlated, indicating that most of the error component is indeed due to random measurement error.

Keywords   Life satisfaction · Domain satisfaction · Well being · Stability · Change · Set point · Disposition · State · Trait

1 Introduction

Well being is typically defined as a life that matches an individual's ideal. The most widely used measures of well being are subjective reports of satisfaction with life in general or

U. Schimmack (✉)
University of Toronto, Mississauga, ON, Canada
e-mail: uli.schimmack@utoronto.ca

P. Krause · G. G. Wagner · J. Schupp
German Socio-Economic Panel Study (SOEP), Berlin, Germany

G. G. Wagner
Berlin University of Technology, Berlin, Germany

J. Schupp
Free University of Berlin, Berlin, Germany

various important life domains. The main goal of empirical studies of well being is to explain variance in well being measures. In other words, if well being were a constant that would not vary across individuals, cultures, or over time, it would be relatively uninteresting to study well being empirically. The current interest in well being research can be attributed to the fact that well being varies across individuals (Diener 1984), over time (Ehrhardt et al. 2000; Schimmack and Oishi 2005), and across cultures (Deaton 2008; Veenhoven and Hagerty 2006). A better understanding of the causes of this variability may help individuals and policy makers to make better decisions that increase well being. The main aim of this article is to examine the contribution of random measurement error, true changes in well being, and the influence of stable dispositions to the total variance in a variety of well being measures.

At the most abstract level of analysis, the total variance in well being reports can be decomposed into three components: Trait variance, state variance, and error variance (Eid and Diener 2004; Kenny and Zautra 1995; Schimmack and Lucas 2007a). Trait variance in well being reveals stability in individuals' well being over long periods of time that would theoretically persist indefinitely. The stable component produces positive retest correlations over short and long retest intervals. State variance in well being reveals systematic changes in individuals' well being over time. These changes produce stability in well being in the short term, but instability over longer time intervals. The third variance component, random error variance, accounts for the fact that empirical studies of well being have to rely on unreliable measures of well being, especially in costly national representative surveys. Although random error variance does not systematically bias the results of empirical studies, it is dangerous to ignore it (Schmidt and Hunter 1996).

To illustrate, consider the common finding that wealth is a strong predictor of well being measures in studies of nations, but a fairly weak predictor in studies of individuals within a nation. The direct comparison of effect sizes at the national level and the individual level fails to take measurement error into account. Comparisons of national averages are based on averages across hundreds or thousands of reports. Thus, random measurement error has a relatively small effect on the correlation at the national level. In contrast, random measurement error can severely attenuate the correlation across individuals within nations.

Measurement error can also severely bias estimates of trait variance due to genetic differences between individuals (Schmidt and Hunter 1996). Twin studies typically find heritability coefficients ranging from .3 to .5 (Lykken and Tellegen 1996; Nes et al. 2006; Stubbe et al. 2005). These coefficients are typically interpreted as evidence that the remaining 50–70% of the variance are explained by environmental factors. However, this interpretation ignores the fact that environmental variance includes measurement error. Thus, environmental influences are overestimated and heritability is underestimated. In short, it is difficult and problematic to interpret observed correlations without taking random error variance into account.

Psychologists typically rely on internal consistency to assess measurement error. However, the use of internal consistency as a measure of reliability rests on several assumptions that are likely to be violated in empirical studies (Schmidt et al. 2003). For example, numerous biases in well being reports are likely to influence multiple items of a well-being scale in the same direction. These shared biases accounted for account 10% of the variance in life satisfaction judgments (Eid and Diener 2004). Retest correlations over short time periods that allow for little true changes in well being provide more accurate information about reliability, especially if the retest interval is sufficiently long to eliminate memory effects (Saris et al. 1998). Another advantage of retest reliability is that it can be

used to assess the reliability of single-item indicators which are more commonly used in large representative panel studies of well being (Ehrhardt et al. 2000; Fujita and Diener 2005).

1.1 Short-Term Retest Correlations

One approach to the assessment of reliability is to ask the same question twice within a very short time interval. If the time interval is sufficiently short to assume that the true level of happiness has not changed (e.g., 20 min or 1 h), retest correlations provide direct information about the amount of measurement error. Furthermore, these retest correlations can still be inflated by shared method variance. For example, memory effects may inflate the correlation between repeated answers to the same question within the same interview (Scherpenzeel and Saris 1997). Thus, estimates of error variance are conservative, and the true amount of error variance is likely to be larger. In an influential article on the measurement of well being, Schwarz and Strack (1999) claimed that "measures of SWB have low test-retest correlations, usually hovering around .40, and not exceeding .60 when the same question is asked twice during a 1-h interview" (p. 62). If this were true, correlations of .40 between MZ twins would suggest that well being is nearly perfectly heritable because even two reports of the same twin correlate .4. Thus, one cannot expect a stronger correlation between the reports of twins.

