Longitudinal Factor Analysis and Measurement Invariance of Sense of Coherence and General Self-Efficacy in Adolescence

Grevenstein, Dennis; Bluemke, Matthias

Longitudinal Factor Analysis and Measurement Invariance of Sense of Coherence and General Self-Efficacy in Adolescence

Dennis Grevenstein and Matthias Bluemke
Psychological Institute, University of Heidelberg, Germany

Abstract. Sense of coherence (SOC) and General Self-efficacy (GSE) are trait-like self-regulatory attributes, supposedly benefitting health. Previous data on their factorial validity and longitudinal stability in adolescent samples have been inconclusive. The present study examined the factor structure, measurement invariance (MI), and stability coefficients of SOC and GSE among German adolescents in a longitudinal design over the course of nine years from age 15 to age 24. Results supported the factorial validity of both scales. GSE parameters were invariant up to the level of strict invariance, whereas for SOC partial scalar and strict invariance were attainable after modifications. Here we document reliability, validity, and factor mean changes of the SOC and GSE scales from adolescence to young adulthood. Interindividual differences in SOC were moderately stable. Though this implies limited sensitivity to intraindividual developmental changes, it qualifies SOC for long-term predictions. GSE was conspicuously less stable, raising questions about its long-term criterion validity.

Keywords: sense of coherence (SOC), general self-efficacy (GSE), stability, longitudinal factor analysis, measurement invariance

Personality traits have frequently been used as predictors of health. Several trait-like constructs are specifically conceptualized to benefit health and buffer against adverse influences. Among the most prominent concepts are Sense of coherence (Antonovsky, 1987, 1998) and General Self-efficacy (Schwarzer & Jerusalem, 1995) as an extension Bandura's (1997) original, situation-specific Self-efficacy concept. Are these two sides of the same coin, or can one of the constructs be assessed better than the other one?

Sense of Coherence

Antonovsky's (1987) salutogenic theory provides a resource-oriented perspective to health. It explains how people keep or regain health. Health is not understood merely as a (binary) outcome; instead there is a fine-grained continuum between health and sickness. Sense of coherence (SOC) is at the heart of the theory and conceptualized as a general resistance resource (resilience, inner strength). As an "orientation-to-life," it buffers people's health in the face of adversities, critical life events, and distress. Three major factors are thought to underlie SOC: *comprehensibility* describes an individual's perception that situations and events are structured and clear; *manageability* denotes an individual's belief that one has the necessary skills to deal with life challenges; and *meaningfulness* represents an individual's belief that life demands and challenges are worthy of investment and engagement.

Plenty of evidence demonstrated associations between SOC and positive health outcomes: It has been linked to good mental health and health-related behavior (Eriksson & Lindström, 2006; Togari, Yamazaki, Takayama, Yamaki, & Nakayama, 2008), general psychological well-being (Nilsson, Leppert, Simonsson, & Starrin, 2010), depression (Haukkala et al., 2013), and anxiety (Moksnes, Espnes, & Haugan, 2013). Beyond health, higher SOC has also shown associations with work-related indices, such as higher perceived control over one's work, lower perceived job demands, and generally more job satisfaction (Holmberg, Thelin, & Stiernström, 2004).

The development of SOC is considered to be a dynamic process up to age 30. Adolescence is seen as a particularly sensitive phase (Rivera, García-Moya, Moreno, & Ramos, 2013). According to Antonovsky, SOC does not stabilize until the age of 30, and before that, is shaped by outside factors. Experiencing consistency (enhancing comprehensibility), load-balancing (enhancing manageability), and participation in decision-making (enhancing meaningfulness) are all supposed to foster SOC in childhood/adolescence (Antonovsky, 1987).

General Self-Efficacy

Self-efficacy (SE) constitutes a core aspect of Bandura's social cognitive theory (Bandura, 1977, 1997, 2001) and represents self-referent thoughts, convictions, and expectations of one's beliefs to overcome obstacles and succeed in a given situation. Accordingly, it too has been found a predictor of health (Egger, 2011; Schwarzer, 1992), and of academic and work-related performance (Robbins et al., 2004; Stajkovic & Luthans, 1998). According to Bandura, however, it is task- or situation-specific.

General or Generalized Self-efficacy (GSE) cuts across specific situations: An individual's general feeling across a broad range of challenging situations that require effort, skill, and perseverance (Schwarzer & Jerusalem, 1995; Tipton & Worthington, 1984). There has been considerable debate whether Bandura's SE concept, based on learning experiences, can be turned into a trait-like personality construct that is stable and consistent enough to predict actual outcomes. Task-specific SE may better account for variance in a task-specific situation (Bandura, 1997). Still, GSE is thought to have at least the potential to predict general outcomes in those situations and contexts where task-specific SE has not yet been developed. Indeed, like SOC, GSE has been related to health and performance outcomes (Andersson, Moore, Hensing, Krantz, & Staland-Nyman, 2014; Luszczynska, Gutiérrez-Doña, & Schwarzer, 2005).

Sense of Coherence (SOC) and General Self-Efficacy (GSE)

Both SOC and GSE cover aspects of self-regulation (Geyer, 1997); clearly SOC's manageability facet closely resembles that of GSE. Unsurprisingly, moderate to high correlations ($r = .45–.61$) between the two have been found (Posadzki, Stockl, Musonda, & Tsouroufli, 2010; Zirke, Schmid, Mazurek, Klapp, & Rauchfuss, 2007). Given their conceptual and empirical similarity, we intended to disambiguate their relative merits. Though both constructs explain health variance, the psychometric quality of the scales warrants deeper inspection. Specifically, when stability of interindividual differences – at young age in particular – and longitudinal predictions are at stake, before any conclusions about the constructs can be drawn, longitudinal measurement invariance needs to be established.

Reliability and Validity

Sense of Coherence (SOC)

The stability of interindividual differences has been documented before, both for adults (Feldt, Leskinen, Kinnunen, & Mauno, 2000; Feldt, Leskinen, Kinnunen, & Ruoppila, 2003) and adolescents (Honkinen et al., 2008). For adults, test-retest reliabilities of .78 over 1 year, .59–.67 over 5 years, and .54 over 10 years have been found (Eriksson & Lindström, 2005). Whereas internal consistency estimates have mostly been satisfying, findings for stability and transitional changes of SOC scale means have been mixed. At the aggregate level, differences between age groups have emerged (Rivera et al., 2013), yet Honkinen and colleagues (2008) showed only minor changes of mean SOC scores from age 15 to age 18 longitudinally.

With regard to the factorial structure of the SOC scale, findings are similarly mixed. Antonovsky (1993) stressed the holistic nature of SOC and advocated the use of a global SOC score. Contrasting his view, several studies replicated the theoretically derived three-factor structure of the SOC scale (Feldt et al., 2000, 2003; Stein, Lee, & Jones, 2006). Yet in most cases, the comprehensibility and manageability factors correlated so highly ($r > .90$) that one can question if they are truly conceptually distinct. Alternatively, a two-factor structure has been proposed with one factor spanning comprehensibility and manageability items, and with meaningfulness constituting the second factor (Zimprich, Allemand, & Hornung, 2006).

