Arens, A. Katrin; Preckel, Franzis

Distinct constructs with a common core. Modelling neuroticism, self-esteem, and generalized self-efficacy using bifactor exploratory structural equation modelling.

Current psychology 44 (2025) 12, S. 12209-12224

Current Psychology (2025) 44:12209–12224
https://doi.org/10.1007/s12144-025-07984-9

Distinct constructs with a common core: modelling neuroticism, self-esteem, and generalized self-efficacy using bifactor exploratory structural equation modelling

A. Katrin Arens[1] · **Franzis Preckel**[2]

Accepted: 15 May 2025 / Published online: 6 June 2025

Abstract

Neuroticism, self-esteem, and generalized self-efficacy are conceptually closely related constructs. This study investigates whether the common core of the three constructs goes back to a conceptual overlap and/or to a hierarchical ordering of the constructs. We tested a confirmatory factor analyses (CFA) model, an exploratory structural equation model (ESEM), a bifactor-CFA model, and an bifactor-ESEM model using a sample of German secondary school students at two measurement waves (t1: $N=1573$; t2: $N=1210$). At both waves, the bifactor-ESEM model emerged to be the best fitting model and showed a well-defined G-factor and well-defined S-factors. The bifactor-ESEM model displayed invariant factor loadings, item thresholds, and item residuals across t1 and t2. At t1 and t2, the G-factor of the bifactor-ESEM model displayed positive and statistically significant relations to general life satisfaction and school satisfaction; the S-factors showed inconsistent relations to these two outcome variables. Neuroticism, self-esteem, and generalized self-efficacy thus seem to be conceptually related as well as hierarchically ordered constructs. The bifactor-ESEM model can reconcile the two opposing views according to which neuroticism, self-esteem, and generalized self-efficacy are distinct versus similar constructs.

Keywords Neuroticism · Self-esteem · Generalized self-efficacy · Bifactor exploratory structural equation modelling · Confirmatory factor analyses · Construct-relevant psychometric multidimensionality

Neuroticism, generalized self-efficacy, and self-esteem are well-known and important constructs in personality research. Neuroticism has mostly been investigated in the context of the Big Five model (McCrae & Costa, 1997), according to which personality consists of five traits – neuroticism, extraversion, conscientiousness, agreeableness, and openness. Neuroticism is defined as the opposite of emotional stability and high levels of neuroticism reflect negative affect and emotional instability. High levels of neuroticism are positively related to psychopathology (e.g., McCrae & John, 1992; Kotov et al., 2010). Self-esteem reflects one's self-worth as well as the acceptance, respect, and value felt against oneself as a person (e.g., Donnellan et al., 2011; Harter, 2006; Rosenberg, 1979). High levels of self-esteem are related to positive affect, life satisfaction, job satisfaction, relationship satisfaction, and mental health (Orth et al., 2012; Orth & Robins, 2022). The construct of generalized self-efficacy originates from social cognitive theory (Bandura, 1986) and is defined as "beliefs in one's capabilities to mobilize the motivation, cognitive resources, and courses of action needed to meet given situational demands" (Wood & Bandura, 1989, p. 408). High levels of generalized self-efficacy are related to various advantageous outcomes including self-regulation, life satisfaction, planning for the future, optimism, or need for achievement (e.g., Azizli et al., 2015; Chen et al., 2004; Luszczynska et al., 2005).

✉ A. Katrin Arens
k.arens@dipf.de

[1] Department on Research on Education and Human Development and IDeA Research Center for Individual Development and Adaptive Education of Children at Risk, DIPF | Leibniz Institute for Research and Information in Education, Rostocker Str. 6, 60323 Frankfurt am Main, Germany

[2] University of Trier, Trier, Germany

The three constructs have been often examined separately, although respective studies pursue similar research questions, such as, for instance, the relation to well-being (for neuroticism: e.g., Gale et al., 2013; Grant et al., 2009; Steel et al., 2008; for self-esteem: e.g., Katsantonis et al., 2023; Muris & Otgaar, 2023; for generalized self-efficacy: e.g., Caprara & Steca, 2005). In addition, the three constructs show strong interrelations and have often served as the outcome of another (e.g., Chen et al., 2004; Marcionetti & Rossier, 2016; Strobel et al., 2011; Weidmann et al., 2018). There has been a debate whether the three constructs form distinct, albeit highly related constructs or whether they have a common core (Judge et al., 2002). In this study, we try to conciliate these points of view by proposing and testing various models including a confirmatory factor analytic (CFA) model, a model based on exploratory structural equation modeling (ESEM), a bifactor-CFA model, and an bifactor-ESEM model.

The distinctiveness of neuroticism, self-esteem, and generalized self-efficacy

Theoretically, neuroticism, self-esteem, and generalized self-efficacy can be distinguished as they originate from different research traditions. Neuroticism is part of the Big Five model, generalized self-efficacy is grounded in social cognitive theory (Bandura, 1986), and self-esteem is the core of self-esteem theories (e.g., Harter, 2006; Rosenberg, 1979). Asendorpf and van Aken (2003) apply the differentiation between surface and core characteristics to neuroticism and self-esteem. Surface characteristics are assumed to be less stable and more malleable due to environmental influences; core characteristics are assumed to be rather stable. Self-esteem is assumed to be a surface characteristic, whereas neuroticism is assumed to be a core characteristic.

Further supporting their distinctiveness, Chen et al. (2004) showed that self-esteem and generalized self-efficacy have different functionalities and vary in their outcome relations. Self-esteem was found to be more strongly related to affective traits (emotional stability, negative affectivity) than generalized self-efficacy. Generalized self-efficacy instead was found to be more strongly related to motivational traits (i.e., conscientiousness, need for achievement, openness, and learning goal orientation) and motivational states (i.e., specific self-efficacy, self-set goals, metacognition, effort, and behavior intentions) compared to self-esteem. Further supporting their distinctiveness in the study by Chen et al. (2004), the factor correlation between self-esteem and generalized self-efficacy was not perfect in CFA models (Sample 1: $r=0.80$; Sample 2: $r=0.79$). Moreover, when the correlation between generalized self-efficacy and self-esteem was set to one, this led to a significant decline in model fit compared to a model with a freely estimated correlation.

Additional evidence of the distinctiveness of the constructs comes from a behavior genetic perspective. Hufer-Thamm and Riemann (2021) presented a twin study integrating self-esteem, life satisfaction, and neuroticism. The majority of the variance shared by self-esteem and life satisfaction did not stem from genetic variance shared with neuroticism. The authors therefore concluded that self-esteem, life satisfaction, and neuroticism do not represent a unified construct.

The common core of neuroticism, self-esteem, and generalized self-efficacy

Despite the arguments for the distinctiveness of neuroticism, self-esteem, and generalized self-efficacy, research has highlighted the similarities of the three constructs. Judge et al. (1997) denominated neuroticism, self-esteem, and generalized self-efficacy as well as locus of control as core evaluations of the self. Core evaluations of the self are traits that share three important attributes: (1) they have an evaluation-focus; i.e., all constructs are related to evaluations, as opposite to descriptions, of the self; (2) they are fundamental; i.e., all constructs are source traits as opposite to surface traits; (3) they have a broad scope; i.e., all constructs are general in nature and encompass a variety of objects and entities.