Several meta-analyses show that this statement misrepresents the actual stability of well being judgments (Saris et al. 1998; Scherpenzeel and Saris 1997; Schimmack and Oishi 2005; Veenhoven 1994). Multiple item-scales reduce measurement error and show retest correlations above .8 over retest intervals of 1 month (Andrews and Withey 1976; Eid and Diener 2004; Headey et al. 1991; Pavot and Diener 1993; Saris et al. 1998; Schimmack and Oishi 2005; Veenhoven 1994). Even single-item measures consistently exceed Schwarz and Strack's (1999) upper limit of $r = .60$ (Schimmack and Oishi 2005). One meta-analysis of large surveys with multiple happiness measures produced an average short-term retest correlation of about .6, and revealed several factors that increase or decrease reliability (e.g., response format; Saris et al. 1998; Scherpenzeel and Saris 1997). In sum, short-term retest correlations show that well being judgments contain a substantial amount of reliable information, although they fail to provide unambiguous evidence about the precise amount of reliable variance in well being judgments.

1.2 Long-Term Retest Correlations

Retest correlations over longer time periods can be used to estimate the amount of state and trait variance in well being reports (Conley 1984; Ehrhardt et al. 2000; Heise 1969; Schimmack and Lucas 2007a; Schimmack and Oishi 2005). Several articles have used retest correlations in one way or another to make inferences about the amount of error, state, and trait variance in well being reports. Unfortunately, researchers have drawn different and sometimes contradictory conclusions from these findings. For example, Costa et al. (1987) reported a retest correlation of $r = .48$ over a 9-year retest interval. Based on this finding and additional analyses, the authors concluded that "environmental influences on subjective well-being seem to be limited in magnitude, duration, and scope." A similar argument was made by Lykken and Tellegen (1996) based on a similar retest correlation over a 10 year retest interval. Other researchers have reached different conclusions based on the same retest correlation. Veenhoven (1994) obtained a very similar estimate for the 10-year stability of SWB of "about 40% after 10 years" (p. 110). However, he concluded

that this finding suggests that environmental factors have a strong effect on well being, and that traits may explain less than 20% of the variance in well being. Importantly, these radically different conclusions are not based on contradictory evidence because all authors based their conclusions on a retest correlation of about .4 to .6 over a 10-year interval. Rather the conclusions rest on different assumptions about the nature of the changing variance component. The trait interpretation assumes that most of the unstable variance reflects random measurement error. In contrast, the environmental interpretation assumes that the same variance reveals the influence of environmental factors on well being. To test these conflicting assumptions, it is necessary to estimate and remove the contribution of random measurement error before retest correlations are used to make inferences about the contribution of state and trait variance to the reliable variance in well being measures (Schimmack and Lucas 2007b).

1.3 The Trait-State-Error Model

To separate trait variance and state variance it is necessary to study long time intervals. Furthermore, a minimum of four assessments is needed if retest correlations are influenced by measurement error (Kenny and Zautra 1995). Not surprisingly, empirical data that fulfill these requirements are rare. One alternative to a single study with repeated assessments is to conduct a meta-analysis of different studies with varying time intervals. Veenhoven (1994) conducted the first meta-analysis of this kind. The meta-analysis revealed that retest stability gradually decreased with increasing time intervals between repeated assessments, but the retest correlations did not drop to zero. Veenhoven (1994) suggested that the asymptote is around .20. An asymptote of .20 would suggest 20% trait variance. However, this estimate of trait variance is attenuated by measurement error. This can be seen by the fact that, the trendline in Veenhoven's (1994) figure crosses the y-axis, representing a time interval of 0, at about .7. Evidently, the true stability for a time interval of 0 should be 1.00. Thus, the observed value for a time interval of zero provides a reliability estimate. Adjusting the asymptote of .20 accordingly, yields a corrected estimate of trait variance of close to 30%. This estimate may still underestimate the amount of trait variance because Veenhoven (1994) had limited information about the long-term stability of well being. For example, the estimate for 40-year stability was based on informant ratings in two very small samples ($N = 81$ combined). Given the small sample sizes, the confidence interval for the asymptote is large, and it is difficult to obtain a precise estimate of the amount of stable trait variance.