The previous studies are not without criticism though. In many cases, items composing the SOC scale had to be adjusted or dropped (Feldt et al., 2003; Stein et al., 2006), or the measurement model had to be modified to include correlated item residuals to achieve reasonable model fit (Feldt et al., 2000; Zimprich et al., 2006). Such modifications more often than not are carried out non-transparently, and they may have even been applied inconsistently across different points of measurement within the same population/publication. Standing in stark contrast, reasonable psychometric qualities have been attested to the SOC scale even in adolescent samples undergoing developmental transitions (Hagquist & Andrich, 2004). Taken together, the facts warrant further inspection of the suitability of SOC as a trait-like construct with a measurement model that actually holds across time.

General Self-Efficacy (GSE)

The GSE scale developed by Schwarzer and Jerusalem (1995), also used in the present research, was originally developed in German. It was subsequently adapted many times and is currently available in 31 languages (see http://userpage.fu-berlin.de/health/selfscal.htm). The scale has demonstrated reasonably good psychometric qualities, typically displaying internal consistencies ranging between .75 and .91. In the context of educational assessment, test-retest reliabilities over a year have amounted to .55 in a sample of adolescent students and .75 for adult teachers (Schwarzer & Jerusalem, 1999). With regard to the factor structure, GSE is considered an essentially unidimensional construct. Also, in a large-scale evaluation across 25 countries, a one-factor solution was supported (Scholz, Doña, Sud, & Schwarzer, 2002).

Despite these encouraging findings, GSE has been heavily criticized in the past (Chen, Gully, & Eden, 2001; Marsh & Grayson, 1994; Stajkovic & Luthans, 1998). Most criticism focused on reliability, construct validity, and factor structure of the GSE measures, though later studies supported the idea of reliable assessment, implying that at least the criticism of GSE on the basis of its reliability may have been overstated (Peter, Cieza, & Geyh, 2014; Scherbaum, Cohen-Charash, & Kern, 2006).

Measurement Invariance

Although construct reliability in a given sample at a given point in time is important, adopting a valid measurement model across different samples and measurement waves is equally crucial (Vandenberg & Lance, 2000). When applying a scale across situational contexts, cultures, or age groups with the intent to compare their scores numerically, most researchers simply assume that the scores reflect the identical construct. Yet, if there are reasons to believe that latent variable scores can change over time, one might alternatively suspect that the understanding of a construct (or items) differs. Stability indices and comparisons of mean scores across different groups – be they related to age, culture, or other grouping features – are only valid as far as measurement invariance can be established (Chen, 2008). One needs to ascertain that differences in scale means are due to true differences in latent means, not different item utilization. In other words, any observed group differences – and likewise relationships or relationship differences between a latent variable and external criteria – might be explained by disparate measurement models that hold for the measurement device.

Although measurement invariance (MI) for the SOC-13 scale has been addressed before (Hittner, 2007; Richardson, Ratner, & Zumbo, 2007; Zimprich et al., 2006), to date no study has investigated *longitudinal* measurement invariance (MI) during the period of adolescence. Even in adult samples, for whom SOC should be rather stable, merely *partial* invariance levels for *some* items could be established over the course of 5 years (Richardson et al., 2007), yet transitions through adolescence might affect the invariance of the whole measurement model. Similarly, the longitudinal invariance of GSE remains unilluminated, so a check is warranted if GSE can actually be considered a relatively stable trait, as implied by vast parts of the health literature.

Study Overview

We first examine the factorial validity of two major health-relevant constructs, SOC and GSE, sampled at age 15 and age 24. Satisfactory measurement models provided, we then test longitudinal MI of SOC and GSE among identical adolescents over a period of 9 years. Finally, assuming at least metric invariance, we report on the longitudinal stability of the latent variables. After general aspects of the data collection, we report the analyses separately for SOC and GSE.

Method

Study Sample

The data are part of a ten-year-longitudinal study on salutogenic factors and substance use (RISA) conducted in Germany from 2003 to 2012 (approved by the Ethics Committee of the University Hospital Heidelberg; #218/2005). The total study included 14 data collection events related to various health-related behaviors. Participants were 318 students (164 female; 51.6% and 154 male; 48.4%) with a median age of 14 at the beginning of the study. Almost two-thirds of the participants (n = 208) grew up in a traditional family setting, which was defined as living with both biological parents up to age 18. Level of education was almost equally distributed across the three-tier German school system.

While there was noticeable sample attrition (n = 134; 42.1%) over the course of 10 years, this participant dropout is comparable to other studies on adolescents' development (Honkinen et al., 2009). There was no *systematic* dropout with regard to the variables examined here. The present research analyzes data from T3 (age 15; n = 286) and T14 (age 24; n = 184). T3 represents the initial reference point, because participants could enter the study until T3. For stability and longitudinal MI analyses, we analyzed a subsample of n = 299 with data available at either age 15 or age 24 (another 19 individuals had completely missing data on the variables of interest).

Measures

Sense of Coherence

The German adaptation of Antonovsky's abbreviated 13-item Orientation to Life questionnaire (SOC-13) provides 5-point rating scales (most of the time ranging from 1 = *very rarely* to 5 = *very often*; Abel, Kohlmann, & Noack, 1995). To achieve comparability with an authoritative German version developed by Schumacher and colleagues later (Schumacher, Gunzelmann, & Brähler, 2000), SOC scores were rescaled to a 7-point rating scale format using a linear transformation. The scale includes four meaningfulness items (e.g., "Do you have the feeling that you don't really care about what goes on around you?"), five comprehensibility items (e.g., "Has it happened in the past that you were surprised by the behavior of people whom you thought you knew well?"), and four manageability items (e.g., "Has it happened that people whom you counted on disappointed you?"). For comparison purposes, the typical SOC sum scores were computed, provided that participants completed all items. Cronbach's alpha amounted to .87/.91 at age 15/24.

General Self-Efficacy

The GSE scale was developed by Schwarzer and Jerusalem (Hinz, Schumacher, Albani, Schmid, & Brähler, 2006; Schwarzer & Jerusalem, 1995). It comprises 10 items such as "If there are challenges, I can find a way to succeed." Answers were provided on 4-point scales (1 = *not true*, 2 = *rarely true*, 3 = *mostly true*, 4 = *completely true*). Typical GSE mean scores were computed (α = .86 and .89 at ages 15 and 24, respectively).

