Social Indicators Research (2006) 75: 241–271
DOI 10.1007/s11205-004-5297-2

OLIVER SCHILLING

DEVELOPMENT OF LIFE SATISFACTION IN OLD AGE: ANOTHER VIEW ON THE "PARADOX"

(Accepted 20 October 2004)

ABSTRACT. Empirical evidence of no age-related decline in life satisfaction (LS) in old age contrasts with frequently observed declines in the objective quality of elder people's lives and has therefore been labelled a "paradox" and interpreted in terms of stability of LS in the respective gerontological discussion. However, as this evidence was mainly derived from cross-sectional age group comparisons, it does neither clearly indicate the absence of age-related mean level change, nor intra-individual stability of LS. Thus, the development of LS in middle and late adulthood was analysed longitudinally by using data from the German Socio-Economic Panel. Based on single item measures of LS across 16 repeated panel waves with one per annum (1984–1999), autoregressive (quasi-Markov) structural equation models were used to estimate true score variances and intra-individual true score stability in one-year intervals. Research questions concerned (a) "monotonic" stability and variance in a subsample of old respondents (born before 1925) as compared to the total sample and (b) change in stability and variances in old age. Results indicate high "monotonic" true score stability of LS over the whole adult life span, whereas mean levels declined slightly in old age. No striking evidence for age-related changes in variance or stability was found.

KEY WORDS: aging, happiness, life satisfaction, longitudinal studies, structural equation modelling, well-being

INTRODUCTION

Old age has frequently shown to be related to multiple psycho-social losses (e.g. Baltes and Mayer, 1999) and increased health risks (e.g. Manton, 1990; Coni et al., 1992; Kolberg, 1999) qualifying late adulthood as a period of decline with regard to objective quality of life. In contrast, several studies provided empirical evidence of no age-related decline in subjective well-being (SWB) in old age (e.g. Larson, 1978; Herzog and Rodgers, 1981; Horley and Lavery, 1995; Smith et al., 1996; Diener and Suh, 1997). The absence of a negative age effect on SWB has been called a "paradox", and theoretical discussion has focussed on explanations for elder people's ability to maintain stable SWB under severe losses (Brandtstädter and Greve, 1994; Baltes and Carstensen, 1996; Staudinger, 2000). Thus, the missing relationship between older age and SWB has been recognized in terms of within-subject changes, drawing attention to the stability of SWB through the process of aging.

A closer look at the empirical evidence for "stability despite loss" in SWB reveals two important limitations: (1) Nonexistence of a negative age effect has been consistently found for measures of life satisfaction (LS), the so-called cognitive component of SWB (Diener et al., 1999), whereas results for affective components of SWB are less clear (Mroczek and Kolarz, 1998; Kunzmann et al., 2000; Ehrlich and Isaacowitz, 2002). Therefore, discussion of the paradox should focus on explanation of stability of LS rather than of SWB in total. (2) The above mentioned findings are largely derived from cross-sectional studies. This means, that no age effect was found in comparisons across different age groups' mean levels at a certain point in time. It is obvious that this does not clearly indicate the absence of an age effect in terms of change in mean levels, since changes related to age and birth cohort may be muddled in cross-sectional comparisons. Thus, these comparisons cannot provide a unique observation of an age effect.

Moreover, evidence for no change in mean levels across old age, whether derived cross-sectionally or longitudinally, cannot definitely indicate stability in terms of within-subject change. On the one hand, in a defined age period mean levels may be stable in the presence of frequent changes in individuals' SWB. On the other hand, changes in mean levels can come along with some kind of within-subject stability. According to Tisak and Meredith (1990), strict, parallel, and monotonic stability could be distinguished. Focussing on change in a variable X measured twice at occasions $t1$ and $t2$, these three types can be characterized in terms of the invariance of certain aspects of the variable's distributions at $t1$ and $t2$: *Monotonic stability* denotes invariance of the rank orders of individuals' values – within-subject change may occur and variances and means may differ, but subjects keep their position within the population in terms of the measured variable. *Parallel stability* could be described as monotonic stability plus invariance of the variances – i.e. the same amount of change occurs for each subject, so that the trajectories of the subjects' values would show as parallel lines from $t1$ to $t2$ in graphic representation. *Strict stability* simply means no change at all and might thus correspond to an intuitive notion of a variable being intraindividually "stable" – implying invariance of rank orders, variances and means. According to this typology, invariance of means is a necessary, though not sufficient condition only for strict stability, whereas invariance of rank orders appears as the basic common condition for all types of stability. Therefore, serial correlation indicating the degree of this invariance has been used as coefficient of stability (Costa et al., 1983).

Altogether, gerontological reasoning about the "paradox" of SWB in older age raises the issue of stability of LS in older age, which cannot be solved from the findings that mostly nurtured this discussion. If no age effect is found, i.e. mean levels revealed to be stable across the old age period or if LS and age do not correlate, this may be the result of older persons' tendencies to maintain their LS throughout the process of aging, but it could not be taken as standalone evidence for these tendencies. To study within-subject change of LS in old age, longitudinal analyses, addressing all facets of stability as mentioned above, are essential.

However, longitudinal studies on this issue are rare. In their review of studies concerning the relation between old age and SWB, Diener and Suh (1997) noticed a lack of longitudinal data as one of the shortcomings in this field of research. Through the past decades, several longitudinal studies in research on ageing have been conducted (for overview see Schaie and Hofer, 2001), but among these, few studies explicitly addressed change in LS (Baur and Okun, 1983; Palmore and Kivett, 1985; Chen, 2001) or dealt with change in overall SWB, including results concerning the stability of satisfaction in older age (Grombach, 1976; Kozma and Stones, 1983; Costa and McCrae, 1984; McNeil et al., 1986; Ferring and Filipp, 1997; Smith and Freund, 2002). Altogether, published longitudinal evidence seems to confirm the nonexistence of age-related changes in mean levels of satisfaction, inasmuch mean differences between subsequent measurements have been found to be small in absolute value and statistically insignificant. Stability of subjects' relative satisfaction over time, as indicated by serial correlations between subsequent measurements of satisfaction or regression coefficients in autoregressive models (predicting satisfaction at some point in time from previously measured satisfaction), seems relatively high: In the studies cited above, shares of common variance in subsequent measures vary, but always appear to be above 25%. Regarding the effect sizes framework given by Cohen (1988), these proportions could be called "large". On the other hand, these results do not support the notion that satisfaction is stable to the extent that changes in subjects' relative satisfaction (i.e. in the rank order of the satisfaction values) are extremely rare.

Drawing definite conclusions about stability of LS from the longitudinal results cited above is hampered by serious methodological limitations: First, differences in the studies' longitudinal design and measurement of LS may be crucial for results on stability. In particular, the duration of the time interval between repeated measurements may have a major influence on the amount of change which could be observed, depending on the speed and causal mechanisms of the changes that occur. For example, constant low speed changes may sum up to minor changes within short intervals, producing high short term stability of LS, but may sum up to substantial changes and thus low stability in the long run. Thus, results concerning stability of LS derived from different studies are not comparable without further consideration of the studies' measurement time intervals.

Second, in analyzing stability, measurement error is important because it may affect serial correlations between repeated measures. According to the typology cited above (Tisak and Meredith, 1990), serial correlations could be regarded as basic facet of stability. If random error components are included in the repeated measurements, means would not be affected, nor would differences in subsequent variances be biased if the variances of the error components do not change across time. But assuming independence of measurement error, both from "true scores" and reciprocally between errors of repeated measurements, serial correlations of the measured satisfaction values will be below true score correlations and therefore underestimate the "true" stability of satisfaction. More advanced models of measurement may complicate this relation but still cause biased estimates of coefficients of stability.

Grombach (1976) reports estimations of true score serial correlations for "satisfaction with present situation", derived from fitting Heise's (1969) earliest version of the autoregressive structural equation model presented in the method section below. These results appear to be somewhat incomprehensible (subsequent two-year serial correlations are -0.26 and -1.0 respectively, overall four-year correlation is 0.9, reliability of measures estimated -0.15), and may be artifacts of too restrictive statistical modelling (implying constant reliabilities and true score variances for all measurements). All the other longitudinal ageing studies mentioned above do not report estimates of error-free coefficients of stability of LS in old age.

Third, in longitudinal studies serious rates of dropout may occur and therefore larger numbers of missing values have to be handled. Listwise deletion of the cases with missing values, as has been applied in all the longitudinal ageing studies mentioned above, can cause bias in the results of statistical analysis (Diggle et al., 2002; Schafer and Graham, 2002).