Schimmack and Oishi (2005) conducted another meta-analysis of the retest stability of happiness measures, which included new data based on larger samples that were published after Veenhoven's (1994) seminal study. In addition, the meta-analysis distinguished between single-item and multiple-item measures of happiness. This meta-analysis suggests a slightly higher asymptote around .25. The main surprising finding in Schimmack and Oishi's (2005) meta-analysis was that the asymptote for multiple item measures was the same as for single-item measures. Given the higher reliability of multiple item measures, this finding implies a lower estimate of the true trait variance based on the multiple item studies. Based on the single-item studies, the estimated true trait variance would be 50%. In contrast, the estimate based on multiple item measures would be 36%. Despite this inconsistency, the meta-analytic evidence suggests that 30–50% of the reliable variance in happiness measures is trait variance. This finding implies that the majority, 50–70% of the reliable variance is state variance, a finding quite consistent with the hypothesis that happiness is not a trait, and that environmental factors have a strong influence on SWB (Diener 1984).

Stones et al. (1995) challenged Veenhoven's (1994) conclusion and claimed that "much of the variance in happiness is due to stable individual differences" and that "happiness is a durable individual difference dimension having limited (but significant) reactivity to situations" (p. 142). A figure that summarizes the results of three studies suggested 80–99% trait variance, compared to 1–20% state variance. The apparent contradiction can be easily resolved. Stones et al. (1995) examined treatment effects in studies with short time intervals in specialized relatively homogeneous populations. Due to the short time intervals, a substantial portion of the trait variance is actually state variance that would have changed if a longer time interval had been examined. Furthermore, the authors included changes not due to treatment in the error component of their model. Thus, Stones et al.'s critique of Veenhoven (1994) is flawed by an inappropriate definition of trait and a lack of empirical data that can distinguish trait variance from state variance that is highly stable over short retest periods (Veenhoven 1998).

In response to Stones et al. (1995), Veenhoven and colleagues provided the first longitudinal study of the amount of error, state, and trait variance in life satisfaction ratings (Ehrhardt et al. 2000). The study relied on annual life-satisfaction ratings over the first ten waves of the German Socio-Economic Panel Study (SOEP). The results of this study were consistent with the meta-analytic findings: reliability of the single life-satisfaction item was about 58% and trait and state variance accounted for approximately half of the reliable variance. This finding was replicated in a more recent analysis of the German SOEP data with more waves (Schimmack and Lucas 2007a, b), and a study of a British panel study (Lucas and Donnellan 2007). In sum, decompositions of total variance in well being into state, trait, and error variance suggest that the amount of error variance depends on the reliability of the measure. Thus, comparisons of state and trait variance have to be corrected for unreliability before meaningful comparisons across studies are possible. After correction, trait and state variance account for roughly equal portions of the reliable variance.

1.4 Limitations of the Trait State Error Model

Ehrhardt et al. (2000) noted one limitation of studies that estimate trait, state, and error variance based on annual retests. Annual assessments of well being cannot distinguish between measurement error and occasion specific state variance; that is state variance due to factors that have a relatively brief influence on well being. To illustrate, somebody may win the lottery shortly after the first assessment. This event could still increase well being nearly 1 year later on the second assessment, but may no longer influence life satisfaction on the third assessment nearly 2 years after the event. In fact, the effect of many life events on well being is often short-lived (Suh et al. 1996). Thus, a model that relies on shared variance across consecutive waves to estimate state variance may overestimate error variance at the expense of state variance. To correct for this bias (Ehrhardt et al. (2000) estimated short-term stability of life satisfaction measures based on studies with shorter retest intervals. Based on an estimate of 23% error variance in this study, the authors concluded that nearly half of the error variance in a study with annual assessments is in fact occasion specific state variance. If this assumption were correct, state variance would account for a much larger portion of the total variance than the trait-state-error model suggests. However, it is problematic to rely on other samples to estimate error variance. Moreover, the assumption that only 23% of the total variance is error variance seems questionable given meta-analytic studies of short-term stability (Saris et al. 1998), and the fact that even repeated assessments of the same question within a 1-h interview suggest

that 30–40% of the variance is error variance (Andrews and Withey 1976; Headey et al. 1991; Schimmack and Oishi 2005).

Another problem is that error variance in panel studies varies across waves, which makes it even more difficult to rely on results from studies with one retest to separate occasion specific state variance from error variance in panel studies with multiple waves (Ehrhardt et al. 2000; Frick et al. 2006). In sum, applications of the trait-state-error model to data with annual retests are likely to underestimate the contribution of state variance to the total variance in well being. However, so far it has been difficult to examine the relative contribution of random measurement error and state variance to the error component in the TSE model.