Statistical Analysis

Apart from *SPSS 21*, we used *Mplus* 7.11 (Muthén & Muthén, 1998–2012). The goodness-of-fit of the models was evaluated by (1) the – ideally nonsignificant – χ^2 test (Bentler & Bonett, 1980) and as low as possible a χ^2/df ratio, ideally as low as 2 (Tabachnick & Fidell, 2007); (2) the comparative fit index (CFI) with values of .90/.95 and above indicating appropriate/good model fit (Bentler, 1990; Hu & Bentler, 1999); (3) the root mean square error of approximation (RMSEA) with values of .00–.05/.06–.08/.09–.10 indicating good/reasonable/poor model fit (Browne & Cudeck, 1993); and (4) the standardized root mean square residual (SRMR) with values less than .08 (Hu & Bentler, 1999) or .05 (Schumacker & Lomax, 2010) considered to reflect good fit. The Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) were used as comparative fit indices. Lower scores indicate better model fit (Akaike, 1987), and differences greater than ±10 supply "strong evidence" against equal fit (Raftery, 1995). Robust Maximum Likelihood (MLR) was used for parameter estimation because GSE uses a coarse 4-point scale and multivariate normality did not hold in our sample (Small's omnibus test: $\chi^2(92) = 299.69$, $p < .001$). Full Information Maximum Likelihood (FIML) was used to handle missing data for the longitudinal CFAs (22.0% across all cells, mostly reflecting the sample attrition).

As the SOC-13 scale lacks essential tau-equivalence and strict unidimensionality, Cronbach's Alpha is an inappropriate indicator of reliability and likely to underestimate the true reliability of the construct (Bollen, 1989). Instead we provide the composite reliability (CR) and the average variance extracted (AVE) as SEM-based reliability estimates (Bacon, Sauer, & Young, 1995; Fornell & Larcker, 1981).

MI is tested within a sequential approach of nested, increasingly restricted confirmatory factor-analytical (CFA) models (Meredith, 1993; Vandenberg & Lance, 2000). For independent age groups, MI is approached by means of multiple-groups CFA, yet for longitudinal comparisons of the same group across age, MI has to be established within a longitudinal CFA framework (Brown, 2006; see Electronic Supplementary Material 1 for the specifics).

Irrespective of whether dependent or independent data are at hand, four increasingly restrictive forms of MI can be tested (Meredith, 1993; Schmitt & Kuljanin, 2008; Vandenberg & Lance, 2000):

(1) Configural MI indicates equal construct dimensionality and equivalent item-to-factor patterns across groups, though item loadings on a factor, intercepts, and residuals can differ.

(2) Under Metric MI, all loadings are constrained to be equal across groups, reflecting the same psychological meaning of the latent variable. Lack of metric MI invalidates longitudinal stability of interindividual differences.

(3) Scalar MI additionally assumes invariant item intercepts across groups. Observed scores are then based on the same unit of measurement, and the groups are equally calibrated. Lack of scalar MI prevents a meaningful interpretation of factor mean differences, in our case the inspection of intra-individual change over time.

(4) Strict MI further requires equality constraints for residuals, indicating equally reliable measures across groups. When strict MI holds across time, all differences on manifest variables are due to true differences on the latent variables, rather than random measurement error.

From this level onward, one can examine the invariance of structural parameters. For instance, latent means can be tested for differences and, if invariance exists, it may be legitimately attributed to developmental changes. Meredith (1993) argued that strict MI is desirable to infer latent mean differences on the basis of observed mean differences. However, lack of strict invariance merely confirms non-invariant residuals, which reflect disparate reliabilities; therefore, scalar MI is sufficient to compare latent means (Little, 1997). We anticipated that latent means would differ after 9 years.

If some parameters are non-invariant across groups, a weaker form of MI, partial invariance, may still hold. For instance, partial metric MI requires the majority, but not all, of the loadings to be invariant across groups. Even if only partial scalar MI holds, because a few intercepts are non-invariant, latent means can still be cautiously compared (Byrne, Shavelson, & Muthén, 1989; Lubke & Dolan, 2003).

Each MI model is compared to the previous one with fewer restrictions on the basis of χ^2-difference tests; Satorra-Bentler scaled χ^2-difference tests are used if scaled χ^2-difference values, which are not distributed as χ^2 themselves, are obtained from MLR (Satorra, 2000; Satorra & Bentler, 2001). As sample size increases, so does the test sensitivity to minor deviations from MI, yet relatively independent from sample size, fit indices inform us on the severity of model impairments after additional parameter restrictions. The adherence to strict cutoffs in examining ΔCFI and ΔRMSEA is increasingly discouraged (Fan & Sivo, 2009; Saris, Satorra, & van der Veld, 2009). Still, a drop in CFI less than or equal to .010 is conventionally deemed acceptable, as long as it is balanced by ΔRMSEA no greater than +.015 (Chen, 2007; Cheung & Rensvold, 2002). Finally, lower BIC values indicate a better trade-off between accuracy and parsimony.

Results

Descriptive Data Analysis

Mean SOC scores significantly increased during the study period, *M*s = 63.57 versus 67.33 (*SD*s = 11.50 vs. 11.66), $t(161) = 4.46$, $p < .001$, $d = 0.33$. The same was true for GSE, *M*s = 2.83 versus 3.09 (both *SD*s = 0.44), $t(166) = 6.41$, $p < .001$, $d = 0.59$. Manifest SOC scores yielded a significantly higher test-retest correlation than

Table 1. Zero-order correlations (below diagonal) and latent variable correlations (above diagonal) at ages 15 and 24 for SOC and GSE

	SOC_{15}	SOC_{24}	GSE_{15}	GSE_{24}
SOC_{15}	–	.58	.57	.30
SOC_{24}	.57	–	.36	.76
GSE_{15}	.47	.28	–	.29
GSE_{24}	.28	.66	.27	–

Notes. Correlations are significant at $p < .001$. Each latent correlation is exclusively based on the two involved variables, yet the final invariance models (with longitudinally correlated residuals) were adopted for same-construct relationships.

GSE, CI_{95} = .49–.64 versus .16–.37 ($Z = 4.98$, $p < .001$); GSE_{15} showed similarly strong associations with GSE_{24} as with SOC_{24} (see also Table 1).

Separate Confirmatory Factor Analyses at Ages 15 and 24

Sense of Coherence (SOC)

Upon testing six reflective measurement models for SOC-13 with CFA (more details can be found in the Electronic Supplementary Material, ESM 1), we accepted a two-factor model with one factor spanning across comprehensibility (#2, 6, 8, 9, 11) and manageability items (#3, 5, 10, 13), while meaningfulness (items #1, 4, 7, 12) formed the second factor. We incorporated the often correlated residuals between items #2 and #3, reflecting interpersonal trust (Feldt et al., 2003; Frenz, Carey, & Jorgensen, 1993; Gana & Garnier, 2001; Hittner, 2007; Richardson et al., 2007).
The model fitted the data well at ages 15 and 24, $\chi^2(63) = 79.53/91.08$, $p = .08/.01$, $\chi^2/df = 1.26/1.45$, RMSEA = .030 [CI_{90} = .000–.049]/.049 [CI_{90} = .024–.071], CFI = .982/.968, SRMR = .038/.045, CR = .88/.92, AVE = .38/.47.

General Self-Efficacy (GSE)

Essentially, GSE is conceived of as a unidimensional construct. Accordingly, a one-factor model explained the data reasonably well at both ages, $\chi^2(35) = 60.26/54.23$, $p < .01/.05$, $\chi^2/df = 1.72/1.55$, RMSEA = .050 [CI_{90} = .028–.071]/.055 [CI_{90} = .022–.083], CFI = .962/.966, SRMR = .041/.043, CR = .86/.89, AVE = .38/.46.