For the study of change in LS in older age, serious bias may be caused especially by selectivity of drop-out in terms of the variable under examination, i.e. LS and/or changes in LS may be predictors of respondents' willingness or ability to take part in the study. In particular, there may be a "terminal decline" such that ongoing impairment of living conditions may reduce SWB and promote dropout. Applying listwise deletion would exclude these "pre-dropout" changes from analysis, whereas methods which make use of the dropout subjects' existing data would at least include the change that happened between the measurements before dropout.

Given the few longitudinal results on age-related change in LS, limited by the methodological difficulties mentioned above, the issue of stability of LS in older age remains disputable. Going beyond the scope of explaining cross-sectional stability of mean levels, the issue of stability of LS in old age refers to the age-related development of intra-individual variability of satisfaction. Thus, the present work aimed to analyze within-subject stability of older respondents' LS ratings in a large, population based longitudinal database, namely the German Socio-Economic Panel (SOEP).

With regard to possible dynamics of satisfaction in old age, different expectations seem theoretically reasonable and analysis of stability should be driven by assumptions about age-related changes in LS among the elderly. Thus, in the analyses presented below, patterns of change in respondents' satisfaction ratings were examined for evidence on the following hypotheses concerning development of LS in old age:

Lifelong stability: LS may be highly stable across the whole adult life span – including old age. Theoretical reasons for lifelong stability of LS descend from two major sources: First, influence of personality has been an important topic of research on SWB and empirical results show strong correlation between stable personality traits and indicators of SWB (for a review see DeNeve and Cooper, 1998; Diener et al., 2003). In short, it could be presumed that evidence supports the assumption of at least some stabilizing influence of personality on LS. Second, the role of adaptation has been largely debated in theories on SWB: Changes in living conditions may change our satisfaction for a while, but in the long run, this reaction may diminish and we might fall back to our previous level. This "hedonic treadmill" could be explained from discrepancy theories of satisfaction (e.g. Michalos, 1986), based on the idea that the evaluation of one's own life implies some kind of comparison of actual with target states of living conditions. Adaptation to "objective" changes may include a readjustment of standards for comparison towards the new actual state of living conditions and thus a reduction of the discrepancy caused by the initial change. The existence of adaptation effects on SWB has been evidenced empirically, but questions concerning importance and occurrence of such effects are still debatable (for a review see Veenhoven, 1991, 1996; Diener et al., 1999). Altogether, influences of stable traits and adaptation may promote intra-individual stability of LS throughout all periods of life.

Increasing stability: In older age, stability of LS may grow for several reasons. In advanced models of psychological resilience, different modes of coping with losses are maintained, e.g. assimilative, accommodative, and immunizing coping styles as described by Brandstädter (1999), or the processes of selection, compensation, and optimization as described by Baltes and Baltes (1990). Staudinger (1999) proposes that life insight, i.e. knowledge about life and how to evaluate it, may promote the choice of successful adaptational modes. Going one step beyond this argument, it could be assumed that if life insight develops across the life span, this should strengthen adaptive competencies, empowering older people to restore their LS after negative events. Thus, stability of LS may be raised in older ages. Another age-related source of stability could be time perspective: People's judgements about life in general may not only focus on the present state, but also include an evaluation of what has been achieved in the past as well as expectations about future achievements. With increasing age, the steadily growing period of past lifetime may become more relevant. Older people simply have more past time to think about, and thus have more "material" to be judged that does not change any more and promotes the judgement's stability. Few empirical findings support this assumption: Brandtstädter et al. (1997) found a shift from future towards past orientation in meaning perspective of the elderly, and Shmotkin (1991), comparing different age groups' factor loadings of satisfaction with past, present, and future life in representing one common factor, found relevance of past time evaluation elevated for older age.

Increasing variance: Some aspects of older age could promote better rather than worse satisfaction. Apart from social and physical losses coming with age, there may be gains in emotional competencies to regulate emotional experience (Kunzmann et al., 2000), supporting positive emotions. Assuming that satisfaction is partly inferred from affect (Schwarz and Strack, 1991; Veenhoven, 1996), this should uplift LS judgements as well. Also, common perceptions of old age as a period of decline could cause positive satisfaction for individuals growing old without encountering expectable losses: Negative stereotypes about old age may influence old people's onset of standards for the comparison of actual with target states of living conditions. The elder's LS may be based on lowered expectations and therefore elevated for those affected only by minor to moderate losses, while on the other hand, LS of those faced with severe losses may decline accordingly. Thus, transition from middle adulthood into old age could possibly produce "winners" and "losers" in LS. This means intra-individual change in both directions, which should result in declining stability of individuals' relative satisfaction and growing variance of satisfaction ratings observed in samples of elders surveyed longitudinally, whereas mean levels may be found stable across time.

Rather than being mutually exclusive, these hypotheses mark different influences that may overlap to some extent. Due to the effects of personality and adaptation, intra-individual stability of satisfaction may be high across the whole lifespan, but decline slightly in old age because of the "winners and losers" dynamics, or it may even get higher in old age due to stabilizing influences as described above. Comparing the assumptions of age-related increases in stability or variance, these seem to contradict each other as the latter predicts declining stability in older age.

To analyse stability of LS in older age, it seems reasonable to compare the elders' stability with the average stability which can be found over the whole adult age range: Are average, non age-specific stability coefficients high enough to support the assumptions of lifelong stability? Are stability coefficients and/or variances found for a subgroup of elder respondents higher compared to average values found across the whole adult age range, thus promoting the assumption of higher stability or variance in old age? Moreover, making use of a longer time series of repeated measurements of LS, analysis of "change in change" of satisfaction in old age, i.e. longitudinally increasing (or declining) stability coefficients or variances within an age group of elders, is crucial to support the assumptions of age-related growth in stability or variance of LS.

Considering the methodological difficulties mentioned above, statistical analysis was mainly based on a longitudinal structural equation model (SEM), which permits estimation of true score change in satisfaction, as well as state of the art treatment of drop-out. In particular, different model versions representing different patterns of change were fitted to find the pattern describing the dynamics of LS most closely. Parameter estimates of the model chosen are then discussed with regard to the hypotheses stated above.

METHODS

Data

For the subsequent analyses data from the West German subsample of SOEP has been used. SOEP was designed initially as a representative sample of households in former West Germany, with exclusion of households belonging by definition to the main foreigner groups living in Germany (an own subsample has been surveyed for these, for close description see SOEP Group, 2001; Haisken-DeNew and

Frick, 2003). The survey includes collection of person level data from all persons belonging to the sample households. Starting in 1984, 9076 respondents (from 4528 households) aged 16 years and older were sampled. Follow-up of this subsample with annually repeated interviews is ongoing. Data over 16 years, ending with 6045 respondents in 1999, was included in analysis.

In comparison with official statistics, the initial West German SOEP sample has been found highly representative for the overall population, with elderly persons being slightly under-represented (Haisken-DeNew and Frick, 2003). As is always the case in panel studies, attrition may undermine representativeness of the subsequent measured waves. Dealing with this problem, the tracking concept of the panel study is crucial: SOEP follows persons within the survey territory (i.e. Germany) in case of residential mobility. Persons moving in an existing SOEP household as well as all children of SOEP respondents once they reached the age of 16 are to be surveyed. Thus, new persons are constantly entering the sample, which should at least partly compensate for changes in the demographic composition of the sample. Rendtel (1993) found no evidence for participation in the SOEP follow-up surveys being influenced by "classic" demographic variables.

SOEP covers annually repeated measures of LS and thus offers an opportunity to study change in satisfaction within one year or larger intervals, as well as the "change in change" within subsequent one year intervals across a longer period of time. LS is measured by a single item asking for present satisfaction with life to be rated on a 0–10 numerical scale (0 = "absolutely dissatisfied", 10 = "absolutely satisfied").

Statistical analyses presented here were run on two groups of respondents: (1) the total sample of all respondents aged 18 and older (named "ALL"), and (2) for the subgroup of those aged 60 and older at initial survey in 1984, i.e. those born 1924 or earlier ("OLD"). Development of sample sizes (numbers of nonmissing values) is presented below in Table I.