1.5 New Evidence

The present study goes beyond previous studies in several ways. First, the study not only examines global life satisfaction ratings, but also examines the amount of error, state, and trait variance in judgments of domain satisfaction in four domains as well as the average domain satisfaction in the four domains. Second, the study combines information about long-term stability based on repeated annual assessments over 15 years with information about short-term reliability based on a 6-week retest interval. Third, the study takes advantage of recent advances in structural equation modeling to obtain more precise parameter estimates than previous studies. Fourth, the study goes beyond the study of single constructs and examines for the first time the correlation between trait, state, and error variances of life satisfaction and domain satisfaction. Importantly, correlations of error variances in life-satisfaction and domain satisfaction reveal that some of the error variance is not random, and may indeed reflect occasion specific state variance. Finally, the study compares long-term participants in a panel study to participants who answered well being questions for the first time. This comparison tests practice effects on reliability estimates (Ehrhardt et al. 2000; Frick et al. 2006).

2 Method

The data were obtained in the context of the SOEP conducted by the German Institute for Economic Research (DIW). The panel data have been used in numerous longitudinal studies of well being (Ehrhardt et al. 2000; Lucas and Donnellan 2007; Schimmack and Lucas 2007b). In addition to the panel data, several additional data were collected. First, a subsample of the participants in the regular study with annual retests completed a retest after 6 weeks. Second, a second national representative sample was recruited and completed the same questions 6 weeks apart. The measures used for this study are the well being measures that are routinely assessed in the SOEP, namely several questions about domain satisfaction (household income, health, housing, leisure) and a global life-satisfaction item. Participants respond with ratings on an 11-point scale ranging from 0 = totally dissatisfied to 10 = totally satisfied.

3 Results

Representativeness of retest SOEP sub-sample Retest correlations in earlier waves (Waves 1–22) in the panel sample were first compared to those in the total SOEP sample, excluding

Table 1 Observed retest correlations of life-satisfaction and domain satisfaction ratings

Measure	SOEP	Retest sample				New sample
	1–22*	1–22*	22–23a	22–23b	23a–23b	23a–23b
Life	.55	.48	.62	.57	.71[a]	.55[a]
Health	.61	.51	.62	.55	.68	.60
Household income	.59	.60	.63	.63	.79[a]	.68[a]
Housing	.56	.44	.56	.58	.75[a]	.61[a]
Recreation	.52	.48	.60	.57	.59	.53

SOEP, results for all SOEP participants; Retest sample, SOEP participants who completed 6-week retest; New sample, first time respondents not participating in SOEP

1–22, waves 1–22; 23a, first assessment; 23b, second assessment after 6-week retest

* Average observed annual stability

[a] coefficients in the same row are significantly different from each other

the panel sample (Table 1). Correlations were comparable and ranged from .34 to .60, mean $r = .49$, for the retest sample, and from .47 to .64, mean $r = .55$, for the total sample. Thus, the results for the retest sample are likely to generalize to the total sample.

Practice effects The first analyses compared the short-term retest correlations of the participants in the SOEP panel to those who completed the same questionnaire for the first time (Table 1). As expected, correlations were always higher in the panel sample, and four of the six comparisons were significantly different. The difference for the life satisfaction item was substantial. The retest correlation in the first-time sample suggests only a reliability of 55%, compared to 71% reliable variance in the panel sample. This finding is consistent with findings that reliability increases in the panel study (Ehrhardt et al. 2000). A reliability estimate of 55% is similar to the 58% reliability estimate for the first SOEP wave (Ehrhardt et al. 2000), and Sari's estimate of about 60% reliability for a single life satisfaction item. Thus, practice effects may increase reliabilities by about 10–15% percentage points. The presence of practice effects in panel studies is important and needs to be taken into account in studies that examine the true stability of well being. Studies that fail to take practice effects into account may erroneously suggest that stability increases when participants get older, whereas models that allow for increasing reliability show no age effects on reliability (Schilling 2006).