In addition to these theoretical arguments, Judge et al. (2002) presented four empirical studies demonstrating that the constructs of neuroticism, self-esteem, generalized self-efficacy, and locus of control share a common core. Study 1 was a meta-analysis showing high population correlations among the constructs (average correlation of $r=0.60$). In Study 2, neuroticism, self-esteem, generalized self-efficacy, and locus of control were subject to a CFA. When stating separate first-order factors for each of the constructs (i.e., a first-order factor model), the factors were found to be highly correlated (average correlations: $r=0.63$ in Study 2a, $r=0.51$ in Study 2b). The first-order factors were found to form a higher-order factor given an improved model fit of a higher-order factor model compared to a first-order factor model. The four first-order factors displayed statistically significant and high loadings on the higher-order factor (neuroticism: $\lambda=-0.63/-0.74$; self-esteem: $\lambda=0.93/0.84$; generalized self-efficacy: $\lambda=0.77/0.75$; locus of control: $\lambda=0.82/0.54$ in Studies 2a and 2b, respectively). Study 3 investigated the discriminant validity using the multitrait-multimethod technique. The corresponding findings indicated a lack of discriminant validity among the constructs. Specially, the correlations between different constructs

measured by the same method (heterotrait–monomethod correlations) were higher than the correlations between the same constructs measured by different methods (monotrait–heteromethod correlations). For instance, the correlation between self-reported self-esteem and neuroticism was higher than the correlation between self-reported self-esteem and peer-reported self-esteem. Study 4 considered the correlations between neuroticism, self-esteem, generalized self-efficacy, and locus of control and various outcome variables assessed in seven empirical studies. The results demonstrated that neuroticism, self-esteem, generalized self-efficacy, and locus of control displayed a similar pattern of outcome relations. In fact, the four constructs similarly showed significant relations to the Big Five traits (agreeableness, conscientiousness, extraversion, and openness), job satisfaction, life satisfaction, and happiness. In addition, the four constructs all had similar relations to stress and strain. This finding again indicated a lack of discriminant validity of the constructs as "(…) each trait measure separately contributes little beyond the contribution of their common core " (Judge et al., 2002, p. 704; see also Judge & Bono, 2001). Based on these studies, Judge et al. (2002) concluded that neuroticism, self-esteem, generalized self-efficacy, and locus of control have so much in common that they form a higher-order construct. Still, Judge et al. (2002) did not advocate completely abandoning the treatment of the constructs as separate.

Beyond this evidence from personality psychology, arguments for the common core of neuroticism and self-esteem also stem from clinical psychology. Hankin et al. (2007) examined whether three cognitive vulnerability factors for depression are distinct or similar constructs. They conducted exploratory factor analyses (EFA) and CFA including negative cognitive style (based on hopelessness theory), dysfunctional attitudes (based on Beck's theory), and rumination (based on response-styles theory) as three cognitive vulnerability factors for depression together with self-esteem and neuroticism. The CFA findings indicated a four-factor model. Self-esteem and neuroticism, together with depression, loaded onto a common factor that was denominated as "mood". The other three factors of the four-factor model proved to be separate, albeit correlated factors for each of the three cognitive vulnerability factors for depression.

The source of the common core

As outlined so far, the three constructs of neuroticism, self-esteem, and generalized self-efficacy seem to share a common core. The relevant question addresses the source of this common core. To provide an answer to this question, we refer to the concept of psychometric multidimensionality (Morin et al., 2020; Swami et al., 2023).[1] Construct-relevant psychometric multidimensionality occurs when items designed to measure different constructs have something in common– a common core– that is not explained by the factors the items are assumed to measure. Construct-relevant psychometric multidimensionality in our case would mean that the items purported to measure neuroticism, self-esteem, and generalized self-efficacy do not only reflect their corresponding factors, but also reflect a common core that is not explained by the factors. This common core might be present because the constructs are (1) conceptually related, (2) hierarchically ordered, or (3) both.

Conceptually related constructs

In the case of construct-relevant psychometric multidimensionality due to conceptually related constructs, the items show content overlap. That is, items assumed to measure a specific construct are also associated with other constructs. For instance, items measuring neuroticism might have also associations with the constructs of self-esteem and generalized self-efficacy. This content overlap can be addressed by EFA as EFA allows item cross-loadings on non-target factors in addition to the main loadings on the target factors. Studies on the distinctiveness of neuroticism, self-esteem, and generalized self-efficacy, however, mostly rely on CFA. CFA, yet, suffers from the independent cluster model (ICM) according to which items are constrained to load only on a single factor and all non-target loadings are constrained to be zero. Therefore, CFA models cannot unveil psychometric multidimensionality due to conceptually related constructs because they do not include cross-loadings. EFA is therefore better suited to disclose psychometric multidimensionality due to conceptually related constructs given the inclusion of cross-loadings. However, EFA (in contrast to CFA) is not based on structural equation modelling (SEM) although SEM has many advantages as SEM distinguishes between manifest item indicators and latent constructs; SEM models can thus be evaluated by goodness-of-fit indicators and allow invariance testing.

ESEM is a combination of CFA and EFA. Similar to CFA, ESEM is based on SEM, and similar to EFA, ESEM allows cross-loadings (for an introduction to ESEM, see Morin et

[1] One can distinguish between construct-irrelevant psychometric multidimensionality and construct-relevant psychometric multidimensionality. Construct-irrelevant psychometric multidimensionality often orginates from item wording effects. When an item subset has parallel wordings or has to be reversely coded, these items are associated with each other given shared method variance. Hence, in the case of construct-irrelevant psychometric multidimensionality, the associations between items originate from methodological reasons and not from content-related reasons the latter representing construct-relevant psychometric multidimensionality as described in the text.

al., 2020). In ESEM, each item has a main loading on its target factor and cross-loadings on the other, non-target factors. By applying the target rotation procedure, within an item set, it is possible to a priori specify which items should show near-zero cross-loadings and which items should have freely estimated main loadings. Therefore, ESEM models with target rotation get a confirmatory character; in our case, it is possible to test whether items measuring neuroticism, self-esteem, and generalized self-efficacy load on other (non-target) factors beyond their target factors.

ESEM has been found to lead to decreased factor correlations compared to CFA (Asparouhov et al., 2012; Marsh et al., 2014; Morin et al., 2020). Due to the ICM in CFA, conceptual overlap between items can be only expressed through factor correlations leading to higher factor correlations; in ESEM, conceptual overlap can be expressed through factor correlations and cross-loadings leading to lower factor correlations.

So far, the factor structure of neuroticism, self-esteem, and generalized self-efficacy has not been examined using ESEM. The consistently found high factor correlations among the three constructs in CFA studies (e.g., Barańczuk, 2021; Diseth et al., 2014; Judge et al., 1998; Robins et al., 2001) might express the conceptual overlap between the constructs and might decrease when cross-loadings are allowed in ESEM.

Hierarchically-ordered constructs

In the case of hierarchically ordered constructs, items are assumed to measure a specific dimension as well as a more global construct. In the case of neuroticism, self-esteem, and generalized self-efficacy, items measuring neurotisicm might assess the construct of neuroticism itself as well as a global construct simultaneously underlying neuroticism, self-esteem, and generalized self-efficacy. Hierarchical factor analyses are commonly applied to examine the presence of construct-relevant psychometric multidimensionality due to hierarchically ordered constructs (Rindskopf & Rose, 1988). There has already been empirical evidence for the existence of construct-relevant psychometric multidimensionality due to hierarchically ordered constructs in the case of neuroticism, self-esteem, and generalized self-efficacy along with locus of control. In fact, Judge et al. (2002) demonstrated the adequacy of a higher-order factor model with neuroticism, self-esteem, generalized self-efficacy, and locus of control as first-order factors. However, higher-order factor models suffer from the proportionality constraint according to which the ratio of variance proportional to the first-order factors to variance proportional to the higher-order factor is restricted to be the same across a given set of items measuring the constructs considered (Brunner et al., 2012; Gignac, 2016). Therefore, bifactor models have been discussed as a more flexible approach to display a hierarchical structure of constructs (Chen et al., 2006; Reise et al., 2010). In a bifactor model, all item indicators are assumed to define a global factor (G-Factor) and a variety of specific factors (S-factors) whereby all factors are orthogonal (uncorrelated). Therefore, the total covariance existent among the item indicators is partitioned into the G-factor and a variety of S-factors. The G-factor explains the covariance shared among all item indicators, and the S-factors explain the residual covariance of the item subset not explained by the G-factor. In our case, for instance, items assumed to measure neuroticism might load on the G-factor together with the items measuring self-esteem and self-efficacy, but the items assumed to measure neuroticism are also assumed to define a S-factor of neuroticism.

Finally, the common core of neuroticism, self-esteem, and generalized self-efficacy might be traced back to the simultaneous existence of both forms of psychometric multidimensionality. This can be tested through a bifactor-ESEM model that combines ESEM (testing for psychometric multidimensionality due to conceptually related constructs) and bifactor modelling (testing for psychometric multidimensionality due to hierarchically ordered constructs). In bifactor-ESEM models, all items are assumed to have substantial main loadings on the G-factor and on their corresponding S-factor (see the bifactor model). In addition, all items are assumed to have smaller cross-loadings on their non-corresponding S-factors (see the ESEM model) (Morin et al., 2016, 2020; Swami et al., 2023). Hence, in our case, for instance, items assumed to measure neuroticism are supposed to display main loadings on the G-factor and on their corresponding S-factor of neuroticism in addition to non-target cross-loadings on the other S-factors of self-esteem and generalized self-efficacy.