Statistical Modelling

To estimate true score stability of LS, the autoregressive structural equation model termed "quasi-Markov simplex" (QMS) by Jöreskog

TABLE I

Sample sizes, sample moments and first order autocorrelations of observed LS

	ALL					OLD				
	N	$\bar{X}$	S	skew.	R	N	$\bar{X}$	S	skew.	R
1984	9039	7.45	2.09	−0.97	–	1954	7.42	2.28	−0.94	–
1985	8343	7.26	2.01	−0.90	0.46	1652	7.30	2.24	−0.86	0.49
1986	7980	7.33	1.89	−0.90	0.48	1449	7.45	2.13	−1.02	0.53
1987	7842	7.13	1.93	−0.86	0.47	1356	7.23	2.15	−0.97	0.49
1988	7450	7.04	1.96	−0.84	0.51	1208	7.03	2.23	−0.83	0.59
1989	7174	7.10	1.94	−0.85	0.53	1095	7.06	2.15	−0.72	0.59
1990	7005	7.31	1.78	−0.94	0.54	1033	7.13	2.13	−0.77	0.54
1991	6898	7.40	1.70	−1.01	0.55	945	7.32	2.05	−0.94	0.60
1992	6780	7.30	1.70	−0.93	0.55	873	7.10	2.06	−0.80	0.57
1993	6718	7.23	1.78	−0.99	0.56	809	7.12	2.13	−0.86	0.60
1994	6595	7.12	1.78	−0.96	0.57	737	6.98	2.10	−0.82	0.59
1995	6522	7.07	1.80	−0.92	0.58	654	6.77	2.22	−0.77	0.53
1996	6442	7.12	1.75	−0.97	0.59	579	6.97	2.12	−0.77	0.59
1997	6366	6.98	1.79	−0.87	0.60	540	6.83	2.15	−0.71	0.65
1998	6169	7.05	1.77	−1.00	0.58	485	6.80	2.13	−0.77	0.65
1999	6033	7.08	1.77	−1.03	0.58	424	6.75	2.23	−0.72	0.62

Note: N = number of nonmissing values; $\bar{X}$ = sample means; S = standard deviation; skew. = skewness; R = first order autocorrelation with previous year.

(1970) was chosen. For the purpose of modelling the 16 subsequent satisfaction measurements, it can be specified easily with two equations:

$$X_{ti} = \mu_t + T_{ti} + E_{ti} \quad (t = 1, \ldots, 16) \tag{1}$$

$$T_{ti} = \beta_{t-1} T_{(t-1)i} + D_{ti} \quad (t = 2, \ldots, 16) \tag{2}$$

In Equations (1) and (2), X_{ti} names the observed LS value of subject i at panel wave t, whereas T_{ti} names the latent true score of subject i at panel wave t. E_{ti} denotes the measurement error contained in X_{ti}, and D_{ti} denotes the residual of T_{ti} predicted from previous true score $T_{(t-1)i}$. Thus, apart from intercept μ_t, Equation (1) specifies the simple measurement model of classical test theory, whereas Equation (2) specifies a first order autoregressive model for true scores. As proposed by Rudinger and Rietz (1998), fitting the mean structure of the observed variables was incorporated adding intercept μ_t into Equation (1). Figure 1 depicts the "covariance part" of the model resulting

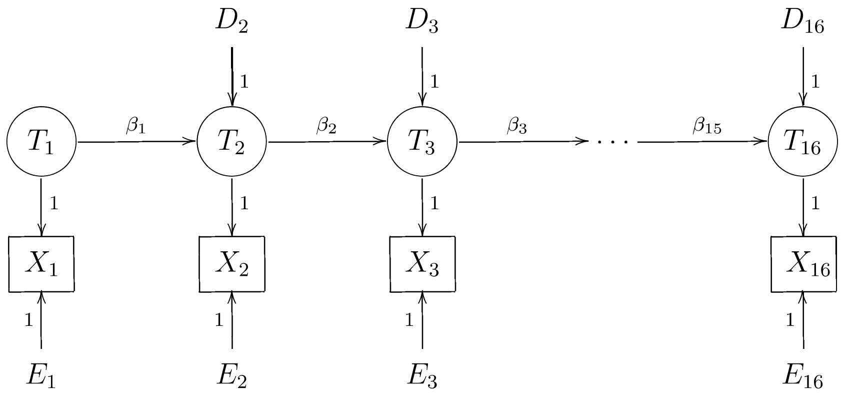

Figure 1. Quasi-Markov simplex.

from Equations (1) and (2) (leaving out the intercepts μ_t). Note that measurement errors are uncorrelated with each other and with true scores.

QMS parameters to be estimated are the unstandardized regression weights β_t, error variances σ^2_{Et}, initial true score variance σ^2_{T1} and residual variances σ^2_{Dt}. Based on these, stability coefficients ρ_t (lag-1 true score correlations, i.e. standardized regression coefficients), true score variances σ^2_{Tt} ($t \geq 2$), and reliabilities can be computed. It has been shown that error variances, true score variances, and unstandardized regression weights between corresponding true scores are identified for all but the first and last measures (Jöreskog, 1970; Werts et al., 1971). Thus, without any further constraints true score variances $\sigma^2_{T2}, \ldots, \sigma^2_{T15}$ and stability coefficients $\rho_2, \ldots, \rho_{14}$ can be estimated. For the implied mean vector, of course all 16 means μ_t are identified.

Within the QMS-framework, stability and exogenous true score variance are estimated for each time interval uniquely, thus the model permits "change in change" and is not restricted to a single pattern of one-year change repeated over 15 time intervals (as, for example, the widely used linear latent growth curve model). Different patterns of change in terms of true score variances and/or stability coefficients can be specified as different QMS model versions by constraining model parameters. Thus, the pattern best characterizing development of LS in the sample investigated can be found through comparisons of nested model versions.

Different patterns of change were ''translated'' into QMS versions as follows:

M1: Unrestricted QMS, i.e. no equality constraints for variances, stability coefficients, or means. Note that for this model, only the ''inner'' coefficients ρ_2 to ρ_{14} and true score variances σ^2_{T2} bis σ^2_{T15} are estimable.

M2: Perfect monotonic stability of LS, i.e. all true score correlations $\rho_t = 1$. This is specified by removing true score random variances, i.e. constraining $\sigma^2_{Dt} = 0$.

M3: Pattern of one-year change invariant over subsequent time intervals, specified by constraining equality of 15 true score random variances and 15 unstandardized stability coefficients, i.e. $\sigma^2_{Dt} = \sigma^2_D$ and $\beta_t = \beta$. Note that these constraints do not imply perfect stability, thus M3 is less restrictive than M2. Note also, that M3 does not imply invariance of stability coefficients or true score variances, but also allows for monotonous increase or decrease in these parameters over the subsequent intervals.

M4: Second order stationarity (e.g. Hershberger et al., 1996), i.e. equality of true score variances and stability coefficients. This is specified by constraining $\sigma^2_{Tt} = \sigma^2_T$ and $\beta_t = \beta$ (these constraints imply ρ_t as well). Note that the first of these constraints is set upon ''indirect'' model parameters, as for $t \geq 2$ true score variances are not estimated directly in fitting the model, but must be computed from estimated values of σ^2_{T1}, σ^2_{Dt}, and β_t. Thus, these constraints are more complicated to handle (see explanations below).

M2 to M4 may be seen as representing fundamental change characteristics (perfect stability, invariance of change pattern, second-order stationarity). To round off constraining true score change, three more model versions were added: If M2 is accepted, also higher restricted models of perfect stability may hold, namely perfect monotonous stability combined with invariant pattern of change (constraining $\sigma^2_{Dt} = 0$ and $\beta_t = \beta$) or perfect parallel stability ($\sigma^2_{Dt} = 0$ and $\beta_t = 1$). If perfect stability and invariance of the change pattern are rejected, a less restrictive model of invariant true score variances may hold ($\sigma^2_{Tt} = \sigma^2_T$), allowing for all changes in the pattern of change which are consistent with constant true score variance, including changing coefficients of stability. For reasons of conciseness, results for this additional model versions will not be reported in detail.

Constraining equality of true score variances and stability coefficients (as in M4) deserves special attention: For $t \geq 2$ true score variance σ^2_{Tt} results from σ^2_{Dt} plus a nonlinear combination of all "previous" true score parameters, increasing in complexity with t. Equalizing two such variances would result in nonlinear equations as well and the same applies to equality constraints for stability coefficients, as $\rho_t = \beta_t \sigma_{Tt} / \sigma_{T(t+1)}$. Unfortunately, at time of computation most existing SEM software packages could not handle the large number of nonlinear constraints that would be necessary to specify invariance of variances and/or stabilities over the 16 panel waves. Rudinger and Rietz (1998) presented an approach to overcome this problem by the use of phantom variables (Rindskopf, 1984). Applying this approach, additional latent predictor variables for T_{ti} with fixed variances were included such that the mathematical composition of σ^2_{Tt} results in the equality of σ^2_{Tt} with a single model parameter, estimable "directly" and controllable by use of simple equality constraints. Given equality of true score variances, equality of stability coefficients could be easily constrained by setting the unstandardized regression coefficients β_t to be equal. A more detailed description of this procedure is given by Schilling (2003).

In addition to models M2 to M4, representing restrictions on the covariance structure of the data, a model version with a restricted mean structure was specified to be compared with unrestricted M1:

M5: Invariance of means, i.e. no mean change over the 15 time intervals or $\mu_t = \mu$. True score variances and stabilities are left unconstrained.