Annual versus 6-week stability Comparisons of annual and 6-week stability were based only on the last waves of the SOEP. The reason is that comparisons based on earlier waves in the SOEP would underestimate the difference between short-term and long-term stabilities due to practice effects and increasing reliabilities over time. Annual stability was estimated based on the correlation between the assessment in the previous year with the assessment in the current year. In addition, the correlation between the previous year and the 6-week retest in the current year was also examined because it also covers approximately the same retest interval of about 1 year (plus 6 weeks). Not surprisingly, both correlations were virtually identical, .45–.63, mean $r = .57$, for the first assessment in the current year, and .43–.63, mean $r = .56$, for the retest in the current year. In comparison, the retest correlations in the current year over the much shorter 6-week retest interval were notably higher, ranging from .59 to .79, mean $r = .72$. According to Cohen (1988), the difference between these two correlations is moderate to large, $q = .55$. The difference between 6-week and annual stability for life satisfaction was about 15% points (.57 vs. .72). This finding suggests that 15% of

the variance in life-satisfaction is state variance rather than random measurement error, as assumed by a model that estimates reliability on the basis of annual retest correlations.

3.1 Trait-State-Error Model

The data of the SOEP panel sample were analyzed with Kenny and Zautra's (1995) trait-state-error model (see Schimmack and Lucas 2007a, for computational details). In short, the model tries to explain the covariations among repeated assessments of the same constructs with four parameters (Fig. 1). First, the total variance is decomposed into three independent variance components that represent trait, state, and error variance. In Fig. 1, a standard measurement model is used to first distinguish error variance (E) and reliable variance (R).

$$\text{Total variance} = \text{Reliable variance} + \text{Error variance} \tag{1}$$

Then reliable variance is decomposed into trait variance and state variance.

$$\text{Reliable variance} = \text{Trait variance} + \text{State variance}. \tag{2}$$

Substituting reliable variance in the Eq. 1 with Eq. 2 yields

$$\text{Total variance} = \text{Trait variance} + \text{State variance} + \text{Error variance} \tag{3}$$

The fourth parameter estimates the stability of state variance between retest intervals (S).

If the model is restricted to four parameters, it assumes that the stability parameter for the state variance and the amount of new variance remains constant. Thus, the model implies the non-linear constrained that

$$\text{State variance t2} = \text{State variance t1} \times \text{stability}^2 + \text{New state variance t2}.$$

It is possible to relax some assumptions of the model and to estimate more parameters. Due to the presence of practice effects, it is most important to allow error variances to vary across waves (Schilling 2006). It is also important to recognize that the stability estimate of state variance depends on the retest interval between assessments. This requires an additional nonlinear constraint because the relation between stabilities for different retest intervals is a power function (Kenny and Zautra 1995).

To equate stabilities of state variances for a 1-year and a 6-week retest interval, the following constraint was used:

Fig. 1 The trait-state-error model. *Note*: The figure does not show the intermediate waves between waves 2 and n. T, trait variance; S, state variance; N, new state variance; β oldS, 1-year stability of state variance; β olds*, 6-week stability of state variance; R, reliable variance; E, error variance; O, observed (Total) variance. The figure shows the complete model; there are no additional residual variances
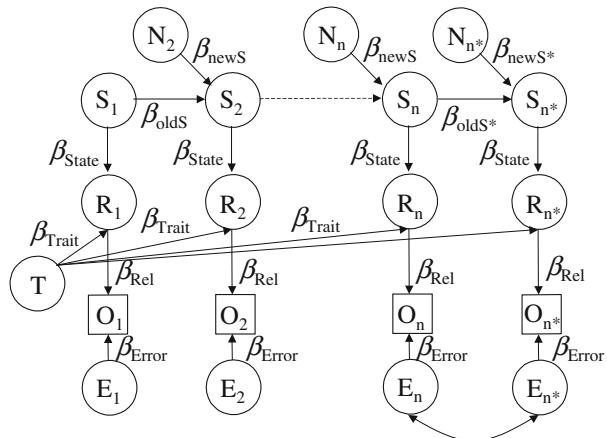
$$\text{6-Week stability} = 1 - (1\text{-Year stability})^{6/52}.$$

Another non-linear equation was used to obtain reliability estimates on each occasion.

$$\text{Reliability} = (\text{State} + \text{Trait})/(\text{State} + \text{Trait} + \text{Error}_{\text{wave}})$$

Because error variances varied across waves, an average reliability estimate was obtained by averaging the separate reliabilities of each wave.

Finally, the error variances of the short-term retest assessments were allowed to correlate to allow for the presence of reliable occasion-specific variance that is lost in reliability estimates based on annual retests.