The present study

To explore the source of the common core of neuroticism, self-esteem, and generalized self-efficacy, we test for the presence of both forms of psychometric multidimensionality. To this aim, we follow the taxonomy of models as proposed by Swami et al. (2023). First, we state a CFA model as a baseline model (Fig. 1a). Second, we test for the presence of psychometric multidimensionality due to conceptually related constructs by stating an ESEM model (Fig. 1b). Support for the appropriateness of ESEM models is evident when ESEM models have a superior fit to the data and lead to reduced factor correlations relative to CFA models. In addition, the pattern of factor loadings should include small to moderate cross-loadings of the items on their non-target

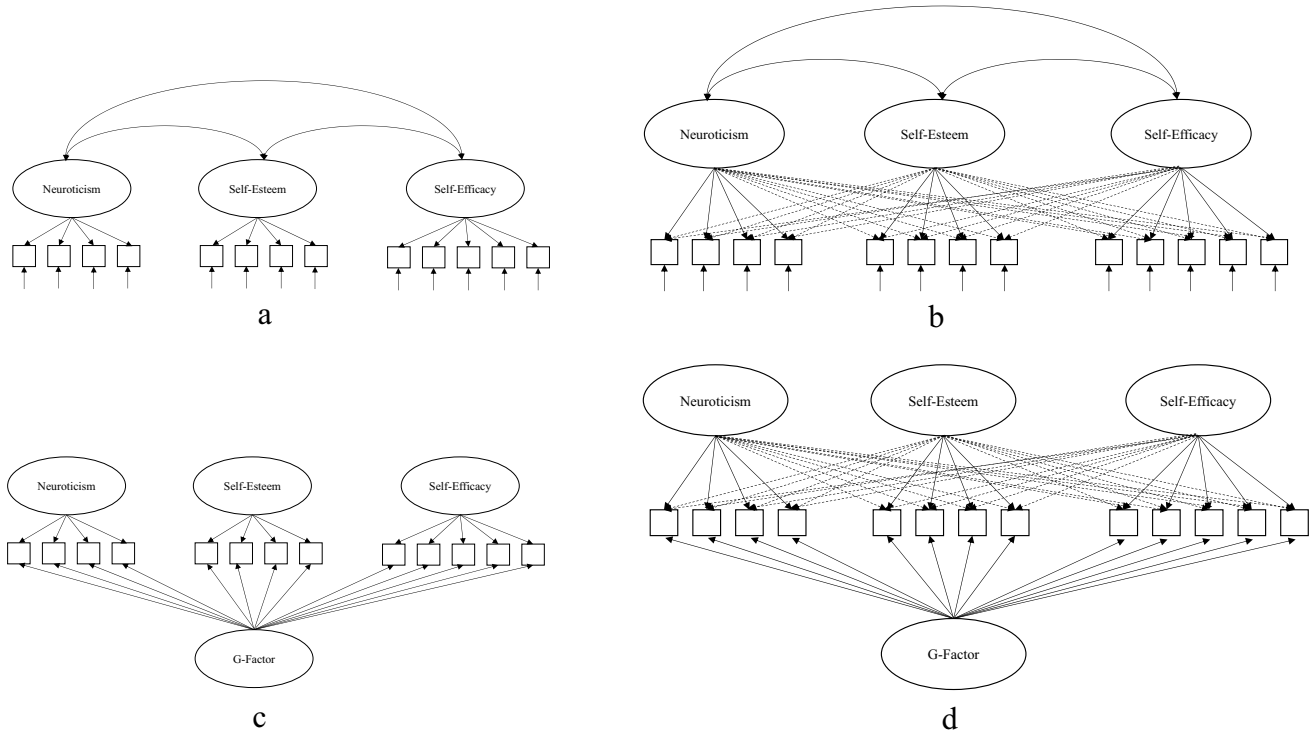

Fig. 1 **a**. The CFA model of neuroticism, self-esteem, and generalized self-efficay. **b**. The ESEM model of neuroticism, self-esteem, and generalized self-efficay. *Note.* Solid lines represent main loadings; dashed lindes represent cross-loadings. **c**. The bifactor-CFA model of neuroticism, self-esteem, and generalized self-efficay. **d**. The bifactor-ESEM model of neuroticism, self-esteem, and generalized self-efficay. *Note.* Solid lines represent main loadings; dashed lindes represent cross-loadings. Error terms of the items are omitted for the sake of readability of the figure

factors (ideally lower than 0.200, at least lower than 0.300) and higher main loadings of the items on their target factors (at least exceeding 0.300, ideally exceeding 0.500; Morin et al., 2020). Third, we test for the presence of psychometric multidimensionality due to a hierarchical ordering of the constructs by stating a bifactor model (Fig. 1c). Bifactor models are supported when they have an improved level of fit to the data compared to alternative modelling approaches, and when the G-Factor and S-factors are well-defined. More concretely, all items should have substantial loadings on the G-factor and on their corresponding S-factors (Morin et al., 2016, 2020). Fourth, we test for the simultaneous existence of both forms of psychometric multidimensionality by applying a bifactor-ESEM model (Fig. 1d). The bifactor-ESEM model is supported if it has a superior fit to the data relative to the other models. In addition, the G-Factor and S-factors should be well-defined so that all items have substantial loadings on the G-factor and on their corresponding S-factors (at least exceeding 0.300, ideally exceeding 0.500). In addition, the pattern of factor loadings should include small to moderate cross-loadings of the items on their non-target S-factors (ideally lower than 0.200, at least lower than 0.300 (Morin et al., 2016, 2020).

After exploring the modelling approach to neuroticism, self-esteem, and generalized self-efficacy, we investigate the pattern of relations with two outcome variables, namely life satisfaction and school satisfaction. Life satisfaction is defined as the cognitive aspect of subjective well-being depicting individuals' overall evaluation of the quality of one's life (Pavot & Diener, 2008). Life satisfaction has been found to be negatively related to neuroticism (Gale et al., 2013), and positively related to self-esteem (Diener & Diener, 1995) and generalized self-efficacy (Azizli et al., 2015; see also Marcionetti & Rossier, 2021). School satisfaction is related to students' perceived quality of experiences and life at school. In fact, school satisfaction covers students' enjoyment and affective experiences at school (Zullig et al., 2011). School satisfaction has been commonly explored with regard to school-related outcomes (e.g., school absence, academic self-efficacy, achievement, and behaviour problems; Daily et al., 2020; DeSantis King et al., 2006; Huebner & McCullough, 2000); little research has been done with respect to its relation to neuroticism, self-esteem, and generalized self-efficacy. Life satisfaction and school satisfaction are closely related (Suldo et al., 2006); therefore, the pattern of relations between school

satisfaction and neuroticism, self-esteem, and generalized self-efficacy might be similar to the pattern of relations with life satisfaction.

The sample of the present study encompassed students from the academic track of the the German educational school system (Gymnasium). This is the highest ability track of the German secondary school system and graduation after 12 or 13 school years (depending on the federal state) allows entry to university (Becker et al., 2016). Our sample was repeatedly investigated at two measurement waves; the first measurement wave took part in Grade 11, the second measurement wave took part in Grade 13, i.e., just before leaving school. The longitudinal study design provides an optimal opportunity to replicate our findings across two measurement waves and to test for their generalizability. We thus first conduct the analyses separately at t1 and t2; afterwards, we test for the longitudinal measurement invariance (i.e., invariance of the factor pattern, factor loadings, item intercepts, and item residuals) of the final retained model (Millsap, 2011).