Longitudinal Factor Analyses and Measurement Invariance

Based on these adopted measurement models, we examined longitudinal MI of SOC and GSE (see OSM for details on the model specifications). To legitimately investigate longitudinal stability, at least (partial) metric invariance has to hold. To interpret longitudinal changes of factor means, at least (partial) scalar invariance is required.

Sense of Coherence (SOC)

The longitudinal factor analysis (Figure 1) comprised six latent variables. Meaningfulness and a combined comprehensibility/manageability factor loaded on the global SOC factor, repeatedly so for ages 15 and 24, with SOC_{15} predicting SOC_{24}. Each time, the residuals of items #2 and #3 were correlated. The dependent nature of the data

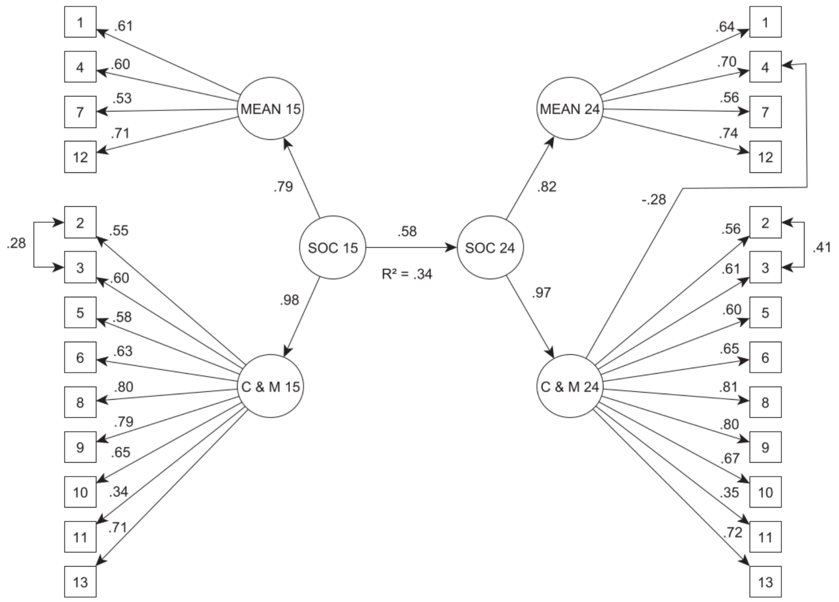

Figure 1. Longitudinal factor-analytical model of Sense of coherence (SOC-13) comprising factors meaningfulness (MEAN) and comprehensibility/manageability (C&M) at ages 15 and 24 with fully standardized path coefficients. Variance invariance was assumed (Model 5), yet longitudinal residual correlations are not displayed.

was accounted for by modeling item-specific residual covariances, that is, correlated residuals between corresponding items at different time points (Brown, 2006). The fact that the global SOC factor was split only in two sub-factors means that all MI testing had to be done on the two first-order factors.

MI testing (cf. Table 2) confirmed that configural MI held across age. For SOC_{24} the initial configural MI model yielded a negligible, nonsignificant negative residual variance ($\varepsilon = -0.003$), which we subsequently fixed to 0, gaining one *df*, which then further along became unnecessary in the metric MI model. When we tested metric MI by fixing factor loadings to be equal across age, model fit did not decline, supporting metric MI. We observed a noticeable drop when testing the equivalence of item intercepts, that is, scalar MI.

We gradually examined whether this lack of MI was caused by specific parameters. Relaxing one candidate parameter at a time on the basis of Modification Indices (ModInd; Byrne et al., 1989), we checked if model fit could be improved. Without strictly binding rules, this exploratory procedure is permissible if only a few parameters are adjusted and if they make sense. Researchers are advised to look for even more substantial improvements than ModInd close to 3.84 (statistically just significant at $p = .05$; Brown, 2006).

The fit of the scalar MI model improved noticeably after relaxing the intercept of item #11 ("When something happened, have you generally found that: . . . you overestimated or underestimated its importance vs. you saw things in the right proportions" (ModInd = 29.69, $\Delta\chi^2 = 31.93$). Apparently item #11 was easier to endorse at age 24. Hence, SOC enjoyed partial scalar invariance.

Within this context, we constrained the item residuals to equality for a test of strict MI, but observed poorer model fit. To achieve strict MI a cross-loading of item #4 at age 24 ("Until now your life has had: . . . clear goals") on the comprehensibility-manageability factor was suggested (ModInd = 10.26, $\Delta\chi^2 = 9.39$). Although this parameter is not a focal residual parameter, following Byrne and colleagues (1989), we checked if modifying the model in this manner would help or impoverish other parameters. The respective cross-loading ($\lambda_4 = -.28^{**}$) rather improved the factor loadings of all items on the meaningfulness factor. Apparently, SOC_{24} tended to be lower than SOC_{15} the more rigidly people entertained goals in life. This finding parallels the known negative influence of rigidity, specifically intolerance of uncertainty, on mental health (Ciarrochi, Said, & Deane, 2005). The strict MI model, adjusted for one cross-loading, showed a nonsignificant negative residual variance for SOC_{24} ($\varepsilon = -0.022$), which we fixed to 0 here (but which was positive in the next step).

After the four basic MI levels we examined the invariance of structural properties. First, we constrained all (co-)variances to be equal across age without significantly harming model fit. But when testing the equality of factor means, the fit of this full MI model decreased considerably, although the drop in fit might be well within acceptable limits. Once again, a nonsignificant negative residual variance for SOC_{24} had to be fixed to 0.

Table 2. Measurement invariance for SOC and GSE across ages 15 and 24 according to Longitudinal Confirmatory Factor Analysis (LCFA)

LCFA comparison	Equal loadings	Equal intercept	Equal residuals	Equal variances	Equal means	*df*	χ^2	Δdf	$\Delta\chi^2$	*p*	CFI	RMSEA [CI_{90}]	SRMR	AIC	BIC
SOC															
1 Configural invariance						280	345.98**	-	-	-	.967	.028 [.017-.037]	.051	14,445	14,804
2 Metric (weak) invariance	×					290	366.25**	10	21.30	.019	.961	.030 [.019-.039]	.064	14,446	14,768
3 Scalar (strong) invariance	×	×				301	430.85***	11	72.39	< .001	.934	.038 [.030-.046]	.077	14,493	14,775
3a Partial scalar invariance	×	×				300	398.92***	10	35.89	< .001	.950	.033 [.024-.042]	.072	14,460	14,745
4 Residual (strict) invariance	×	×	×			314	437.88***	14	38.50	< .001	.937	.036 [.028-.044]	.085	14,476	14,709
4a Partial strict invariance	×	×	×			313	428.49***	13	28.97	< .001	.942	.035 [.026-.043]	.080	14,468	14,705
5 Variance Invariance	×	×	×	×		**315**	**434.48*****	**2**	**5.39**	**.068**	**.940**	**.036 [.027-.044]**	**.083**	**14,471**	**14,700**
6 Full structural invariance	×	×	×	×	×	317	447.39***	2	11.76	.003	.934	.037 [.028-.044]	.096	14,482	14,704
GSE															
1 Configural invariance						166	278.83***	-	-	-	.920	.048 [.038-.057]	.076	7,817	8,054
2 Metric (weak) invariance	×					175	289.26***	9	10.07	.345	.919	.047 [.037-.056]	.087	7,810	8,013
3 Scalar (strong) invariance	×	×				184	297.67***	9	7.48	.587	.920	.045 [.036-.055]	.087	7,799	7,970
4 Residual (strict) invariance	×	×	×			187	309.91***	3	10.29	.016	.913	.047 [.037-.056]	.102	7,809	7,968
5 Variance Invariance	×	×	×	×		**188**	**308.20*****	**1**	**0.47**	**.494**	**.915**	**.046 [.037-.055]**	**.110**	**7,809**	**7,964**
6 Full structural invariance	×	×	×	×	×	189	338.72***	1	27.93	< .001	.894	.051 [.043-.060]	.147	7,842	7,993