Choice among model versions was based on the logic of pairwise nested model comparisons, accepting the more parsimonious model (with higher degrees of freedom) if no substantial reduction of model fit was observed. Figure 2 depicts the strategy of model comparisons: Initially, unconstrained version M1 has to be accepted with good fit. Regarding the pattern of change in the covariance structure, M1 has to be compared with M2 to M4: First, M1 can be compared with M2. If perfect stability is rejected, M1 can be compared with M3 (if M2 is accepted, M1 can be compared to the

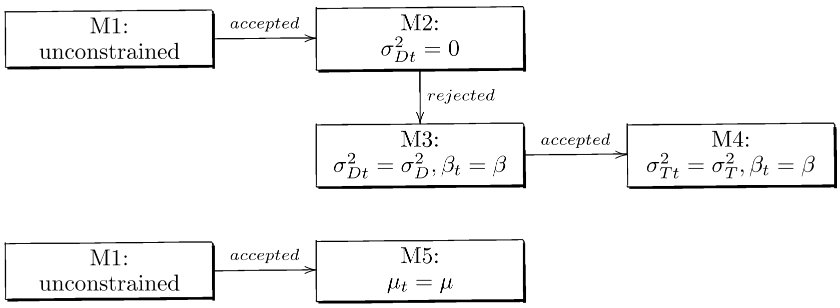

Figure 2. Nested model comparisons.

higher restrictive versions of perfect stability mentioned above). If invariance of change pattern holds, M1 can be compared with M4 (if M3 is rejected, M1 can be compared to the invariance of true score variance model mentioned above). Regarding mean change, M1 is compared with M5.

In pairwise comparisons, decisions were *not* based on the chi-square difference test (likelihood ratio test, see Bollen, 1989). Chi-square testing of model fit is recognized to be associated with serious problems, namely rejection of true models due to high statistical power (Hu and Bentler, 1995) or nonnormality of the data (West et al., 1995). As the chi-square difference test is based on the chi-square statistics of the models compared, it could be argued that it is affected by these problems as well, provoking rejection of restrictions that virtually do not reduce fit to a substantial extent. Among a plethora of fit indices to evaluate model fit, the root mean squared error of approximation (RMSEA), indicating "discrepancy per degree of freedom" (Browne and Cudeck, 1993), has become widely used. Unlike other fit indicators, RMSEA offers an interval estimation approach for assessing model fit (Steiger, 2000), and it showed excellent sensibility to model misspecification in a simulation study presented by Fan et al. (1999). Thus, RMSEA was chosen as basic index for model comparisons. Additionally, considering results and recommendations from Hu and Bentler (1998), Marsh et al. (1996), Hoyle and Panter (1995), incremental fit indices NFI, TLI, CFI, and IFI have been regarded (the namings are these used by the AMOS 4.0 software, for explanation see Arbuckle and Wothke, 1999). Altogether, conditions for accepting the more restricted model in nested model comparisons were set as follows:

- Differences in absolute NFI-, TLI-, CFI-, IFI-, and RMSEA-values of the models compared are below 0.01.
- The RMSEA 90% confidence intervals of the models compared do overlap.

To deal with longitudinal dropout, full information maximum likelihood (FIML) estimation was applied (Arbuckle, 1996), using AMOS 4.0 software (Arbuckle and Wothke, 1999). The development of FIML, as well as other maximum likelihood algorithms for use with missing data (Enders, 2001a), was inspired by the typology for the probabilities of missingness presented by Rubin (1976), who showed that maximum likelihood based missing data procedures could yield unbiased estimates under the less restrictive "missing at random" (MAR), as well as under the "missing completely at random" (MCAR) condition, whereas the widely used listwise deletion of missing data and other "older" data-editing procedures yield unbiased estimates only under MCAR. Compared to these procedures, FIML proved to be more efficient even if MCAR holds (Arbuckle, 1996; Wothke, 2000; Enders and Bandalos, 2001). Reviewing methods for missing data, Schafer and Graham (2002) recommended the use of maximum-likelihood based procedures to deal with longitudinal dropout caused by death of respondents.

FIML estimation could be challenged by the implied assumption of multivariate normal distribution of the data, which may be violated due to "true" nonnormality of latent continuous satisfaction values as well as the observed variables being "coarsely categorized" on an 11-point rating-scale (West et al., 1995). Nonnormality may cause biased significance tests, yielding to many significant results in tests of model parameters and chi-square test of overall model fit. Thus, chi-square testing may produce overrejection of true models when data does not meet the normality assumption. Tests of multivariate normality (Mardia, 1985), as well as standard solutions to overcome these problems, such as the use of the ADF estimator (Browne, 1984) and/or the use of categorized data procedures (e.g. Lee et al., 1995), are applicable only with listwise deletion treatment of missing values. Thus, using FIML may trade off bias caused by suboptimal missing data treatment against bias due to nonnormality of the data. Concerning robustness of FIML estimation against violations of the normality assumption, Enders (2001b) found negligible bias in FIML estimates under moderate violations of normality, similar to results reported for conventional maximum likelihood estimation under nonnormality conditions (Raykov and Widaman, 1995; West et al., 1995). Recommended conditions to trust in conventional maximum likelihood estimates may hold for FIML as well. These include moderate univariate skewness and kurtosis of the observed variables, at least moderate sample sizes and numbers of categories in categorized variables (West et al., 1995).

RESULTS

Descriptive Statistics

Table I shows basic descriptive statistics of the 16 observed LS variables for the total sample of all respondents aged 16 or older (ALL) and the subsample of respondents aged 60 or older in 1984 (OLD). In view of the old respondents' subsequent means, a downward tendency appears, which does not show that clearly for the total sample of all respondents. Comparing both samples' standard deviations, those in the old subsample appear on a higher overall level, which seems relatively stable over subsequent measurements, while in the total sample a slight decrease across the early panel waves is visible.

Univariate skewness values have been printed and demonstrate that violations of normality, as assumed in application of FIML estimation, may be within a moderate range; the same appears in the kurtosis values not printed in Table I, with maximum value 1.71 in the ALL sample (no other value above 1.5), and 1.12 in the OLD sample. Nevertheless, it should be noted that all tests for univariate normality (ALL: Kolmogorov test, OLD: Shapiro–Wilk test, see SAS Institute Inc., 1990) showed significant deviation from normal distribution ($p \leq 0.01$).

Referring to the effect size framework given by Cohen (1988), most first-order autocorrelations printed in Table I could be rated as "large" (≥ 0.5). Of course, all correlations were statistically significant ($p \leq 0.001$). It should be noted that the total correlation matrix, including higher order correlations, shows the simplex structure, i.e. a decline of correlation values related to growing time lag (Jöreskog, 1970).

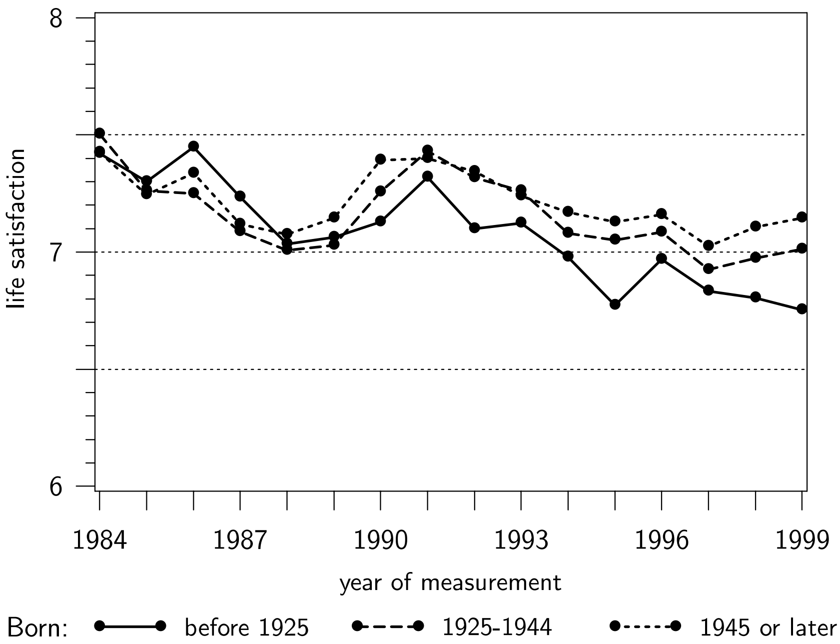

Figure 3. LS means of respondents aged <40, 40–59, and 60+ in 1984.