Method

Sample and study design

The data came from the longitudinal project "Evaluation of a Gifted Education Program: Effects of ability grouping of gifted students in special gifted classes in secondary schools in Rhineland-Palatinate" (Preckel & Schmidt, 2014). The project was approved by the Supervision and Services Directorate of Rhineland-Palatinate ("Aufsichts- und Dienstleistungsdirektion"; protocol number: 32-03 405/29/05). For the present study, we considered the sixth and seventh measurement wave of the project (from now on denominated as t1 and t2) when the students attended Grade 11 and Grade 13 because neuroticism, self-esteem, and generalized self-efficacy were only assessed at these measurement waves. We considered those students who had at least one valid item on the neuroticism, self-esteem, and generalized self-efficacy measures. At t1, the sample consisted of $N=1573$ students ($n=746$ male, $n=824$ female, $n=3$ no gender indicated). The mean age of the students at t1 was 16.90 years ($SD=0.79$). A number of $n=617$ students had German as a native language, whereas $n=70$ students had another language than German as the native language, and for $n=886$, there was no information on native language. At t2, the sample consisted of $N=1210$ students ($n=564$ male, $n=641$ female, $n=5$ no gender indicated). The mean age at t2 was 18.71 years ($SD=0.86$). A number of $n=570$ students had German as a native language, $n=63$ students had another language than German as the native language, and for $n=577$, there was no information on native language. For the longitudinal measurement invariance tests, we considered a sample of $N=1895$ students ($n=920$ male, $n=967$ female, $n=8$ no gender indicated) who provided at least one valid item on the neuroticism, self-esteem, and generalized self-efficacy measures at t1 or t2. In the longitudinal sample, $n=733$ students had German as a native language, $n=84$ students had another language than German as the native language, and for $n=817$, there was no information on native language. Information on fathers' and mothers' professions can be retrieved from Table S1 of the Online Supplements.

The present study was not preregistered. Due to data protection requirements, we cannot offer the full data. For replicability of the findings, we present all variables and the correlation matrices at both measurement waves under https://osf.io/26zyv/?view_only=7a6d1d248cc74bb58772b65ca45905c1.

Instruments

Neuroticism

Neuroticism was assessed with the short version of the Big Five Inventory (BFI-K); the BFI-K is a short German instrument to measure the Big Five (Rammstedt & John, 2005). The neuroticism scale consists of four items ("I get easily depressed, downhearted; I am relaxed and don't let stress upset me (to be reverse coded); I worry a lot; I get easily nervous and insecure"). The items were responded on a five-point Likert scale ranging from "fully incorrect" to "fully correct". The reliability estimates were good for this scale at both measurement waves when considering both the coefficient alpha (α) as well as the McDonald's Omega (ω) estimates of reliability: t1: $\alpha=0.753$, $\omega=0.752$; t2: $\alpha=0.764$, $\omega=0.768$.

Self-esteem

Self-esteem was assessed with four items from the German version of Rosenberg's self-esteem scale (Von Collani & Herzberg, 2003) ("I think I have a number of good qualities; I can do most things just as well as other people; I have a good opinion of myself; Overall, I'm happy with myself"). Higher scores on all items consistently represent higher levels of self-esteem. Responses were given on a five-point Likert scale ranging from "does not apply at all" to "fully applies". The reliability estimates were good for this scale at both measurement waves: t1: $\alpha=0.776$, $\omega=0.792$; t2: $\alpha=0.776$, $\omega=0.785$.

Generalized self-efficacy

Generalized self-efficacy was measured by five items from Schwarzer and Jerusalem (1999) ("I am confident that I

could deal efficiently with unexpected events; Thanks to my resourcefulness, I know how to handle unforeseen situations; I can solve most problems if I invest the necessary effort; If I am in trouble, I can usually think of a solution; I can usually handle whatever comes my way"). Higher scores on all items consistently represent higher levels of generalized self-efficacy. The response scale was the same as used for measuring self-esteem. The reliability estimates were good for this scale at both measurement waves: t1: $\alpha = 0.753$, $\omega = 0.747$; t2: $\alpha = 0.780$, $\omega = 0.783$.

Life satifsation

Life satisfaction was measured by a single item that was worded "How satisfied are you with your life, all things considered? " This item was retrieved from the Socio-Economic Panel (Entringer et al., 2022). The students responded to this item on a seven-point Likert scale ranging from "very unsatisfied" to "very satisfied".

School satisfaction

School satisfaction was measured by a single item that was worded "When you think about all your experiences at school, how satisfied are you with school?") This item was retrieved from the "Linzer Questionnaire of School- and Classroom Climate for Grades 8 to 13" (Eder, 1998). The students responded to this item on a seven-point Likert scale ranging from "very unsatisfied" to "very satisfied".

Statistical analyses

The analyses were conducted within the SEM framework using Mplus 8.10 (Muthén & Muthén, 1998, 2023). All models were estimated using the Weighted Least Squares (WLSMV) estimator to adequately handle ordinal data emerging from the Likert scales used in the instruments of this study (Lei & Shiverdecker, 2020). We followed the sequence of models as proposed by Morin et al., (2020; see also Swami et al., 2023) and present all corresponding syntaxes under https://osf.io/26zyv/?view_only=7a6d1d248cc74bb58772b65ca45905c1. Accordingly, we started with a traditional CFA model (Brown, 2015; Fig. 1a). In this model, the four items measuring neuroticism were assumed to load on a neuroticism factor, the four items measuring self-esteem were assumed to load on a self-esteem factor, and the five items measuring generalized self-efficacy were assumed to load on a generalized self-efficacy factor. To obtain model identification, the unstandardized loading of the first item on a factor was set to 1.

The second model was an ESEM model (Fig. 1b) that was estimated with oblique target rotation (Asparouhov & Muthén, 2009; Marsh et al., 2011). Therefore, the items of the three scales (neuroticism, self-esteem, and generalized self-efficacy) were specified to have main loadings on the factors the items were assumed to measure but also cross-loadings on the other factors. The cross-loadings were a priori specified to be close to zero, whereas the main loadings were freely estimated. The factors were allowed to correlate among each other.

The third model was the bifactor-CFA model (Fig. 1c). Here, all items were assumed to load on the G-factor; in addition, each item was assumed to load on its matching S-factor. Along with the common specification of bifactor models, the factors were orthogonal, thus, they did not correlate among each other. To obtain model identification, the unstandardized loading of the first items on a factor were freely estimated, while the factor variances were set to 1.

Finally, we estimated a bifactor-ESEM model. Similar to the ESEM model, we used target rotation and differentiated between main loadings and cross-loadings. Similar to the bifactor model, we assumed a G-factor and three S-factors. Each item was specified to have freely estimated main loadings on the G-factor and its matching S-factor as well as near-zero cross-loadings on the other S-factors. Along with the common specification of bifactor models, the factors were orthogonal.

We then tested for longitudinal measurement invariance using the most adequate model resulting from the model sequence described above. The sequence of longitudinal measurement invariance testing started with a model of configural invariance where the same factor model structure using the same item indicators for defining the same factors at t1 and t2 was stated (Millsap, 2011). Afterwards, we tested for the invariance of factor loadings across measurement waves (weak or metric invariance). We then tested for the invariance of item thresholds (strong or scalar invariance), and finally, we tested for the invariance of item residuals (strict invariance). In all models of longitudinal measurement invariance testing, we included correlated uniquenesses between repeatedly used items (i.e., the same items used at t1 and t2 for measuring neuroticism, self-esteem, and generalized self-efficacy, respectively) to control for potential shared method variance (Marsh & Hau, 1996).

In a final step, we added life satisfaction and school satisfaction as outcome variables at t1 and t2 to the most adequate model. Both outcomes (life satisfaction and school satisfaction) were estimated as single-item factors at both t1 and t2.

The amount of missing values on the item level was small (t1: range from 0.4%

to 3.8%; t2: range from 0.2% to 1.7%). Missing values were handled by the Full Information Maximum Likelihood (FIML) approach implemented in Mplus. The FIML

Table 1 Goodness-of-fit indices

Model	χ^2	df	CFI	TLI	RMSEA	SRMR
t1						
First-order CFA	673.938	62	.951	.938	.079	.037
First-order ESEM	255.006	42	.983	.968	.057	.018
Bifactor-CFA	377.486	52	.974	.961	.063	.027
Bifactor-ESEM	118.830	32	.993	.983	.042	.012
t2						
First-order CFA	490.909	62	.961	.951	.076	.036
First-order ESEM	187.439	42	.987	.975	.053	.017
Bifactor-CFA	226.815	52	.984	.976	.053	.024
Bifactor-ESEM	97.250	32	.994	.986	.041	.013
Longitudinal Measurement Invariance Tests						
Configural Invariance	365.786	204	.994	.990	.020	.018
Metric Invariance	452.509	240	.992	.989	.022	.022
Scalar Invariance	546.213	288	.990	.988	.022	.023
Strict Invariance	574.509	288	.989	.987	.023	.023

CFA confirmatory factor analysis; *ESEM* exploratory structural equation modelling; *CFI* comparative fit index; *TLI* Tucker-Lewis index; *RMSEA* root mean square error of approximation; *SRMR* standardized root mean squared residual. All χ^2 values are statistically significant ($p<.001$)

approach is known to be reliable and to lead to unbiased parameter estimates for handling missing data (Enders, 2010; Graham, 2009).