Notes. Model fit: $^{**}p < .01$. $^{***}p < .001$. Accepted models printed in bold. χ^2 values are Satorra-Bentler scaled.

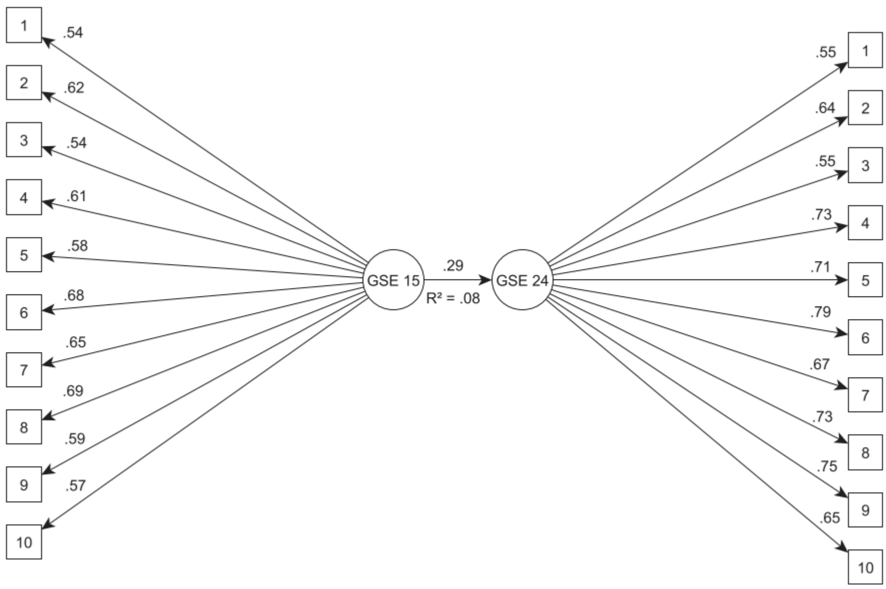

Figure 2. Longitudinal factor-analytical model of General Self-efficacy (GSE) at ages 15 and 24 with fully standardized path coefficients. Variance invariance was assumed (Model 5), yet longitudinal residual correlations are not shown.

We cautiously attribute the misfit to unequal latent means (ModInd = 11.78, $\Delta\chi^2 = 12.91$). Rather than claiming full MI, it is more likely that SOC scores increased as the outcome of a developmental trend. Consequently, unstandardized factor means increased significantly over time for comprehensibility and manageability (Estimate = 0.62, $SE = 0.17$, $p < .001$), yet the increase was not significant with regard to meaningfulness (Est. = 1.51, $SE = 2.83$, $p = .59$). This pattern is in fact in line with the findings presented by Zimprich and colleagues (2006) who reported latent mean differences exclusively for the combined comprehensibility and manageability factor.

The longitudinal stability of SOC across 9 years can be estimated on the basis of the (fitting) metric MI model, alternatively on the basis of the variance invariance model (Figure 1). The standardized regression weights for the longitudinal SOC path were .59 and .58. Hence, SOC_{15} explained about 34-35% of the variance of SOC_{24}.

General Self-Efficacy (GSE)

The longitudinal factor model (Figure 2) included two unidimensional latent variables, one for GSE_{15} and one for GSE_{24}. Furthermore, correlated residuals between corresponding items across time accounted for the nonindependence of data, and GSE_{15} predicted GSE_{24}. We appreciated configural, metric, scalar, and strict MI (Table 2). We acknowledge, though, that at the final level SRMR indicated poor model fit beyond conventionally accepted boundaries. Other indices indicated good fit as in previous steps.

Adopting strict MI, we tested the equality of structural parameters. Equal factor variances had almost no effect on model fit, again with the exception of SRMR. The assumption of equal factor means (full invariance), however, was not supported. A noticeable drop in model fit occurred, because factor means were not equal (ModInd = 28.31, $\Delta\chi^2 = 30.66$). An examination of factor mean differences indicated an unstandardized longitudinal increase of 0.57 points ($SE = 0.10$, $p < .001$) – about half a scale point on GSE's 4-point scale.

On the basis of metric MI and the variance invariance model, we estimated the stability of latent GSE over the course of nine years. The unattenuated correlation from GSE_{15} to GSE_{24} was $r = .33/.29$, so merely 8–11% of the variance were explained about a decade later.

Discussion

The present research investigated the factor structure, measurement invariance, and longitudinal stability of Sense of coherence (SOC) as measured by the SOC-13 scale, and General Self-efficacy (GSE) in adolescent participants across nine years – the maximum span available in our data, which provides the most stringent stability test. A single-factor model for GSE was adopted. For SOC a two-factor model (meaningfulness and manageability-comprehensibility) best reflected the data, closely replicating the findings by Zimprich and colleagues (2006). As found previously, model fit was better with correlated residuals between items #2 and #3 (Feldt et al., 2003, 2007; Hittner, 2007; Richardson et al., 2007).

Our findings have repercussions for psychological hypothesis testing and theory-building. To answer the question raised by previous researchers whether SOC can be compared across adolescent and adult samples (Hagquist & Andrich, 2004), our results indicate that this is the case. A qualified answer is that it may depend on the intended type of comparison. Whereas the GSE scale exhibited metric, scalar, and strict MI without any need for modification, the last unaltered level of invariance attained for SOC-13 was metric MI. Essentially the same meaning of both constructs was assessed at puberty and young adulthood.

Hence, one can legitimately examine and interpret covariance structures across age, for instance, in terms of stability of interindividual differences or in terms of (age-dependent changes of) correlations with other variables.

Regrettably, we could not establish scalar invariance for SOC unconditionally, as the identical respondents were calibrated differently for item #11 at age 24 than at age 15. The item was easier to endorse after having reached adulthood. The factor loading of item #11 was also noticeably low. In the context of the present study, it is impossible to find an explanation for this finding, yet one could speculate that the complex item wording might be a disadvantage. However, as this one non-invariant intercept could be identified, we can confidently announce partial scalar invariance. Strictly speaking, a numerical comparison of observed item scores and scale sum scores across age introduces a statistically non-negligible amount of bias. Conventionally, partial scalar MI is still considered "good enough" for an interpretation of mean structures if the number of non-invariant parameters is low (Byrne et al., 1989, Steenkamp & Baumgartner, 1998; see Steinmetz, 2013, for more critical view on this issue). Whether one considers the extent to be relevant or not, mean structures can always be examined at the latent level, so that age-dependent factor means and time trends may be legitimately studied.