Parameters were estimated with MPLUS4.2 (Muthén and Muthén 2007), which allows users to specify non-linear constraints. Furthermore, MPLUS4.2 provides confidence intervals for all parameters. The parameter estimates are reported with 99% confidence intervals. Parameters are statistically significant if the confidence interval does not include zero. More important, confidence intervals are useful to compare parameter estimates to previous studies because it is unlikely that parameters are exactly equivalent across studies. The results are reported in Table 2.

The reliability estimate based on the TSE model for life satisfaction is also consistent with previous studies (Ehrhardt et al. 2000; Schilling 2006; Schimmack and Lucas 2007b). A comparison of different domains shows relatively high reliability for household income, and relatively low reliability for recreation and housing. The reliability of life satisfaction falls in the middle. Thus, there is no evidence that global life satisfaction judgments are more difficult and therefore less reliable than domain satisfaction judgments.

Table 2 shows that annual stability is quite similar across different domains. At least in this small sample, none of the differences are statistically significant. Table 2 also shows that the 6-week stability is nearly perfect. This finding shows that studies of short time periods cannot distinguish between the trait and state component of the TSE model because state variance hardly changes.

The final columns show the new evidence about correlations between the error variances over a 6-week retest interval. Confidence intervals for these estimates are quite large because this parameter estimate is based on a single data point, whereas the other estimates are based on multiple observations. Nevertheless, all parameters except the one for recreation are significant. Thus, the TSE model overestimates error variance because it cannot detect occasion-specific state variance. Although the data are insufficient to provide precise estimates of the amount of occasion-specific state variance, the parameter estimates seem

Table 2 Model parameters and 99% confidence intervals of stability and reliability

Measure	Annual S	6-week S	Rel.	Error r
Life	.89 [.84|.94]	.99 [.98|.99]	.63 [.56|.70]	.23 [.03|.43]
Health	.93 [.89|.97]	.99 [.99|1.00]	.59 [.52|.66]	.21 [.01|42]
Household income	.91 [.86|.95]	.99 [.98|1.00]	.68 [.62|.75]	.30 [.10|.50]
Housing	.89 [.82|.97]	.99 [.98|1.00]	.51 [.43|.59]	.42 [.25|.59]
Recreation	.92 [.88|.95]	.99 [.99|.99]	.54 [.48|.60]	.03 [−.21|.27]
Average domain	.92 [.87|.97]	.99 [.98|1.00]	.64 [.57|.71]	.32 [.13|.51]

Annual S, annual stability of state component; 6-week S, 6 week stability of state component; Rel., reliability estimate (1 − error variance); Error r, correlated error between error variance in last wave and 6-week retest after the last wave

lower than one would expect based on Ehrhardt et al.'s assumption that 50% of the error variance in the TSE model represents occasion-specific state variance. For life satisfaction, the estimate of .23 suggests that a quarter of the variance may be a better estimate. Given roughly 40% error variance, this would again lead to an estimate of about 10% of the total variance. Adding 10% occasion-specific state variance to the total state variance has a small effect on the relative amount of state variance versus trait variance, which increases from 50 to 55%.

3.2 Bivariate TSE Model

The next analysis uses a slightly modified version of the bivariate TSE model to examine the relation between life satisfaction and domain satisfaction. Essentially, the bivariate model estimates traits, state, error variance and state stability for each construct. In addition, it estimates the covariations between the four parameters. Kenny and Zautra (1995) also included additional parameters to examine cross-lagged causal effects between the state components of two variables. These parameters were omitted in the present model for two reasons. First, cross-lagged regression analysis is especially problematic when two variables are highly correlated with each other on the same occasion, which creates the problem of multicollinearity. Second, the main focus of the analysis was on the correlation between the error variances, which reveal occasion-specific covariations between life satisfaction and domain satisfaction. The bivariate model examined the relation between life satisfaction and average domain satisfaction. The results of these analyses are similar to those of a more complex model with individual domains as separate predictors because differences in the weights of domains have small effects on the amount of explained variance in life satisfaction (Andrews and Withey 1976).