For model evaluation and model selection, we considered the goodness-of-fit indices and the pattern of factor loadings (Morin et al., 2020; Swami et al., 2023). We considered the fit of the different models by inspecting several commonly accepted descriptive goodness-of-fit indices: the comparative fit index (CFI), the Tucker-Lewis index (TLI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR). For the CFI and TLI, values between 0.90 and 0.95 are commonly accepted as indicators of good model fit (Hu & Bentler, 1999). For the RMSEA, values below 0.05 are indicative of a close fit, values between 0.05 and 0.08 are indicative of a reasonable fit, and values greater than 0.10 are indicative of a poor fit (Browne & Cudeck, 1993). For the SRMR, values below 0.05 indicate good model fit and values below 0.08 indicate adequate model fit (Diamantopoulos & Siguaw, 2000; Hu & Bentler, 1999). We further considered the factor loadings: ESEM would be supported by small to moderate cross-loadings of items on the non-target factors and well-defined target factors; bifactor representations would be supported by a well-defined G-factor, well-defined S-factors, and small cross-loadings of items on the non-target S-factors (Morin et al., 2020). For evaluation of the longitudinal measurement invariance, we relied on the guideline proposed by Cheung and Rensvold (2002). Accordingly, the CFI should not drop by more than $\Delta CFI \leq -0.01$ between two nested models (models with more and less invariance constraints) to state invariance.

Table 2 Factor correlations

	Neuroticism	Self-esteem
t1		
Neuroticism		
Self-esteem	-.651*/-.594*	
Generalized self-efficacy	-.598*/-.475*	.613*/.537*
t2		
Neuroticism		
Self-esteem	-.540*/-.503*	
Generalized self-efficacy	-.664*/-.556*	.605*/.528*

The coefficients before the slash refer to the confirmatory factor analysis model; the coefficients behind the slash refer to the exploratory structural equation modelling approach

* $p<.05$

Results

Model comparisons

At both measurement waves, the goodness-of-fit indices indicated an adequate fit of the CFA model (Table 1). The factors were well defined at t1 and t2 as the factor loadings ranged between 0.594 and 0.857 at t1 and between 0.529 and 0.870 at t2 (Table 3). The factor correlations were theoretically plausible given negative correlations between neuroticism and self-esteem and between neuroticism and generalized self-efficacy, and positive correlations between self-esteem and generalized self-efficacy (Table 2).

Regarding the ESEM model, the goodness-of-fit indices indicated a better model fit compared to the CFA model at both measurement waves (Table 1). The adequacy of the ESEM representations was supported by the pattern of factor loadings (Table 3). The main loadings were all statistically significant and high ranging from 0.477 to 0.907 at t1

Table 3 Standardized factor loadings

	Neuroticism				Self-Esteem			
	CFA	ESEM	Bifactor-CFA	Bifactor-ESEM	CFA	ESEM	Bifactor-CFA	Bifactor-ESEM
Neuro_1	.761*/.748*	**.621*/.707***	.374*/.468*	**.490*/.547***		-.112*/-.088*		-.152*/-.123*
Neuro_2	.604*/.558*	**.612*/.511***	.377*/.323*	**.397*/.332***		.070*/.053		.065*/.062
Neuro_3	.646*/.710*	**.850*/.875***	.651*/.654*	**.610*/.662***		.039/.030		-.027/-.039
Neuro_4	.761*/.820*	**.585*/.638***	.324*/.406*	**.445*/.482***		-.064*/.002		-.104*/-.040
Sesteem_1		.104*/.109*		.029/.037	.707*/.735*	**.778*/.835***	.480*/.636*	**.512*/.639***
Sesteem_2		.078*/.134*		.074*/.071*	.594*/.529*	**.510*/.495***	.236*/.348*	**.286*/.376***
Sesteem_3		-.053*/-.054*		-.121*/-.074*	.857*/.870*	**.907*/.863***	.605*/.632*	**.647*/.660***
Sesteem_4		-.138*/-.132*		-.145*/-.154*	.801*/.841*	**.693*/.746***	.380*/.525*	**.476*/.600***
Gself_1		-.072*/.016		-.079*/-.006		-.114*/-.107*		-.016/-.065*
Gself_2		.004/-.058*		-.030/-.007		-.036/.036		.045/-.004
Gself_3		-.188*/-.213*		-.010/.053*		.166*/.188*		-.039/-.045*
Gself_4		.071*/.056*		.101*/.023		.040/-.010		-.027/.007
Gself_5		.021/.010		.167*/.051		.057/.065*		-.134*/.008
	Generalized Self-Efficacy				G-Factor			
	CFA	ESEM	Bifactor-CFA	Bifactor-ESEM	Bifactor-CFA	Bifactor-ESEM		
Neuro_1		-.054*/.017		-.062*/-.007	-.624*/-.571*	-.516*/-.496*		
Neuro_2		-.098*/-.117*		.033/.075*	-.478*/-.441*	-.489*/-.452*		
Neuro_3		.136*/.110*		.127*/.048	-.460*/-.491*	-.454*/-.456*		
Neuro_4		-.149*/-.204*		-.106*/-.104	-.638*/-.656*	-.540*/-.592*		
Sesteem_1		.029/-.009		.004/-.023	.543*/.457*	.526*/.448*		
Sesteem_2		.192*/.170*		.038/.084*	.520*/.382*	.513*/.368*		
Sesteem_3		-.099*/-.036		-.040*/-.067*	.669*/.599*	.604*/.560*		
Sesteem_4		-.009/-.009		-.021/.014	.676*/.621*	.596*/.538*		
Gself_1	.594*/.681*	**.662*/.816***	.498*/.588*	**.516*/.493***	.402*/.520*	.415*/.607*		
Gself_2	.646*/.757*	**.716*/.702***	.551*/.382*	**.586*/.273***	.442*/.651*	.455*/.695*		
Gself_3	.788*/.780*	**.477*/.458***	.258*/.081	**.098 /-.226**	.680*/.783*	.756*/.897*		
Gself_4	.633*/.634*	**.687*/.710***	.512*/.471*	**.384*/.408***	.440*/.496*	.539*/.555*		
Gself_5	.617*/.624*	**.610*/.604***	.415*/.287*	**.201*/.207***	.461*/.550*	.631*/.592*		

*Neuro_* Items referring to neuroticism; *Sesteem_* Items referring to self-esteem; *Gself_* Items referring to general self-efficacy. The coefficients before the slash refer to t1, the coefficients after the slash refer to t2. Main loadings are printed in bold. CFA = confirmatory factor analysis. ESEM = exploratory structural equation modelling.

*$p < .05$

and from 0.458 to 0.875 at t2. The cross-loadings were considerably smaller (t1: range from −0.188 to 0.192; t2: range from −0.213 to 0.188). At both measurement waves, the factor correlations were smaller in the ESEM models than in the CFA models while maintaining the theoretical plausible pattern (Table 2). The adequacy of the ESEM models indicated that construct-relevant psychometric multidimensionality due to conceptually related constructs seems to be at play when measuring neuroticism, self-esteem, and generalized self-efficacy.