To round off interpretation of QMS results (see discussion below), the means of the old group are depicted in Figure 3 and contrasted against the means of all respondents aged 40–59 (born 1925–1944) and younger than 40 (born 1945 or later) at the first panel wave in 1984. The Figure shows some common trend in the mean curves for all three groups, namely a slight decline in the mean levels over the first years until 1988, followed by a decrease over the next three years, and again decline from 1991 towards the end of the period under study. Note that this latter downward tendency appears somewhat stronger in the old group as compared to the other groups.

Additionally, to further qualify the changes in the mean levels, effect size measures d have been computed as proposed by Dunlap et al. (1996). In the groups depicted in Figure 3, as well as in the total sample, these are below the threshold of 0.2, proposed as indicating a "small" effect, for all 15 subsequent one-year mean differences. On the other hand, the slight one-year changes sum up to a substantial magnitude over the years: Computed for the overall 1984–1999 mean change $d = 0.43$ for the old group, $d = 0.35$ for the group born 1925–1944, and $d = 0.28$ for those born 1945 or later. Computed for the 1991–1999 period of decline visible in Figure 3, for the old group $d = 0.44$, for the 1925–1944 born $d = 0.27$ and for the younger cohorts $d = 0.21$. For the total sample, the 1984–1999 and 1991–1999 changes yield $d = 0.32$ and $d = 0.26$ respectively. Thus, compared to the changes in the younger groups, effect sizes of the old group show more pronounced decline in mean levels across the whole period analysed and in particular over the decade of the nineties. In sum, the total changes observed are at most "small" in terms of the effect size framework. It should be noted, that these effect size values are limited by pairwise deletion of cases with incomplete data, i.e. only respondents with measured LS values at both panels waves are included in the computations of the 1984–1999 and 1991–1999 effect sizes.

QMS Analyses

Table II shows the FIML estimation model fit parameters for QMS versions M1 to M5. It should be noted that all chi-square tests of overall model fit were significant with $p \leq 0.001$ for groups ALL and OLD. Regarding the incremental fit indexes and RMSEA values, versions M1, M3, M4, and M5 could be accepted as good fitting models in both groups.

Following the decision rules to compare covariance structure versions M1 to M4 (see Figure 2) leads to acceptance of M4 for both groups. This was true even regarding other model versions mentioned in the methods section above (perfect parallel stability, perfect monotonous stability with invariant pattern of change, invariance of true score variances only). Thus, a second order stationary model with true score variances and stability coefficients not changing over the 16 panel waves fits the data well.

Comparing M1 and M5, for all respondents as well as for the old respondents M5 has to be rejected due to an increase in RMSEA ≥ 0.01 and non-overlapping RMSEA 90% confidence intervals. Thus, a model of stable means does not hold, indicating that even in the old group there is an amount of chance in means that is not negligible.

For unconstrained QMS version M1 and the finally accepted version M4, estimates of stability coefficients and true score variances

TABLE II

Model fit of quasi-Markov simplex models

	M1		M2		M3		M4		M5	
	ALL	OLD	ALL	OLD	ALL	OLD	ALL	OLD	ALL	OLD
df	91		104		117		118		106	
$y\chi^2$	344	**173**	5801	**850**	531	**235**	533	**254**	1377	**459**
TLI	0.999	**0.997**	0.979	**0.979**	0.999	**0.997**	0.999	**0.997**	0.995	**0.990**
NFI	0.999	**0.996**	0.984	**0.982**	0.999	**0.995**	0.999	**0.994**	0.996	**0.990**
IFI	0.999	**0.998**	0.984	**0.984**	0.999	**0.997**	0.999	**0.997**	0.996	**0.992**
CFI	0.999	**0.998**	0.984	**0.984**	0.999	**0.997**	0.999	**0.997**	0.996	**0.992**
RMSEA	0.014	**0.021**	0.063	**0.059**	0.016	**0.022**	0.016	**0.024**	0.029	**0.040**
–L90	0.013	**0.016**	0.061	**0.055**	0.015	**0.018**	0.014	**0.020**	0.028	**0.036**
–U90	0.016	**0.025**	0.064	**0.062**	0.017	**0.026**	0.017	**0.027**	0.031	**0.044**

Note: df = degrees of freedom; –L90/–U90 = lower/upper limit of RMSEA 90% confidence interval.

are shown in Table III. Most notably, in both groups the estimated first order autocorrelations between true scores of subsequent panel waves appear to be near the theoretical limit of 1, indicating monotonic true score stability high above the stability of the observed scores (cf. correlations shown in Table I). Compared to the high stability in the total sample, estimated stability within the group of old respondents is not elevated to a mentionable degree. Estimated true score variances within the old group appear to be higher than within the total sample. In the M4 results, the increase in ‘‘old variance’’ amounts to about 60% of the variance for all respondents. Regarding change in stabilities or true score variances, in both groups no tendencies of systematic changes over the panel waves are visible in the M1 results, confirming acceptance of M4 with one stability and one variance value to characterize change over the 15 one-year intervals.

Regarding the QMS measurement model (see Equation 1), single-item measures reliabilities can be computed from ‘‘direct’’ parameter estimates. As the focus of this article is not on reliability, these are not listed in detail. Based on the M4 solution, reliability estimates range between 0.55 and 0.68 in the old group and between 0.45 and 0.63 for all respondents. Total sample reliabilities follow the tendency to grow from a low initial level over the early waves, as already reported and related to a learning effect (Landua, 1993; Schräpler, 1995): Values

TABLE III

M1 and M4 results: stability coefficients and true score variances

	ALL				OLD			
	ρ_{M1}	ρ_{M4}	σ^2_{M1}	σ^2_{M4}	ρ_{M1}	ρ_{M4}	σ^2_{M1}	σ^2_{M4}
1984	–	–	n.i.	1.98	–	–	n.i.	3.18
1985	n.i.	0.93	2.25	⋮	n.i.	0.94	3.07	⋮
1986	0.88	⋮	1.94	⋮	0.93	⋮	2.56	⋮
1987	0.89	⋮	1.97	⋮	0.88	⋮	2.79	⋮
1988	0.94	⋮	2.15	⋮	1.00	⋮	3.34	⋮
1989	0.93	⋮	2.21	⋮	0.97	⋮	3.18	⋮
1990	0.92	⋮	1.92	⋮	0.91	⋮	3.30	⋮
1991	0.93	⋮	1.68	⋮	0.95	⋮	2.93	⋮
1992	0.95	⋮	1.72	⋮	0.96	⋮	2.85	⋮
1993	0.94	⋮	2.02	⋮	0.96	⋮	3.55	⋮
1994	0.93	⋮	1.94	⋮	0.94	⋮	3.25	⋮
1995	0.95	⋮	2.04	⋮	0.94	⋮	3.14	⋮
1996	0.94	⋮	1.97	⋮	0.99	⋮	3.52	⋮
1997	0.95	⋮	2.15	⋮	0.94	⋮	2.90	⋮
1998	0.93	⋮	2.01	⋮	0.97	⋮	3.52	⋮
1999	n.i.	↓	n.i.	↓	n.i.	↓	n.i.	↓

Note: ρ_{M1}/ρ_{M4} = M1/M4 estimates of coefficients of stability (true score correlation with previous year); $\sigma^2_{M1}/\sigma^2_{M4}$ = M1/M4 estimates of true score variances; n.i. = not identified.

start with minimum value 0.45 in 1984 and reach the 0.6-level in 1990, with all 1984–1989 reliabilities <0.6 and all 1990 to 1999 reliabilities ≥ 0.6. Old group reliabilities follow the same, but less articulate, pattern of early increase, again starting with lowest reliability 0.55 in 1984.

As could be expected in view of the statistical procedure, the FIML estimates of the means differ marginally from the sample means listed in Table I. Thus, estimated means are not shown in detail.

DISCUSSION

With regard to the hypotheses on development of LS in old age proposed in the introductory section, discussion of results will focus on four aspects: (1) Overall, lifelong monotonic stability of LS; (2)

development of true score variances in older age; (3) development of monotonic stability in old age; and (4) development of mean levels in old age.

Overall Monotonic Stability of LS

Results reported in the previous section indicate high monotonic stability of LS within one-year time intervals through the adult life span. Given the stability coefficients reported, over 85% of "true variance" of satisfaction at some point in time could be explained from the previous measurement's satisfaction. Thus, subjects revealed strong tendencies to keep their relative "position" in satisfaction. This conclusion holds for the total sample analysed, including all respondents aged 18 years or older. Stability found in the "old" subsample of individuals born 1924 or earlier was not estimated higher than those in the total sample. Thus, high monotonic stability seems to be the fundamental developmental characteristic of LS over all ages.

Most notably, high stability coefficients have been obtained by modelling true score stability rather than using the observed satisfaction autocorrelations as indicators of monotonic stability. The latter would involve the conclusion of medium stability and a large amount of "exogenous" change in relative satisfaction values. Assuming the QMS as model for longitudinal development of satisfaction, most of this change is attributed to measurement error and not to true change of subjects' LS. Thus, the results undermine the importance of modelling true score change in research on stability of SWB.