The parameter estimates for life satisfaction and domain satisfaction in the bivariate model were similar to those in the univariate model (Table 3). The correlations between the two variables varied across components (Table 4). Trait variances were correlated nearly perfectly. This finding suggests that the stable variance in life satisfaction and average domain satisfaction is caused by the same factors. The correlation for state variance was high, but lower than the trait correlation. This finding suggests that at the state level life satisfaction and average domain satisfaction are distinct constructs that change in the same direction. Correlations among the error variance were much lower. Furthermore,

Table 3 Model parameters and 99% confidence intervals of stability and reliability

Measure	Annual true S	6-week true S	TSE Rel.	Corr. error
Life	.88 [.83|.93]	.98 [.98|.99]	.63 [.56|.71]	.23 [.03|.44]
Average domain	.92 [.89|.95]	.99 [.99|1.00]	.64 [.58|.70]	.32 [.15|.51]

Table 4 Correlations between life satisfaction and average domain satisfaction

Component	Correlation
Trait	.97 [.81|1.00]
Initial state	.77 [.50|1.00]
New state	.65 [.52|.79]
Error-concurrent	.18 [.11|.26]
Error-6-week lag	.07 [−.08|.21]

the cross-construct-cross occasion correlation over the 6-week interval was small and not statistically significant in the present sample. This finding suggests that most of the error variance is either occasion specific or measure specific, which is consistent with an interpretation of this variance as error variance.

4 Discussion

Previous studies of the stability of SWB were limited by comparisons of long-term and short-term stability across different samples. This study addressed this limitation by combining information about the long-term stability over 15 years, annual stability, and 6-week stability within a single sample for measures of global life satisfaction, satisfaction with individual domains, and average domain satisfaction. The findings solidify previous estimates that about half of the reliable variance in well being ratings is trait variance, whereas the other half is state variance. Single-item indicators of well being have a reliability of about 60% when respondents answer well being questions for the first time. In panel studies reliability can increase by about 10% points due to practice effects. Furthermore, reliability estimates based on retest correlations over a 1-year period are attenuated due to the presence of reliable state variance that does not last for more than a year. However, the amount of occasion-specific state variance is relatively small, and it is possible that a notable portion of this reliable variance is due to occasion-specific method variance rather than short-term fluctuations in the state variance of well-being. Although these conclusions may not generalize to other populations, the results are consistent with findings in other studies and can serve as a benchmark for future studies.

One major limitation of this study and previous studies is that most studies have focused on cognitive measures of well being, such as life-satisfaction judgments. It is possible that different results would be obtained for affective measures of well being. Individual differences in affect may be more directly influenced by biological dispositions. Thus, trait variance may account for more of the reliable variance in affect measures than in cognitive measures of well being. This hypothesis needs to be tested in panel studies that include measures of life satisfaction and affective well-being.

Another major limitation is that self-report measures of well being can be influenced by systematic rather than mere random measurement error. Thus, future studies need to examine whether the present results would replicate in multi-method studies. For example, studies could use self-ratings and informant ratings to measure well being.

References

Andrews, F. M., & Withey, S. B. (1976). *Social indicators of well-being*. New York: Plenum.

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

Conley, J. J. (1984). The hierarchy of consistency: A review and model of longitudinal findings on adult individual differences in intelligence, personality and self-opinion. *Personality and Individual Differences, 5*(1), 11–25. doi:10.1016/0191-8869(84)90133-8.

Costa, P. T., McCrae, R. R., & Zonderman, A. B. (1987). Environmental and dispositional influences on well-being: Longitudinal follow-up of an American national sample. *The British Journal of Psychology, 78*, 299–306.

Deaton, A. (2008). Income, health, and well-being around the world: Evidence from the Gallup world poll. *The Journal of Economic Perspectives, 22*(2), 53–72. doi:10.1257/jep.22.2.53.

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

Ehrhardt, J. J., Saris, W. E., & Veenhoven, R. (2000). Stability of life-satisfaction over time: Analysis of change in ranks in a national population. *Journal of Happiness Studies, 1*(2), 177–205. doi:10.1023/A:1010084410679.

Eid, M., & Diener, E. (2004). Global judgments of subjective well-being: Situational variability and long-term stability. *Social Indicators Research, 65*(3), 245–277. doi:10.1023/B:SOCI.0000003801.89195.bc.

Frick, J. R., Goebel, J., Schechtman, E., Wagner, G. G., & Yitzhaki, S. (2006). Using analysis of Gini (ANOGI) for detecting whether two subsamples represent the same universe—The German socio-economic panel study (SOEP) experience. *Sociological Methods & Research, 34*(4), 427–468. doi:10.1177/0049124105283109.

Fujita, F., & Diener, E. (2005). Life satisfaction set point: Stability and change. *Journal of Personality and Social Psychology, 88*, 158–164. doi:10.1037/0022-3514.88.1.158.

Headey, B., Veenhoven, R., & Wearing, A. (1991). Top-down versus bottom-up theories of subjective well-being. *Social Indicators Research, 24*(1), 81–100. doi:10.1007/BF00292652.