Establishing strict MI for SOC proved difficult, given that a negative cross-loading from item #4 on the comprehensibility-manageability factor was required for appropriate model fit. When assuming equal residuals, SOC levels depended negatively on having clear life goals in early adulthood, but not during early adolescence. Being fully aware of the exploratory nature of this modification, it corresponds with Antonovsky's conception of SOC, according to which SOC can only be cultivated in the face of life's difficulties and obstacles. Unable to pursue one's goals, SOC may get strengthened as one learns to flexibly give up goals, whereas rigidity may reflect a maladjustment to unattainable goals. In sum, equal reliability across age can only be achieved if a fine-tuned measurement model is accepted. We conclude that SOC scores cannot be compared straightforwardly across age groups, unless this item is excluded from manifest scale sums (which then implies that the content of the scale changes too), or handled appropriately in SEM.

At the level of structural invariance, equality of variances and covariances held for both SOC and GSE. The equality of latent means across age groups was not supported. Previous examinations of SOC shifts across adolescence provided mixed findings (Rivera et al., 2013). Some have observed stable SOC means (Honkinen et al., 2008), yet we found an increase in SOC over the course of nine years ($d = 0.33$). This upward shift supports salutogenic theory, purporting developmental gains of general resistance resources. There was similar evidence for increasing self-efficacy beliefs ($d = 0.59$). Given the quality of both scales, prior inconsistent findings cannot be traced back to inadequate psychometric properties, or severe incompatibility of people's scores at different age. The changes across time are real and, given their sizes, potentially relevant.

The different magnitude of the SOC and GSE changes is supported by their unlike stability coefficients. The latent test-retest reliability for SOC closely matches a previously reported figure of $r = .54$ over a span of ten years for adult samples (Eriksson & Lindström, 2005). Considering participants' young age at the study onset, and in light of hypothesized SOC fluctuations throughout adolescence, the stability of individual differences across nearly a decade is remarkable. Despite its favorable unidimensional factor structure, GSE exhibited only half of SOC's temporal stability. A sizeable latent state component may have gone unnoticed so far, challenging the conception of GSE as a stable latent trait pertaining to one's generalized abilities and expectations of success.

With regard to longitudinal criterion correlations, SOC is likely to outperform GSE by magnitudes. Miyoshi (2012) observed mutual influences between GSE and task-specific self-efficacy in college students over a period of 2 months. Hence, GSE appears to be rather volatile, thus limiting its utility as a long-term predictor. Indeed, our own data showed that GSE had almost no incremental validity over SOC when predicting mental health and drug use over several years (Grevenstein, Bluemke, & Kroeninger-Jungaberle, 2015). The causes of GSE's instability are beyond the scope of the present work and remain to be addressed in the future. For instance, whether, and how, systematic GSE shifts are related to individual and aggregate changes of stress-buffering, coping strategies, and other psychological resources remains unbeknown.

Conclusions

Our findings speak favorably about the use of the SOC-13 and GSE scales with adolescents, with a grain of salt though. Apart from factorial validity, both scales exhibited metric and (at least partial) scalar invariance, warranting the longitudinal examination of covariance structures as well as mean structures. We doubt that problematic SOC items should be completely dropped from the scales. The invariance of one item intercept ("estimating the importance of something happening") does not yet invalidate scale use; another item ("having clear goals in life") was unexpectedly but plausibly sensitive to developmental processes.

Considering their rather low factor loadings, the GSE items might be in much greater need for refinement than SOC. Though construct reliability was not substantially lower than for SOC, GSE showed very low longitudinal stability, placing serious constraints on how suitable a long-term predictor it is. Yet, on a theoretical note, the stability of SOC may come as an even bigger challenge for future research: How does salutogenic theory accommodate the fact that adolescents do not – transitorily – acquire their resources in line with theory, but appear to be steadily endowed with them?

Acknowledgments
This research is part of a longitudinal study on salutogenesis and drug consumption patterns funded by the German Research Council (DFG) from 2002 to 2013 within its Collaborative Research Centre (Sonderforschungsbereich) 619.

Declaration of Interest
The authors declare that they have no conflict of interest.

Electronic Supplementary Material
The electronic supplementary material is available with the online version of the article at http://dx.doi.org/10.1027/1015-5759/a000294

ESM 1. Text and Tables (word file).
Exploration of the factorial validity of the scales assessing Sense of coherence (SOC) and General Self-efficacy (GSE).
ESM 2. MPlus file (.inp)
Mplus input file for SOC configural invariance
ESM 3. MPlus file (.inp)
Mplus input file for SOC metric invariance
ESM 4. MPlus file (.inp)
Mplus input file for SOC scalar invariance
ESM 5. MPlus file (.inp)
Mplus input file for SOC partial scalar invariance
ESM 6. MPlus file (.inp)
Mplus input file for SOC strict invariance
ESM 7. MPlus file (.inp)
Mplus input file for SOC partial strict invariance
ESM 8. MPlus file (.inp)
Mplus input file for SOC variance invariance
ESM 9. MPlus file (.inp)
Mplus input file for SOC full invariance
ESM 10. MPlus file (.inp)
Mplus input file for GSE configural invariance
ESM 11. MPlus file (.inp)
Mplus input file for GSE metric invariance
ESM 12. MPlus file (.inp)
Mplus input file for GSE scalar invariance
ESM 13. MPlus file (.inp)
Mplus input file for GSE strict invariance
ESM 14. MPlus file (.inp)
Mplus input file for GSE variance invariance
ESM 15. MPlus file (.inp)
Mplus input file for GSE full invariance