Psychometric multidimensionality due to hierarchically ordered constructs was tested by bifactor-CFA models. At both measurement waves, the bifactor-CFA models resulted in a better model fit compared to the CFA models (Table 1). The bifactor-CFA models were further supported by the factor loadings: The G-factor was well defined by all items due to statistically significant loadings. The items assumed to measure neuroticism had negative loadings on the G-factor (t1: −0.460 to −0.624; t2: −0.441 to −0.656), the items assumed to measure self-esteem and generalized self-efficacy had positive loadings on the G-factor (self-esteem: t1: 0.520 to 0.676, t2: 0.382 to 0.621; generalized self-efficacy: t1: 0.402 to 0.680, t2: 0.496 to 0.783). In addition, the corresponding items also showed positive and statistically significant loadings on their S-factors (except one item measuring generalized self-efficacy at t2) implying additional well-defined S-factors (neuroticism: t1: 0.324 to 0.651, t2: 0.323 to 0.654; self-esteem: t1: 0.236 to 0.605, t2: 0.348 to 0.632; generalized self-efficacy: t1: 0.258 to 0.551, t2: 0.081 to 0.588). The adequacy of the bifactor-CFA models indicated that construct-relevant psychometric multidimensionality due to hierarchically-related constructs seems to be at play when measuring neuroticism, self-esteem, and generalized self-efficacy.[2]

[2] We also conducted higher-order factor models using neuroticism, self-esteem, and generalized self-efficacy as first-order factors that are assumed to form a higher-order factor. The results can be found in

Table 4 Correlations of the bifactor-ESEM model

	Neuroticism	Self-esteem	Generalized self-efficacy	G-factor	General Life Satisfaction
t1					
General life satisfaction	-.164*	.152*	-.089	.514*	
School satisfaction	.100*	-.033	-.018	.284*	.237*
t2					
General life satisfaction	-.401*	.401*	.176*	.275*	
School satisfaction	-.010	.128*	.007	.120*	.243*

*$p<.05$

The bifactor ESEM model displayed a very good model fit at both measurement waves; the fit was better than the fits for the CFA model, the first-order ESEM model, and the bifactor-CFA model (Table 1). The adequacy of the bifactor-ESEM model was further substantiated by the pattern of factor loadings (Table 3). The G-factor was well-defined as all items had high and statistically significant loadings: The items for measuring neuroticism had negative loadings on the G-factor (t1: −0.454 to −0.540; t2: −0.452 to −0.592) and the items for measuring self-esteem and generalized self-efficacy had positive loadings (self-esteem: t1: 0.513 to 0.604, t2: 0.368 to 0.560; generalized self-efficacy: t1: 0.415 to 0.756, t2: 0.555 to 0.897). In addition, the S-factors were also well defined: Most items– except two items measuring generalized self-efficacy at t1 and t2, one item measuring generalized self-efficacy at t2, and one item measuring self-esteem at t1– had statistically significant loadings larger than 0.300 on their target S-factors. In addition, the cross-loadings of the items on their non-target S-factors were small and no cross-loading was larger than 0.300: t1: −0.152 to 0.167; t2: −0.219 to 0.084. The adequacy of the bifactor-ESEM models indicated that both forms of construct-relevant psychometric multidimensionality (i.e., due to conceptually related constructs and due to hierarchically ordered constructs) seem to be at play when measuring neuroticism, self-esteem, and generalized self-efficacy. The bifactor-ESEM model is thus kept as the final model that most adequately reflects the underlying structure of neuroticism, self-esteem, and generalized self-efficacy.

Longitudinal measurement invariance testing

For testing longitudinal measurement invariance of the bifactor-ESEM model, we started with a model of configural invariance that states the same factor structure (Table 1). Relative to this baseline model, the CFI values of the metric, scalar, and strict models assuming factor loading invariance, item threshold invariance, and item residual invariance did not drop by more than $\Delta CFI \leq -0.01$, respectively. Therefore, we could state full longitudinal measuring invariance of the bifactor-ESEM model, i.e., up to strict longitudinal measurement invariance.

Outcome relations

In the bifactor-ESEM model, the G-factor had statistically significant and positive relations to both outcomes at both measurement waves (t1: general life satisfaction: $r=0.514$, school satisfaction: $r=0.284$; t2: general life satisfaction: $r=0.275$, school satisfaction: $r=0.120$; for all $p<0.05$; Table 4).[3] The S-factors for neuroticism, self-esteem, and generalized self-efficacy showed significant relations to general life satisfaction at t1 and t2 (except for a significant relation between the S-factor for generalized self-efficacy and general life satisfaction at t1). The relations between the S-factors and school satisfaction were inconsistent: The S-factor for neuroticism had a significant positive relation to school satisfaction ($r=0.100$) at t1; the S-factor for self-esteem displayed a significant positive relation to school satisfaction ($r=0.128$) at t2; all other relations between the S-factors for neuroticism, self-esteem, and generalized self-efficacy and school satisfaction were nonsignificant at t1 and t2.

Discussion

Modelling approaches

Previous research has indicated conceptual overlap and strong associations between neuroticism, self-esteem, and generalized self-efficacy implying a common core of the constructs. The present study investigated whether the common core goes back to a conceptual overlap and/or to a hierarchical ordering of the constructs. To this aim, we followed the taxonomy of models proposed by Swami et al. (2023) to test for the presence of construct-relevant psychometric multidimensionality. After stating a CFA model as a baseline model, ESEM was applied to probe for the presence of

Table S2 to Table S3 of the Online Supplements. The fit was inferior compared to the CFA, ESEM, bifactor-CFA, and bifactor-ESEM models at both measurement waves.

[3] The outcome relations for the CFA model, the ESEM model, and the bifactor-CFA model are documented in Table S5 of the Online Supplements. The goodness-of-fit indices of the models including the outcome variables are presented in Table S4 of the Online Supplements.

construct-relevant psychometric multidimensionality due to conceptually related constructs; a bifactor-CFA model was applied to test for the presence of construct-relevant psychometric multidimensionality due hierarchically ordered constructs, and a bifactor-ESEM model was applied to test for the presence of both forms of construct-relevant psychometric multidimensionality. Finally, we included general life satisfaction and school satisfaction as two outcome variables.

The bifactor-ESEM model emerged to be the best fitting model according to the goodness-of-fit indices. The factor loading pattern also supported the bifactor-ESEM solution as it provided a well-defined G-factor and well-defined S-factors. Hence, the constructs of neuroticism, self-esteem, and generalized self-efficacy seem to be conceptually related as well as hierarchically ordered constructs. The common source of the constructs of neuroticism, self-esteem, and generalized self-efficacy thus goes back to the presence of both forms of construct-relevant psychometric multidimensionality. This finding could be replicated across two measurement waves with German secondary school students. In addition, we found full (i.e., strict) longitudinal measurement invariance of the bifactor-ESEM model as the factor structure, factor loadings, item intercepts, and item residuals were the same across both measurement waves.

Outcome relations

At t1 and t2, the bifactor-ESEM model showed a well-defined pattern of outcome relations with the G-factor as the G-factor displayed positive and statistically significant relations to the two outcome variables of general life satisfaction and school satisfaction. At both measurement waves, the G-factor showed stronger relations to general life satisfaction than to school satisfaction. The S-factors for neuroticism, self-esteem, and generalized self-efficacy also showed significant relations to general life satisfaction that were in the expected direction at t1 and t2 (i.e., positive relations for the S-factors of self-esteem and generalized self-efficacy; negative relations for the S-factor of neuroticism) with the exception of a significant relation between the S-factor for generalized self-efficacy and general life satisfaction at t1. The pattern of relations between the S-factors for neuroticism, self-esteem, and generalized self-efficacy and school satisfaction was, however, inconsistent at t1 and t2. Hence, the outcome relations for the G-factor and the S-factors were clearer for general life satisfaction than for school satisfaction. This is theoretically plausible as general life satisfaction has a global nature and does not refer to a specific domain. Thus, general life satisfaction might be conceptually close to neuroticism, self-esteem, and generalized self-efficacy as the latter constructs are also not tied to a specific domain. The construct of school satisfaction, in contrast, is domain-specific in nature and related to the school domain. School-related constructs such as the quality of student–teacher relations (Arens & Niepel, 2019), autonomy support (Ferguson et al., 2011), and academic achievement (Hui & Sun, 2010) might be more important predictors of school satisfaction compared to neuroticism, self-esteem, and generalized self-efficacy.

Strengths, limitations, and directions for future research

To our knowledge, this is the first study showing the adequacy of a bifactor-ESEM model for the constructs of neuroticism, self-esteem, and generalized self-efficacy. We could replicate the adequacy of the bifactor-ESEM model across two measurement waves with German secondary school students emphasizing the robustness of the findings. Bifactor-ESEM models have gained prominence in the last few years and have been successfully applied to several other constructs such as academic self-concept (Arens et al., 2021), academic motivation (Litalien et al., 2017), value beliefs (Fadda et al., 2020), or well-being (Fadda et al., 2017). In this study, we show that bifactor-ESEM models are also a useful methodological approach to model the three constructs of neuroticism, self-esteem, and generalized self-efficacy. This modelling approach might present a promising way to reconcile the opposing views on neuroticism, self-esteem, and generalized self-efficacy according to which some researchers emphasize the distinctiveness of the constructs (Asendorpf & van Aken, 2003; Chen et al., 2004; Hufer-Thamm & Riemann, 2021) whereas others emphasize their common core (Judge et al., 2002). The G-factor reflects the assumed common core of neuroticism, self-esteem, and generalized self-efficacy, while the assumption of the distinctiveness of the constructs "survives" in the S-factors. The bifactor-ESEM models thus seems to represent a good compromise between adherents advocating a "unified view" (emphasizing the common core) and adherents advocating a "distinct view" (emphasizing the separateness) of the constructs of neuroticism, self-esteem, and generalized self-efficacy.