Heise, D. R. (1969). Separating reliability and stability in test–retest correlation. *American Sociological Review, 34*(1), 93–101. doi:10.2307/2092790.

Kenny, D. A., & Zautra, A. (1995). The trait state error model for multiwave data. *Journal of Consulting and Clinical Psychology, 63*(1), 52–59. doi:10.1037/0022-006X.63.1.52.

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.

Lykken, D., & Tellegen, A. (1996). Happiness is a stochastic phenomenon. *Psychological Science, 7*(3), 186–189. doi:10.1111/j.1467-9280.1996.tb00355.x.

Muthén, L. K., & Muthén, B. O. (2007). *Mplus user's guide* (4th ed.). Los Angeles, CA: Muthén & Muthén.

Nes, R. B., Roysamb, E., Tambs, K., Harris, J. R., & Reichborn-Kjennerud, T. (2006). Subjective well-being: Genetic and environmental contributions to stability and change. *Psychological Medicine, 36*(7), 1033–1042. doi:10.1017/S0033291706007409.

Pavot, W., & Diener, E. (1993). Review of the satisfaction with life scale. *Psychological Assessment, 5*(2), 164–172. doi:10.1037/1040-3590.5.2.164.

Saris, W. E., Van Wijk, T., & Scherpenzeel, A. (1998). Validity and reliability of subjective social indicators—the effect of different measures of association. *Social Indicators Research, 45*(1–3), 173–199. doi:10.1023/A:1006993730546.

Scherpenzeel, A. C., & Saris, W. E. (1997). The validity and reliability of survey questions—a meta-analysis of MTMM studies. *Sociological Methods & Research, 25*(3), 341–383. doi:10.1177/0049124197025003004.

Schilling, O. (2006). Development of life satisfaction in old age: Another view on the "Paradox''. *Social Indicators Research, 75*(2), 241–271. doi:10.1007/s11205-004-5297-2.

Schimmack, U., & Lucas, R. E. (2007a). Environmental influences on well-being: A dyadic latent panel analysis of spousal similarity (submitted).

Schimmack, U., & Lucas, R. E. (2007b). Marriage matters: Spousal similarity in life satisfaction. *Journal of Applied Social Science Studies, 127*, 105–111.

Schimmack, U., & Oishi, S. (2005). The influence of chronically and temporarily accessible information on life satisfaction judgments. *Journal of Personality and Social Psychology, 89*(3), 395–406. doi:10.1037/0022-3514.89.3.395.

Schmidt, F. L., & Hunter, J. E. (1996). Measurement error in psychological research: Lessons from 26 research scenarios. *Psychological Methods, 1*(2), 199–223. doi:10.1037/1082-989X.1.2.199.

Schmidt, F. L., Le, H., & Ilies, R. (2003). Beyond alpha: An empirical examination of the effects of different sources of measurement error on reliability estimates for measures of individual-differences constructs. *Psychological Methods, 8*(2), 206–224. doi:10.1037/1082-989X.8.2.206.

Schwarz, N., & Strack, F. (1999). Reports of subjective well-being: Judgmental processes and their methodological implications. In D. Kahneman, E. Diener, & N. Schwarz (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 61–84). New York, NY: Russell Sage Foundation.

Stones, M. J., Hadjistavropoulos, T., Tuuko, H., & Kozma, A. (1995). Happiness has traitlike and statelike properties: A reply to Veenhoven. *Social Indicators Research, 36*(2), 129–144. doi:10.1007/BF01079722.

Stubbe, J. H., Posthuma, D., Boomsma, D. I., & De Geus, E. J. C. (2005). Heritability of life satisfaction in adults: A twin-family study. *Psychological Medicine, 35*(11), 1581–1588. doi:10.1017/S0033291705005374.

Suh, E., Diener, E., & Fujita, F. (1996). Events and subjective well-being: Only recent events matter. *Journal of Personality and Social Psychology, 70*(5), 1091–1102. doi:10.1037/0022-3514.70.5.1091.

Veenhoven, R. (1994). Is happiness a trait? Tests of the theory that a better society does not make people any happier. *Social Indicators Research, 32*(2), 101–160. doi:10.1007/BF01078732.

Veenhoven, R. (1998). Two state-trait discussions on happiness—a reply to Stones et al. *Social Indicators Research, 43*(3), 211–225. doi:10.1023/A:1006867109976.

Veenhoven, R., & Hagerty, M. (2006). Rising happiness in nations 1946–2004: A reply to Easterlin. *Social Indicators Research, 79*(3), 421–436. doi:10.1007/s11205-005-5074-x.