References

Abel, T., Kohlmann, T., & Noack, H. (1995). *Eine deutsche Übersetzung des SOC* [A German translation of the SOC]. Bern, Switzerland: Institut für Sozial- und Präventivmedizin.
Akaike, H. (1987). Factor analysis and AIC. *Psychometrika, 52*, 317–332. doi: 10.1007/BF02294359
Andersson, L. M. C., Moore, C. D., Hensing, G., Krantz, G., & Staland-Nyman, C. (2014). General self-efficacy and its relationship to self-reported mental illness and barriers to care: A general population study. *Community Mental Health Journal, 50*, 712–728. doi: 10.1007/s10597-014-9722-y
Antonovsky, A. (1987). *Unraveling the mystery of health: How people manage stress and stay well.* San Francisco, CA: Jossey-Bass.
Antonovsky, A. (1993). The structure and properties of the sense of coherence scale. *Social Science & Medicine, 36*, 725–733. doi: 10.1016/0277-9536(93)90033-Z
Antonovsky, A. (1998). The sense of coherence: An historical and future perspective. In H. I. McCubbin, E. A. Thompson, A. I. Thompson, & J. E. Fromer (Eds.), *Stress, coping, and health in families: Sense of coherence and resiliency* (pp. 3–20). Thousand Oaks, CA: Sage.
Bacon, D. R., Sauer, P. L., & Young, M. (1995). Composite reliability in structural equations modeling. *Educational and Psychological Measurement, 55*, 394–406. doi: 10.1177/0013164495055003003
Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*, 191–215. doi: 10.1037/0033-295x.84.2.191
Bandura, A. (1997). *Self-efficacy: The exercise of control* (1st print. ed.). New York, NY: Freeman.
Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology, 52*, 1–26. doi: 10.1146/annurev.psych.52.1.1
Bentler, P. M. (1990). Comparative fit indexes in structural models. Psychological Bulletin, 107, 238–246. doi: 10.1037/0033-2909.107.2.238
Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. Psychological Bulletin, 88, 588–606. doi: 10.1037/0033-2909.88.3.588
Bollen, K. A. (1989). Structural equations with latent variables. New York, NY: Wiley.
Brown, T. A. (2006). Confirmatory factor analysis for applied research. New York, NY: Guilford Press.
Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Beverly Hills, CA: Sage.
Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105*, 456–466. doi: 10.1037/0033-2909.105.3.456
Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*, 464–504. doi: 10.1080/10705510701301834
Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology, 95*, 1005–1018. doi: 10.1037/a0013193
Chen, G., Gully, S. M., & Eden, D. (2001). Validation of a new general self-efficacy scale. *Organizational Research Methods, 4*, 62–83. doi: 10.1177/109442810141004
Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 9*, 233–255. doi: 10.1207/S15328007SEM0902_5
Ciarrochi, J., Said, T., & Deane, F. P. (2005). When simplifying life is not so bad: the link between rigidity, stressful life events, and mental health in an undergraduate population. *British Journal of Guidance & Counselling, 33*, 185–197. doi: 10.1080/03069880500132540
Egger, J. W. (2011). Selbstwirksamkeitserwartung – ein bedeutsames kognitives Konstrukt für gesundheitliches Verhalten [Self-efficacy: An important cognitive construct for health behavior]. *Psychologische Medizin, 22*, 43–58.

Eriksson, M., & Lindström, B. (2005). Validity of Antonovsky's sense of coherence scale: *A systematic review. Journal of Epidemiology and Community Health, 59*, 460–466. doi: 10.1136/jech.2003.018085
Eriksson, M., & Lindström, B. (2006). Antonovsky's Sense of Coherence Scale and the relation with health: *A systematic review. Journal of Epidemiology and Community Health, 60*, 376–381. doi: 10.1136/jech.2005.041616
Fan, X., & Sivo, S. A. (2009). Using Δ goodness-of-fit indexes in assessing mean structure invariance. *Structural Equation Modeling, 16*, 54–69. doi: 10.1080/10705510802561311
Feldt, T., Leskinen, E., Kinnunen, U., & Mauno, S. (2000). Longitudinal factor analysis models in the assessment of the stability of sense of coherence. *Personality and Individual Differences, 28*, 239–257. doi: 10.1016/s0191-8869(99)00094-x
Feldt, T., Leskinen, E., Kinnunen, U., & Ruoppila, I. (2003). The stability of sense of coherence: Comparing two age groups in a 5-year follow-up study. *Personality and Individual Differences, 35*, 1151–1165. doi: 10.1016/s0191-8869(02)00325-2
Feldt, T., Lintula, H., Suominen, S., Koskenvuo, M., Vahtera, J., & Kivimäki, M. (2007). Structural validity and temporal stability of the 13-item sense of coherence scale: Prospective evidence from the population-based HeSSup study. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care & Rehabilitation, 16*, 483–493. doi: 10.1007/s11136-006-9130-z
Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*, 39–50. doi: 10.2307/3151312
Frenz, A. W., Carey, M. P., & Jorgensen, R. S. (1993). Psychometric evaluation of Antonovsky's sense of coherence scale. *Psychological Assessment, 5*, 145–153. doi: 10.1037/1040-3590.5.2.145
Gana, K., & Garnier, S. (2001). Latent structure of the sense of coherence scale in a french sample. *Personality and Individual Differences, 31*, 1079–1090. doi: 10.1016/S0191-8869(00)00205-1
Geyer, S. (1997). Some conceptual considerations on the sense of coherence. *Social Science & Medicine, 44*, 1771–1779. doi: 10.1016/S0277-9536(96)00286-9
Grevenstein, D., Bluemke, M., & Kroeninger-Jungaberle, H. (2015). *Incremental validity of sense of coherence, neuroticism, extraversion, and general self-efficacy: Longitudinal prediction of substance use frequency and mental health.* Manuscript submitted for publication.
Hagquist, C., & Andrich, D. (2004). Is the Sense of Coherence-instrument applicable on adolescents? A latent trait analysis using Rasch-modelling. *Personality and Individual Differences, 36*, 955–968. doi: 10.1016/S0191-8869(03)00164-8
Haukkala, A., Konttinen, H., Lehto, E., Uutela, A., Kawachi, I., & Laatikainen, T. (2013). Sense of coherence, depressive Symptoms, cardiovascular diseases, and all-cause mortality. *Psychosomatic Medicine, 75*, 429–435. doi: 10.1097/PSY.0b013e31828c3fa4
Hinz, A., Schumacher, J., Albani, C., Schmid, G., & Brähler, E. (2006). Bevölkerungsrepräsentative Normierung der Skala zur Allgemeinen Selbstwirksamkeitserwartung [Standardization of the General Self-Efficacy Scale in the German population]. *Diagnostica, 52*, 26–32. doi: 10.1026/0012-1924.52.1.26
Hittner, J. B. (2007). Factorial invariance of the 13-item sense of coherence scale across gender. *Journal of Health Psychology, 12*, 273–280. doi: 10.1177/1359105307074256
Holmberg, S., Thelin, A., & Stiernström, E. (2004). Relationship of sense of coherence to other psychosocial indices. *European Journal of Psychological Assessment, 20*, 227–236. doi: 10.1027/1015-5759.20.4.227
Honkinen, P., Aromaa, M., Suominen, S., Rautava, P., Sourander, A., Helenius, H., & Sillanpää, M. (2009). Early childhood psychological problems predict a poor sense of coherence in adolescents: A 15-year follow-up study. *Journal of Health Psychology, 14*, 587–600. doi: 10.1177/1359105309103578
Honkinen, P., Suominen, S., Helenius, H., Aromaa, M., Rautava, P., Sourander, A., & Sillanpää, M. (2008). Stability of the sense of coherence in adolescence. *International Journal of Adolescent Medicine and Health, 20*, 85–91. doi: 10.1515/ijamh.2008.20.1.85
Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55. doi: 10.1080/10705519909540118
Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: practical and theoretical issues. *Multivariate Behavioral Research, 32*, 53–76. doi: 10.1207/s15327906mbr3201_3
Lubke, G. H., & Dolan, C. V. (2003). Can unequal residual variance across groups mask differences in residual means in the common factor model? *Structural Equation Modeling, 10*, 175–192. doi: 10.1207/S15328007SEM1002_1
Luszczynska, A., Gutiérrez-Doña, B., & Schwarzer, R. (2005). General self-efficacy in various domains of human functioning: Evidence from five countries. *International Journal of Psychology, 40*, 80–89. doi: 10.1080/00207590444000041
Marsh, H. W., & Grayson, D. (1994). Longitudinal confirmatory factor analysis: Common, time-specific, item-specific, and residual-error components of variance. *Structural Equation Modeling: A Multidisciplinary Journal, 1*, 116–145. doi: 10.1080/10705519409539968
Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*, 525–543. doi: 10.1007/BF02294825
Miyoshi, A. (2012). The stability and causal effects of task-specific and generalized self-efficacy in college. *Japanese Psychological Research, 54*, 150–158. doi: 10.1111/j.1468-5884.2011.00481.x
Moksnes, U. K., Espnes, G. A., & Haugan, G. (2013). Stress, sense of coherence and emotional symptoms in adolescents. *Psychology & Health, 29*, 32–49. doi: 10.1080/08870446.2013.822868
Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.
Nilsson, K. W., Leppert, J., Simonsson, B., & Starrin, B. (2010). Sense of coherence and psychological well-being: Improvement with age. *Journal of Epidemiology and Community Health, 64*, 347–352. doi: 10.1136/jech.2008.081174
Peter, C., Cieza, A., & Geyh, S. (2014). Rasch analysis of the general self-efficacy scale in spinal cord injury. *Journal of Health Psychology, 19*, 544–555. doi: 10.1177/1359105313475897
Posadzki, P., Stockl, A., Musonda, P., & Tsouroufli, M. (2010). A mixed-method approach to sense of coherence, health behaviors, self-efficacy and optimism: Towards the operationalization of positive health attitudes. *Scandinavian Journal of Psychology, 51*, 246–252. doi: 10.1111/j.1467-9450.2009.00764.x
Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology, 25*, 111–163. doi: 10.2307/271063
Richardson, C. G., Ratner, P. A., & Zumbo, B. D. (2007). A test of the age-based measurement invariance and temporal stability of Antonovsky's sense of coherence scale. *Educational and Psychological Measurement, 67*, 679–696. doi: 10.1177/0013164406292089
Rivera, F., García-Moya, I., Moreno, C., & Ramos, P. (2013). Developmental contexts and sense of coherence in adolescence: A systematic review. *Journal of Health Psychology, 18*, 800–812. doi: 10.1177/1359105312455077
Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin, 130*, 261–288. doi: 10.1037/0033-2909.130.2.261
Saris, W. E., Satorra, A., & van der Veld, W. M. (2009). Testing structural equation models or detection of misspecifications? *Structural Equation Modeling, 16*, 561–582. doi: 10.1080/10705510903203433
Satorra, A. (2000). Scaled and adjusted restricted tests in multi-sample analysis of moment structures. In R. D. H. Heijmans, D. S. G. Pollock, & A. Satorra (Eds.), *Innovations in multivariate statistical analysis. A Festschrift for Heinz Neudecker* (pp. 233–247). London, UK: Kluwer Academic.
Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66*, 507–514. doi: 10.1007/BF02296192
Scherbaum, C. A., Cohen-Charash, Y., & Kern, M. J. (2006). Measuring general self-efficacy: A comparison of three measures using item response theory. *Educational and Psychological Measurement, 66*, 1047–1063. doi: 10.1177/0013164406288171
Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review, 18*, 210–222. doi: 10.1016/j.hrmr.2008.03.003
Scholz, U., Doña, B. G., Sud, S., & Schwarzer, R. (2002). Is general self-efficacy a universal construct? Psychometric findings from 25 countries. *European Journal of Psychological Assessment, 18*, 242–251. doi: 10.1027/1015-5759.18.3.242