A vital question relates to the meaning of the G-factor. To get a presumptive answer to this question, one has to ask about the common characteristics of neuroticism (with emotional stability as its opposite), self-esteem, and generalized self-efficacy. In this study, all items measuring neuroticism, self-esteem, and generalized self-efficacy showed similarly high loadings on the G-factor. Hence, the meaning of the G-factor seems to be a blending of all three constructs. Low levels of neuroticism, high levels of self-esteem, and high

levels of generalized self-efficacy reflect a positive and resilient perception of the self and one's life. Judge et al. (2002) discusses the meaning of the common core of the three constructs as represented in the higher-order factor as a "general neuroticism factor". This "general neuroticism factor" is evocative of other constructs such as dispositional optimism (Carver & Scheier, 2014), resilience (Olsson et al., 2003), or action orientation (Kuhl, 1981). The significance of the G-factor regarding its broad content coverage is also reflected in the association between the G-factor and the two outcome variables of general life satisfaction and school satisfaction. Future studies are necessary to explore the relation between the G-factor and other related constructs and outcome variables including their differentiation and overlap.

A parallel vital question relates to the meaning of the S-factors. The factor loading pattern of the bifactor-ESEM model supports the tenability of the S-factors for the separate constructs of neuroticism, self-esteem, and generalized self-efficacy. Hence, these factors seem to retain some meaning over and above the "general neuroticism factor" as reflected in the G-factor. In the bifactor ESEM model, the S-factors capture the residual variance not explained by the G-factor and the other S-factors (as item cross-loadings are allowed). That is, the S-factors provide information about students' profiles in neuroticism, self-esteem, and generalized self-efficacy, while keeping constant the level of the G-factor and controlling for conceptual overlap between the three constructs. The three S-factors for neuroticism, self-esteem, and generalized self-efficacy displayed correlations as expected with general life satisfaction at t1 and t2 (except for a significant relation between generalized self-efficacy and general life satisfaction at t1). However, the three S-factors showed no significant relations with school satisfaction (except for a significant positive relation with the S-factor of self-esteem at t2 and an unexpected positive correlation with neuroticism at t1). This might be due to global nature of general life satisfaction versus the domain-specific nature of school satisfaction as outlined above. Still, future studies including a broader range of outcome criteria and related constructs are needed to further examine the validity of the S-factors and to preclude the mere reflexion of method artifacts in the S-factors.

The present study provides important new insights into the structure underlying neuroticism, self-esteem, and generalized self-efficacy and may help advance theoretical models of the relation, distinctiveness, and overlap of the three constructs. In terms of practical implications, the present study might help establish interventions to enhance self-esteem and generalized self-efficacy and to mitigate neuroticism. Such interventions might learn from this study that, given the close relations among neuroticism, self-esteem, and generalized self-efficacy, interventions may be more effective if they address all three constructs simultaneously rather than one construct separately.

Despite these strengths and contributions of the present study, there are limitations that should be addressed in further studies. All instruments used were self-reports which might be biased by response styles such as acquiescent or extreme responding (for an overview see Paulhus & Vazire, 2007). Furthermore, only using self-reports might lead to method artifacts in terms of shared method variance in the G-factor and S-factors in the bifactor-ESEM model. It should thus be tested whether the findings of this study can be replicated when the constructs considered in the study are assessed from multiple sources not exclusively relying on self-reports. Moreover, we used a convenience sample of secondary school students attending Grades 11 and 13 of the academic track. This sample might be positively biased in terms of cognitive ability and socioeconomic status (Becker et al., 2012). In addition, important information on the sample such as mental health status or family composition were missing. Hence, broader and more representative samples also including samples from other countries, cultures and age groups should be used in further studies to probe for the generalizability of the findings. Generalizability should also be examined regarding other instruments used for measuring the constructs considered. For instance, the four items used to measure neuroticism in this study might not tap the construct of neuroticism in total. For instance, Costa et al. (1991) differentiate between six subfacets of neuroticism, namely anxiety, hostility, depression, self-consciousness, impulsiveness, and vulnerability. The Cybernetic Big Five Theory as proposed by DeYoung (2015) proposes withdrawal and volatility as two subcomponents of neuroticism. Hence, it should be scrutinized whether the findings of the present study apply to different conceptualizations of the broad construct of neuroticism.

The present study largely relies on the work by Judge et al. (2002). Judge et al. (2002) considered four constructs, namely neuroticism, self-esteem, generalized self-efficacy, and locus of control. The data used in the present study did not include locus of control and we could thus only consider the three constructs of neuroticism, self-esteem, and generalized self-efficacy. The construct of locus of control seems to differ somehow from neuroticism, self-esteem, and generalized self-efficacy as "(…) evidence regarding locus of control was the weakest", because locus of control showed relatively smaller interrelations to the other constructs of neuroticism, generalized self-efficacy, and self-esteem and to outcome relations (Judge et al., 2002, p. 707). Hence, by measuring neuroticism, generalized self-efficacy, and self-esteem in our study, we might have assessed the three constructs for which the common core is most evident. Still, we

concur that more research on the relation between locus of control and neuroticism, generalized self-efficacy, and self-esteem is necessary. Given the findings of the present study, respective research should be completed using ESEM, CFA-bifactor, and bifactor-ESEM models.

Conclusion

This study provides good support for the adequacy of a bifactor-ESEM solution for modeling neuroticism, self-esteem, and generalized self-efficacy. The bifactor-ESEM model integrates opposing points of views that neuroticism, self-esteem, and generalized self-efficacy are distinct and that the three constructs have a common core. Further research is needed regarding the generalizability of our findings, the unsystematic pattern of outcome relations at t2, the inclusion of other constructs such as locus of control, and the meaning of the G-factor.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12144-025-07984-9.

Author contributions A. Katrin Arens: Conceptualization, Formal analysis, Writing—Original Draft.
Franzis Preckel: Writing—Review & Editing, Funding acquisition, Data Curation.

Funding Open Access funding enabled and organized by Projekt DEAL.
The study was funded by the Ministry of Education, Science, Adolescence, and Culture of Rhineland-Palatinate, Germany. Preparation of this article was supported by a Heisenberg fellowship grant from the German Research Foundation to A.K. Arens (AR 877/3–1).

Data availability All variables and syntaxes for the analyses are openly available in the Open Science.
Framework under https://osf.io/26zyv/?view_only=7a6d1d248cc74bb58772b65ca45905c1.
In addition, we provide the correlation matrices for the variables to reproduce our findings. Due to data protection requirements, we cannot offer the full data. The study was not preregistered.

Declarations

Ethics approval The study was approved by the Supervision and Services Directorate of Rhineland-Palatinate ("Aufsichts- und Dienstleistungsdirektion"; protocol number: 32–03 405/29/05).

Informed consent The parents of all participating students gave their written informed consent to participate in the study. Participation was voluntary for students.

Conflict of interest The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Arens, A. K., & Niepel, C. (2019). School attitude and perceived teacher acceptance: Developmental trajectories, temporal relations, and gender differences. *British Journal of Educational Psychology, 89*(4), 689–706. https://doi.org/10.1111/bjep.12252

Arens, A. K., Jansen, M., Preckel, F., Schmidt, I., & Brunner, M. (2021). The structure of academic self-concept: A methodological review and empirical illustration of central models. *Review of Educational Research, 91*(1), 34–72. https://doi.org/10.3102/0034654320972186

Asendorpf, J. B., & van Aken, M. A. G. (2003). Personality-relationship transaction in adolescence: Core versus surface personality characteristics. *Journal of Personality, 71*(4), 629–666. https://doi.org/10.1111/1467-6494.7104005

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 16*(3), 397–438. https://doi.org/10.1080/10705510903008204

Asparouhov, T., Muthén, B., & Morin, A. J. S. (2012). Bayesian structural equation modeling with cross-loadings and residual covariances. *Journal of Management, 41*(6), 1561–1577. https://doi.org/10.1177/0149206315591075

Azizli, N., Atkinson, B. E., Baughman, H. M., & Giammarco, E. A. (2015). Relationships between general self-efficacy, planning for the future, and life satisfaction. *Personality and Individual Differences, 82*, 58–60. https://doi.org/10.1016/j.paid.2015.03.006

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Prentice-Hall, Inc.