Ehrhardt et al. (2000), analysing the 1984–1994 LS data of the West German SOEP sample, reported an unstandardized true score regression coefficient of 0.929 for what they called "shifting equilibrium model", i.e. the QMS with regression coefficients β_t constrained equal across subsequent time intervals. Thus, they found the same regression coefficient as obtained by model M4 (the restriction of equal true score variances implies $\rho_{M4} = \beta_{M4}$). Schräpler (1995), analysing West German satisfaction data of the first seven SOEP waves and comparing different nested QMS model versions, reports stability coefficients very similar to those noted in Table III for M1 (which was the model version he accepted for LS). Apart from slight differences in model restrictions and the inclusion of five/six more SOEP waves, the QMS analyses presented here differ from those reported by Ehrhardt et al. and Schräpler in missing value treatment, as those did not use the FIML estimator and applied listwise deletion of cases with missing values. Thus, the findings of almost equal regression/stability coefficients with very similar or equal QMS versions could be seen as an indication that inclusion of longitudinal dropouts did not reduce estimates of overall stability. It seems that subjects who dropped out of the panel did not face major changes in LS over the waves preceding the dropout. Such changes could be expected if LS is affected by "terminal decline" related to distance to death, as has been evidenced for individuals' cognitive functioning and other behavioural domains (Berg, 1996).

However, in interpreting serial correlations showing "high" or "low" stability, the time span between measurements has to be considered. Given the simplex structure, higher order autocorrelations will be below the year-to-year true score correlation of 0.93 found for M4. For example, in M4 the "total path" from first true score T_{1i} to last true score T_{16i} would be 0.34 (which is the estimated 16-year correlation, since true score variances are constrained equal). Thus, subjects did not reveal a strong tendency to keep their relative position over longer time periods, as one and a half decade of years.

High monotonic stability over one-year intervals may be seen as an indicator for high "habituality" of LS. Within a year, only a few subjects in the sample changed their relative position in satisfaction. That may be because changes in LS are short-term phenomenons and quickly adapted to the initial level and/or because within a year only few changes occur without adaptation. Thus, the high one-year stability coefficient found here can be seen as an evidence for a strong tendency for a habitual level of satisfaction that is not changed easily.

On the other hand, followed over 16 years few subjects that changed this habitual level of LS within each year summed up to a larger amount, so that over a longer period fewer remain to keep it stable. This latter conclusion, which is in line with Ehrhard et al. (2000), may be seen as conflicting with theoretical conceptions suggesting trait-like properties of LS, and supports Veenhoven's (1994) claim that happiness is not a trait. LS as it appears in the analyses presented here cannot be characterized as trait-like in the sense that most individuals keep habitual satisfaction throughout the lifespan, but results indicate that individuals do not change it frequently.

Development of True Score Variances in Old Age

Acceptance of M4, with only one true score variance invariant over time, indicates that true score variances did not change over time to an important degree (reducing model fit of M4 compared to M1). Concerning the ‘‘winners and losers’’ hypothesis claimed above, increasing variances could be expected in the old group. Thus, the good fit of model versions with variances constrained equal in the old age group could be interpreted as evidence against this hypothesis. On the other hand, estimated M4 true score variance in the old subsample was notedly increased compared to the variance in the total sample, which is in line with the assumption of ‘‘winners and losers’’ in old age. Altogether, evidence for growing variance of LS in old age seems ambiguous.

Interpreting this ambiguity, the definition of the old group should be taken under consideration: Including a broad range of birth cohorts (≤ 1924) could reduce age-related change in true score variances if the age-variance relation is nonlinear over the 60+ age range. If, for example, variance grows over early old age (say 60–70) and reaches a stable level for the ‘‘oldest old’’, the presence (and growing proportion) of the latter would reduce and slow down change in variances. Thus, a minor age-related increase in variances over some part of the old age life span might be not detected within the group of respondents born 1924 or earlier. Additional QMS analyses not presented here, based upon ‘‘closer’’ four-year birth cohort groups, also did not reveal evidence for changing variances in older age (Schilling, 2003). Nevertheless, substantial age-related increase in variances, at least over a part of the older age life span, should have caused rejection of model version M4 and should be visible in M1 results, which do not show any clear trend of increasing or declining variances.

On the other hand, elevation of the elders’ M4 true score variance compared to the total sample variance could have been caused by a systematic age effect on LS: If LS does decline (or increase) in older age, then the variance of a 60+ age group at some point in time should be higher than those of a sample covering all adult ages, because the 60+ sample contains a higher proportion of oldest old persons with decreased (or increased) satisfaction. To get some deeper insight into this matter, Figure 3 has been presented: For the old group compared to younger birth cohorts, a steeper downward tendency in mean levels is present over the second half of the time period analysed, suggesting an age-related decline of LS in "older old" age (see discussion below). Thus, rather than the development of "winners and losers" in LS, an age-related decline accelerating over the old age period may at least partly have increased the true score variance within the old subgroup.

Altogether, evidence for increasing variance in LS in older age seems sparse: Variances of the elders did not increase longitudinally and the high variance in the old group may not have been caused by "winners" *and* "losers", but only by "losers" sampled over different sections of the losing track.

Development of Monotonic Stability in Old Age

Regarding longitudinal development of stability coefficients, acceptance of model version M4 with only one coefficient invariant across measurement intervals indicates that one-year stability did not change to a substantial amount (that would reduce model fit of M4 compared to M1). This especially holds for the subsample of older respondents: No substantial age-related increase in monotonic stability, causing rejection of model version M4, was found, nor is it visible in M1 stability coefficients.

Considering the same argument as explained in the discussion of the true score variances, the broad range of birth cohorts ($\leq$1924) included in the old group could also reduce age-related change in stability coefficients if the age-stability relation is nonlinear over the 60+ age range. But additional QMS analyses not presented here, based upon "closer" four-year birth cohort groups, did not reveal evidence for changing stabilities in older age (Schilling, 2003).

Comparing the M4 stability coefficient of the old group to that of the total sample, it appears negligibly elevated. Thus, no evidence supporting the assumption of increased stability of LS in older age was found. Evidence for decreasing stability in older age, supporting the "winners and losers" assumption, was not found as well.

Development of Mean Levels in Old Age

Rejection of M5 may be taken as evidence that changes in means found in the old group are not negligible, for modelling invariance of means leads to substantial reduction of the model fit. Notably, this finding holds for the total sample as well.

Regarding the trends in mean level changes as depicted in Figure 3, it seems that these can be characterized by three phases: First, an initial decline from 1984 to 1988, followed by a second phase of increase from 1988 to 1991, and ending in a third phase of decline from 1991 to 1999. There are plausible interpretations for this trends. The decline in means may reflect an effect of repeated measurement which causes respondents, starting near the upper end of the satisfaction scale, to shift their responses towards the center of the scale. Also, it may be seen as reflecting some real societal phenomenon of growing dissatisfaction, i.e. a period effect is at work. Considering the short phase of increase visible in Figure 3, it seems remarkable that the historic process of German reunification took part within this time period and may have affected subjective evaluations of life in a positive way. However, these interpretations go far beyond the topics of the present work and deserve further investigation on their own.

Compared to the means of the younger birth cohort groups, a more pronounced decline is visible for the old group across the 1991–1999 period, which may be seen as evidence for a decline in LS especially related with older age. Keeping in mind the aging of the old group (mean age of the old group was 70.1 in 1984, 74.8 in 1991, and 80.3 in 1999), such a substantial age related decline seems to happen not across the whole old age period, but only in the older old (say 70+) ages.

It should also be noted, on the other hand, that fit indices reported for M5 show good fit as well: Constraining *no* mean level changes does not produce dramatic misfit of the model, as it would be the case if the mean changes were larger in amount. As mentioned in the results section, effect sizes are below the "small" level for all one-year changes observed and are "small" for the total change across the whole period analysed. Thus, it must be concluded that the mean level changes discussed here should be interpreted as minor, showing only slight shifts in overall LS across the years.

CONCLUSION

Concerning the development of LS in old age, it seems that there is nothing special at all. In terms of monotonic stability, LS seems to be highly stable across the adult lifespan, i.e. in younger as well as in older ages. This means that individuals of all ages exhibit a strong tendency to maintain their relative level of LS. No evidence was found that this tendency increases in old age. The assumption, that growing old produces "winners and losers" in LS, i.e. that some may improve their SWB as others will decline, was not strongly supported by the results discussed above, as the greater variance in elder's satisfaction may have been caused by some slight age-related decline in mean satisfaction levels over the 60+ age range.