Schumacher, J., Gunzelmann, T., & Brähler, E. (2000). Deutsche Normierung der Sense of Coherence Scale von Antonovsky [Standardization of Antonovsky's Sence of Coherence Scale in the German population]. *Diagnostica, 46*, 208–213. doi: 10.1026/0012-1924.46.4.208
Schumacker, R. E., & Lomax, R. G. (2010). *A beginner's guide to structural equation modeling*. New York, NY: Routledge.
Schwarzer, R. (1992). *Self-efficacy: Thought control of action*. Washington, DC: Hemisphere Publishing.
Schwarzer, R., & Jerusalem, M. (1995). Generalized Self-Efficacy scale. In J. Weinman, S. Wright, & M. Johnston (Eds.), *Measures in health psychology: A user's portfolio. Causal and control beliefs* (pp. 35–37). Windsor, UK: NFER-Nelson.
Schwarzer, R., & Jerusalem, M. (1999). *Skalen zur Erfassung von Lehrer- und Schülermerkmalen. Dokumentation der psychometrischen Verfahren im Rahmen des Wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen* [Scales for the assessment of teacher and student charecteristics]. Berlin, Germany: Freie Universität Berlin.
Stajkovic, A. D., & Luthans, F. (1998). Self-efficacy and work-related performance: A meta-analysis. *Psychological Bulletin, 124*, 240–261. doi: 10.1037/0033-2909.124.2.240
Steenkamp, J. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25*, 78–90. doi: 10.1086/209528
Stein, J. A., Lee, J. W., & Jones, P. S. (2006). Assessing cross-cultural differences through use of multiple-group invariance analyses. *Journal of Personality Assessment, 87*, 249–258. doi: 10.1207/s15327752jpa8703_05
Steinmetz, H. (2013). Analyzing observed composite differences across groups: Is partial measurement invariance enough? *Methodology, 9*, 1–12. doi: 10.1027/1614-2241/a000049
Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Allyn & Bacon/Pearson Education.
Tipton, R. M., & Worthington, E. L. (1984). The measurement of generalized self-efficacy: A study of construct validity. *Journal of Personality Assessment, 48*, 545–548. doi: 10.1207/s15327752jpa4805_14
Togari, T., Yamazaki, Y., Takayama, T. S., Yamaki, C. K., & Nakayama, K. (2008). Follow-up study on the effects of sense of coherence on well-being after two years in Japanese university undergraduate students. *Personality and Individual Differences, 44*, 1335–1347. doi: 10.1016/j.paid.2007.12.002
Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*, 4–69. doi: 10.1177/109442810031002
Zimprich, D., Allemand, M., & Hornung, R. (2006). Measurement invariance of the abridged sense of coherence scale in adolescents. *European Journal of Psychological Assessment, 22*, 280–287. doi: 10.1027/1015-5759.22.4.280
Zirke, N., Schmid, G., Mazurek, B., Klapp, B. F., & Rauchfuss, M. (2007). Antonovsky's Sense of Coherence in psychosomatic patients: A contribution to construct validation. *GMS Psycho-Social-Medicine, 4*, 1–9.

Date of acceptance: May 7, 2015
Published online: November 30, 2015

Dennis Grevenstein
Psychological Institute
University of Heidelberg
Hauptstraße 47-51
69117 Heidelberg
Germany
Tel. +49 6221 547318
E-mail dennis.grevenstein@psychologie.uni-heidelberg.de