There are some limitations to the results presented. As this study was aimed at analysing overall stability of LS in old age, no attempt was made to consider variables potentially influencing LS, such as gender, educational level, or health status. Relating stability of LS to such variables may be an interesting topic of further research. Also, it could be asked to what extend results would vary under changing cultural or societal conditions. Results found for a West German sample across the 1984–1999 period may not hold in other populations or times surveyed. Finally, attrition appears as serious problem in almost every longitudinal study, causing the risk that results are biased due to selectivity of the dropout. In the present study, the problem was eased by choice of statistical methodology which allows existing data from dropped out cases to be included into analysis. However, it cannot be avoided in total.

Finally, the findings suggesting an age-related decline in mean levels of measured LS values challenge the assumption of the "paradox" of LS in old age. Gerontological discussion, focussed on stability of mean levels found in cross-sectional age group comparisons, must take into consideration that this stability may not hold in longitudinal studies. In cross-sectional designs, age-related decline may be overlayed and hidden by cohort effects, thus, further research is needed to disentangle and estimate the effects of age and cohort on LS (Schilling, 2003). Notably, the age-related decline seemed to accelerate in older old ages. Thus, the "paradox" of SWB in old age may hold only in early old ages (see also Kunzmann et al., 2000), but more research is needed concerning age-related decline in LS over the 70+ ages.

REFERENCES

Arbuckle, J. L.: 1996, 'Full information estimation in the presence of incomplete data', in G. A. Marcoulides and R. E. Schumacker (eds.), Advanced Statistical Modeling (Erlbaum, Mahwah, NJ), pp. 243–277.

Arbuckle, J. L. and W. Wothke: 1999, AMOS 4.0 User's Guide (SmallWaters, Chicago).

Baltes, P. B. and M. M. Baltes: 1990, 'Psychological perspectives on successful aging: The model of selective optimization with compensation', in P. B. Baltes and M. M. Baltes (eds.), Successful Aging: Perspectives From the Behavioral Sciences (Cambridge University Press, Cambridge), pp. 1–34.

Baltes, M. M. and L. L. Carstensen: 1996, 'The process of successful aging', Ageing and Society 16, pp. 397–422.

Baltes, P. B. and K. U. Mayer: 1999, The Berlin Aging Study: Aging from 70 to 100 (Cambridge University Press, New York).

Baur, P. A. and M. A. Okun: 1983, 'Stability of life satisfaction in late life', The Gerontologist 23, pp. 261–265.

Berg, S.: 1996, 'Aging, behavior, and terminal decline', in J. E. Birren and K. W. Schaie (eds.), Handbook of the Psychology of Aging, 5th edn. (Academic Press, San Diego, CA), pp. 323–337.

Bollen, K. A.: 1989, Structural Equations with Latent Variables (Wiley, New York).

Brandtstädter, J.: 1999, 'Sources of resilience in the aging self', in T. M. Hess and F. Blanchard–Fields (eds.), Social Cognition and Aging (Academic Press, San Diego, CA), pp. 123–141.

Brandtstädter, J. and W. Greve: 1994, 'The aging self: Stabilizing and protective processes', Developmental Review 14, pp. 52–80.

Brandtstädter, J., D. Wentura and U. Schmitz: 1997, 'Veränderungen der Zeit-und Zukunftsperspektive im Übergang zum höheren Alter: quer- und längsschnittliche Befunde' (Age-related changes in future time perspectives: Cross-sectional and longitudinal findings), Zeitschrift für Psychologie 205, pp. 377–395.

Browne, M. W.: 1984 'Asymptotically distribution-free methods for the analysis of covariance structures', British Journal of Mathematical and Statistical Psychology 37, pp. 62–83.

Browne, M. W. and R. Cudeck: 1993, 'Alternative ways of assessing model fit', in K. A. Bollen and J. S. Long (eds.), Testing Structural Equation Models (Sage, Newbury Park, CA), pp. 136–162.

Chen, C.: 2001, 'Aging and life satisfaction', Social Indicators Research 54, pp. 57–79.

Cohen, J.: 1988, Statistical Power Analysis for the Behavioral Sciences (Lawrence Erlbaum, Hillsdale).

Coni, N., W. Davison and S. Webster: 1992, Ageing: The Facts (Oxford University Press, Oxford).

Costa, P. T., R. R. McCrae and D. Arenberg: 1983, 'Recent longitudinal research on personality and aging', in K. W. Schaie (ed.), Longitudinal Studies of Adult Psychological Development (Guilford Press, New York), pp. 222–265.

Costa, P. T. and R. R. McCrae: 1984, 'Personality is a lifelong determinant of wellbeing', in C. Z. Malatesta and C. E. Izard (eds.), Emotion in Adult Development (Sage, Beverly Hills), pp. 141–157.

DeNeve, K. M. and H. Cooper: 1998, 'The happy personality: A meta-analysis of 137 personality traits and subjective well–being', Psychological Bulletin 124, pp. 197–229.

Diener, E. and E. Suh: 1997, 'Age and subjective well–being: An international analysis', Annual Review of Gerontology and Geriatrics 17, pp. 304–324.

Diener, E., S. Oishi and R. E. Lucas: 2003, 'Personality, culture, and subjective wellbeing: Emotional and cognitive evaluations of life', Annual Review of Psychology 54, pp. 403–425.

Diener, E., E. M. Suh, R. E. Lucas and H. L. Smith: 1999, 'Subjective well-being: Three decades of progress', Psychological Bulletin 125, pp. 276–302.

Diggle, P. J., P. Heagerty, K.-Y. Liang and S. L. Zeger: 2002, Analysis of Longitudinal Data (Oxford University Press, Oxford).

Dunlap, W. P., J. M. Cortina, J. B. Vaslow and M. J. Burke:1996, 'Meta-analysis of experiments with matched groups or repeated measures designs', Psychological Methods 1, pp. 170–177.

Ehrhardt, J. J., W. E. Saris and R. Veenhoven: 2000, 'Stability of life-satisfaction over time', Journal of Happiness Studies 1, pp. 177–205.

Ehrlich, B. S. and D. M. Isaacowitz: 2002, 'Does subjective well-being increase with age?', Perspectives in Psychology 5, pp. 20–26 (Internet document, 20.01.2004: http://dolphin.upenn.edu/ upsych/journal.html).

Enders, C. K.: 2001a, 'The impact of nonnormality on full information maximum likelihood estimation for structural equation models with missing data', Psychological Methods 6, pp. 352–370.

Enders, C. K.: 2001b, 'A primer on maximum likelihood algorithms available for use with missing data', Structural Equation Modeling 8, pp. 128–141.

Enders, C. K. and D. L. Bandalos: 2001, 'The relative performance of full information maximum likelihood estimation for missing data in structural equation models', Structural Equation Modeling 8, pp. 430–457.

Fan, X., B. Thompson and L. Wang: 1999, 'Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes', Structural Equation Modeling 6, pp. 56–83.

Ferring, D. and S.-H. Filipp: 1997, 'Subjektives Wohlbefinden im Alter: Struktur- und Stabilitätsanalysen', (Subjective well-being in old age: Analyses of stability and structure), Psychologische Beiträge 39, pp. 236–258.

Grombach, H.: 1976, 'Consistency and change of personality variables in late life', in H. Thomae (ed.), Patterns of Aging (Karger, Basel), pp. 61–67.

Haisken-DeNew, J. P. and J. R. Frick: 2003, 'Desktop companion to the German Socio-Economic Panel Study (GSOEP)', version 7.0, Berlin: German Institute for Economic Research (Internet document, 20.08.2004: http://www.diw.de/english/sop/service/dtc/dtc.pdf).

Heise, D. R.: 1969, 'Separating reliability and stability in test-retest correlation', American Sociological Review 34, pp. 93–101.

Hershberger, S. L., P. C. M. Molenaar and S. E. Corneal: 1996, 'A hierarchy of univariate and multivariate structural time series models', in G. A. Marcoulides and R. E. Schumacker (eds.), Advanced Statistical Modeling (Lawrence Erlbaum, Mahwah, NJ), pp. 159–194.

Herzog, A. R. and W. L. Rodgers: 1981, 'Age and satisfaction: Data from several large surveys', Research on Aging 3, pp. 142–165.

Horley, J. and J. J. Lavery: 1995, 'Subjective well-being and age', Social Indicators Research 34, pp. 275–282.

Hoyle, R. H. and A. T. Panter: 1995, 'Writing about structural equation models', in R. H. Hoyle (ed.), Structural Equation Modeling (Sage, Thousand Oaks, CA), pp. 158–176.

Hu, L. and P. M. Bentler: 1995, 'Evaluating model fit', in R. H. Hoyle (ed.), Structural Equation Modeling (Sage, Thousand Oaks, CA), pp. 76–99.

Hu, L. and P. M. Bentler: 1998, 'Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification', Psychological Methods 3, pp. 424–453.

Jöreskog, K. G.: 1970, 'Estimation and testing of simplex models', British Journal of Mathematical and Statistical Psychology 23, pp. 121–145.

Kolberg, K. J. S.: 1999, 'Biological development and health risk', in T. L. Whitman, T. V. Merluzzi and R. D. White (eds.), Life-Span Perspectives on Health and Illness (Erlbaum, Mahwah, NJ), pp. 23–45.

Kozma, A. and M. J. Stones: 1983, 'Predictors of happiness', Journal of Gerontology 38, pp. 626–628.

Kunzmann, U., T. D. Little and J. Smith: 2000, 'Is age-related stability of subjective well–being a paradox? Cross–sectional and longitudinal evidence from the Berlin Aging Study', Psychology and Aging 15, pp. 511–526.

Landua, D.: 1993, 'Veränderungen von Zufriedenheitsangaben in Panelbefragungen: Eine Analyse über nicht beabsichtigte Effekte der Längs-schnitt designs' (Changes in satisfaction reports in panel surveys: An analysis of non–intentional effects of longitudinal designs), Kölner Zeit-schrift für Soziologie und Sozialpsychologie 45, pp. 553–571.

Larson, R.: 1978, 'Thirty years of research on the subjective well–being of older Americans', Journal of Gerontology 33, pp. 109–125.

Lee, S.-Y., W.-Y. Poon and P. M. Bentler: 1995, 'A two–stage estimation of structural equation models with continuous and polytomous variables', British Journal of Mathematical and Statistical Psychology 48, pp. 339–358.

Manton, K. G.: 1990. 'Mortality and morbidity', in R. H. Binstock and L. K. George (eds.), Handbook of Aging and the Social Sciences, 3rd ed. (Academic Press, San Diego, CA), pp. 64–89.

Mardia, K. V.: 1985, 'Mardia's test of multinormality', in S. Kotz and N. L. Johnson (eds.), Encyclopedia of Statistical Sciences, Vol. 5 (Wiley, New York), pp. 217–221.

Marsh, H. W., J. R. Balla and K.-T. Hau: 1996, 'An evaluation of incremental fit indices: A clarification of mathematical and empirical properties', in G. A. Marcoulides and R. E. Schumacker (eds.), Advanced Statistical Modeling (Erlbaum, Mahwah, NJ), pp. 315–353.

McNeil, J. K., M. J. Stones and A. Kozma: 1986, 'Longitudinal variation in domain indicators of happiness', Social Indicators Research 18, pp. 119–124.

Michalos, A. C.: 1986, 'An application of multiple discrepancies theory (MDT) to seniors', Social Indicators Research 18, pp. 349–373.

Mroczek, D. K. and C. M. Kolarz: 1998, 'The effect of age on positive and negative affect: A developmental perspective on happiness', Journal of Personality and Social Psychology 75, pp. 1333–1349.

Palmore, E. and V. Kivett: 1985, 'Change in life satisfaction', in E. Palmore, E. W. Busse, G. L. Maddox, J. B. Nowlin and I. C. Siegler (eds.), Normal Aging III (Duke University Press, Durham), pp. 373–380.

Raykov, T. and K. F. Widaman: 1995, 'Issues in applied structural equation modeling research', Structural Equation Modeling 2, pp. 289–318.

Rendtel, U.: 1993, 'Über die Repräsentativität von Panelstichproben. Eine Analyse der feldbedingten Ausfälle im Sozio-oekonomischen Panel (SOEP)' (On the representativeness of panel studies. An analysis of dropouts in the Socio-Economic Panel, SOEP), German Institute for Economic Research, Berlin, Discussion Paper 70.

Rindskopf, D.: 1984, 'Using phantom and imaginary latent variables to parameterize constraints in linear structural equation models', Psychometrika 49, pp. 37–47.

Rubin, D. B.: 1976 'Inference and missing data', Biometrika 63, pp. 581–592.

Rudinger, G. and C. Rietz: 1998, 'The neglected time dimension? Introducing a longitudinal model testing latent growth curves, stability, and reliability as time bound processes', Methods of Psychological Research Online 3, pp. 109–130 (Internet document, 20.01.2004: http://www.mpr-omline.de).

SAS Institute Inc.: 1990, SAS Procedures Guide, Version 6. (SAS Insitute Inc., Cary, NC).

Schafer, J. L. and J. W. Graham: 2002, 'Missing data: Our view of the state of the art', Psychological Methods 7, pp. 147–177.

Schaie, K. W. and S. M. Hofer: 2001, 'Longitudinal studies in aging research', in J. E. Birren and K. W. Schaie (eds.), Handbook of the Psychology of Aging, 5th edition (Academic Press, San Diego, CA), pp. 53–77.

Schilling, O.: 2003, 'Längsschnittliche Analysen zur Entwicklung der Zufriedenheit im höheren Lebensalter' (Longitudinal analyses on the development of satisfaction in older age), doctoral dissertation, University of Heidelberg, Faculty for Behavioral and Cultural Studies.

Schräpler, J. P.: 1995 Die Schätzung von Reliabilität und Stabilität der Zufriedenheitsangaben im Sozio-oekonomischen Panel (The Estimation of Reliability and Stability of Satisfaction Reports in the Socio-Economic Panel), Ruhr University Bochum, Faculty for the Social Sciences, Discussion Paper 95–13.

Schwarz, N. and F. Strack: 1991, 'Evaluating one's life: A judgement model of subjective well–being', in F. Strack, M. Argyle and N. Schwarz (eds.), Subjective Well–being (Pergamon Press, Oxford), pp. 27–47.

Shmotkin, D.: 1991, 'The role of time orientation in life satisfaction across the life span', Journal of Gerontology: Psychological Sciences 46, pp. P243–P250.

Smith, J., W. Fleeson, B. Geiselmann, R. Settersten and U. Kunzmann: 1996, 'Sources of well-being in very old age', in P. B. Baltes and K. U. Mayer (eds.), The Berlin Aging Study: Aging from 70 to 100 (Cambridge University Press, New York), pp. 450–471.

Smith, J. and A. M. Freund: 2002, 'The dynamics of possible selves in old age', Journal of Gerontology: Psychological Sciences 57B, pp. P492–P500.

SOEP Group: 2001, 'The German Socio-Economic Panel (GSOEP) after more than 15 years – overview', Vierteljahrshefte zur Wirtschaftsforschung 70, pp. 7–14.

Staudinger, U. M.: 1999, 'Social cognition and a psychological approach to an art of life', in T. M. Hess and F. Blanchard-Fields (eds.), Social Cognition and Aging (Academic Press, San Diego, CA) pp. 343–375.

Staudinger, U. M.: 2000, 'Viele Gründe sprechen dagegen, und trotzdem geht es vielen Menschen gut: Das Paradox des subjektiven Wohlbefindens' (Many reasons speak against it, yet many people feel good: The paradox of subjective well-being) Psychologische Rundschau 51, pp. 185–197.

Steiger, J. H.: 2000, 'Point estimation, hypothesis testing, and intervall estimation using the RMSEA: Some comments and a reply to Hayduk and Glaser', Structural Equation Modeling 7, pp. 149–162.

Tisak, J. and W. Meredith: 1990, 'Descriptive and associative developmental models', in A. von Eye (ed.), Statistical Methods in Longitudinal Research, Vol. II (Academic Press, San Diego, CA), pp. 387–406.

Veenhoven, R.: 1991, 'Is happiness relative?', Social Indicators Research 24, pp. 1–34.

Veenhoven, R.: 1994, 'Is happiness a trait?', Social Indicators Research 32, pp. 101–160.

Veenhoven, R.: 1996, 'Developments in satisfaction research', Social Indicators Research 37, pp. 1–46.

Werts, C. E., K. G. Jöreskog and R. L. Linn: 1971, 'Comment on "The estimation of measurement error in panel data"', American Sociological Review 36, pp. 110–113.

West, S. G., J. F. Finch and P. J. Curran: 1995, 'Structural equation models with nonnormal variables', in R. H. Hoyle (ed.), Structural Equation Modeling (Sage, Thousand Oaks, CA), pp. 56–75.

Wothke, W.: 2000, 'Longitudinal and multigroup modeling with missing data', in T. D. Little, K. U. Schnabel and J. Baumert (eds.), Modeling Longitudinal and Multilevel Data: Practical Issues, Applied Approaches, and Specific Examples (Erlbaum, Mahwah, NJ), pp. 219–240.

German Centre for Research on Ageing
at the University of Heidelberg
Bergheimer Str. 20
69115 Heidelberg
Germany
E-mail: schilling@dzfa.uni-heidelberg.de