Barańczuk, U. (2021). The Five-Factor Model of personality and generalized self efficacy: A meta-analysis. *Journal of Individual Differences, 42*(4), 183–193. https://doi.org/10.1027/1614-0001/a000345

Becker, M., Lüdtke, O., Trautwein, U., Köller, O., & Baumert, J. (2012). The differential effects of school tracking on psychometric intelligence: Do academic-track schools make students smarter? *Journal of Educational Psychology, 104*(3), 682–699. https://doi.org/10.1037/a0027608

Becker, M., Neumann, M., & Dumont, H. (2016). Recent developments in school tracking practices in Germany: An overview and outlook on future trends. *Orbis Scholae, 10*(3), 9–25. https://doi.org/10.14712/23363177.2017.8

Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford.

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). Sage.

Brunner, M., Nagy, G., & Wilhelm, O. (2012). A tutorial on hierarchically structured constructs. *Journal of Personality, 80*(4), 796–846. https://doi.org/10.1111/j.1467-6494.2011.00749.x

Caprara, G. V., & Steca, P. (2005). Affective and social self-regulatory efficacy beliefs as determinants of positive thinking and happiness. *European Psychologist, 10*(4), 275–286. https://doi.org/10.1027/1016-9040.10.4.275

Carver, C. S., & Scheier, M. F. (2014). Dispositional optimism. *Trends in Cognitive Sciences, 18*(6), 293–299. https://doi.org/10.1016/j.tics.2014.02.003

Chen, G., Gully, S. M., & Eden, D. (2004). General self-efficacy and self-esteem: Toward theoretical and empirical distinction between correlated self-evaluations. *Journal of Organizational Behavior, 25*(3), 375–395. https://doi.org/10.1002/job.251

Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research, 41*(2), 189–225. https://doi.org/10.1207/s15327906mbr4102_5

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

Costa, P. T., McCrae, R. R., & Dye, D. A. (1991). Facet scales for agreeableness and conscientiousness: A revision of the NEO Personality Inventory. *Personality and Individual Differences, 12*(9), 887–898. https://doi.org/10.1016/0191-8869(91)90177-D

Daily, S. M., Smith, M. L., Lilly, C. L., Davidov, D. M., Mann, M. J., & Kristjansson, A. L. (2020). Using school climate to improve attendance and grades: Understanding the importance of school satisfaction among middle and high school students. *Journal of School Health, 90*(9), 683–693. https://doi.org/10.1111/josh.12929

DeSantis King, A. L., Huebner, S., Suldo, S. M., & Valois, R. F. (2006). An ecological view of school satisfaction in adolescence: Linkages between social support and behavior problems. *Applied Research in Quality of Life, 1*(3–4), 279–295. https://doi.org/10.1007/s11482-007-9021-7

DeYoung, C. G. (2015). Cybernetic Big Five Theory. *Journal of Research in Personality, 56*, 33–58. https://doi.org/10.1016/j.jrp.2014.07.004

Diamantopoulos, A., & Siguaw, J. A. (2000). Introducing LISREL. *Sage Publications*. https://doi.org/10.4135/9781849209359

Diener, E., & Diener, M. (1995). Cross-cultural correlates of life satisfaction and self-esteem. *Journal of Personality and Social Psychology, 68*(4), 653–663. https://doi.org/10.1037/0022-3514.68.4.653

Diseth, Å., Meland, E., & Breidablik, H. J. (2014). Self-beliefs among students: Grade level and gender differences in self-esteem, self-efficacy and implicit theories of intelligence. *Learning and Individual Differences, 35*, 1–8. https://doi.org/10.1016/j.lindif.2014.06.003

Donnellan, M. B., Trzesniewski, K. H., & Robins, R. W. (2011). Self-esteem: Enduring issuesand controversies. In T. Chamorro-Premuzic, S. von Stumm, & A. Furnham (Eds.), *The Wiley-Blackwell handbook of individual differences* (pp. 718–746). Wiley-Blackwell.

Eder, F. (1998). *Linzer Fragebogen zum Schul- und Klassenklima für die 8. bis 13. Klasse (LFSK 8–13)* [Linzer Questionnaire of School- and Classroom Climate for Grades 8 to 13]. Hogrefe.

Enders, C. K. (2010). *Applied missing data analysis*. Guilford.

Entringer, T., Griese, F., Zimmermann, S., & Richter, D. (2022). SOEP scales manual (updated for SOEP-Core v37). *SOEP Survey Papers 1217 Series C. Berlin.* DIW/SOEP. https://www.diw.de/documents/publikationen/73/diw_01.c.862242.de/diw_ssp1217.pdf

Fadda, D., Scalas, L. F., Meleddu, M., & Morin, A. J. S. (2017). A bifactor-ESEM representation of the Questionnaire for Eudaimonic Wellbeing. *Personality and Individual Differences, 116*, 216–222. https://doi.org/10.1016/j.paid.2017.04.06

Fadda, D., Scalas, L. F., Morin, A. J. S., Marsh, H. W., & Gaspard, H. (2020). Value beliefs about math: A bifactor-ESEM representation. *European Journal of Psychological Assessment, 36*(2), 259–268. https://doi.org/10.1027/1015-5759/a000513

Ferguson, Y. L., Kasser, T., & Jahng, S. (2011). Differences in life satisfaction and school satisfaction among adolescents from three nations: The role of perceived autonomy support. *Journal of Research on Adolescence, 21*(3), 649–661. https://doi.org/10.1111/j.1532-7795.2010.00698.x

Gale, C. R., Booth, T., Mõttus, R., Kuh, D., & Deary, I. J. (2013). Neuroticism and Extraversion in youth predict mental wellbeing and life satisfaction 40 years later. *Journal of Research in Personality, 47*(6), 687–697. https://doi.org/10.1016/j.jrp.2013.06.005

Gignac, G. E. (2016). The higher-order model imposes a proportionality constraint: That is why the bifactor model tends to fit better. *Intelligence, 55*, 57–68. https://doi.org/10.1016/j.intell.2016.01.006

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60*, 549–576. https://doi.org/10.1146/annurev.psych.58.110405.085530

Grant, S., Langan-Fox, J., & Anglim, J. (2009). The big five traits as predictors of subjective and psychological well-being. *Psychological Reports, 105*(1), 205–231. https://doi.org/10.2466/pr0.105.1.205-231

Hankin, B. L., Lakdawalla, Z., Carter, I. L., Abela, J. R. Z., & Adams, P. (2007). Are neuroticism, cognitive vulnerabilities and self-esteem overlapping or distinct risks for depression? Evidence from confirmatory factor analyses. *Journal of Social and Clinical Psychology, 26*(1), 29–63. https://doi.org/10.1521/jscp.2007.26.1.29

Harter, S. (2006). The development of self-esteem. In M. H. Kernis (Ed.), *Self-esteem issues and answers: A sourcebook of current perspectives* (pp. 144–150). Psychology Press.

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

Huebner, E. S., & McCullough, G. (2000). Correlates of school satisfaction among adolescents. *The Journal of Educational Research, 93*(5), 331–335. https://doi.org/10.1080/00220670009598725

Hufer-Thamm, A., & Riemann, R. (2021). On the link of self-esteem, life satisfaction, and Neuroticism. *Journal of Personality, 89*(5), 998–1011. https://doi.org/10.1111/jopy.12632

Hui, E. K. P., & Sun, R. C. F. (2010). Chinese children's perceived school satisfaction: The role of contextual and intrapersonal factors. *Educational Psychology, 30*(2), 155–172. https://doi.org/10.1080/01443410903494452

Judge, T. A., & Bono, J. E. (2001). Relationship of core self-evaluations traits—self-esteem, generalized self-efficacy, locus of control, and emotional stability—with job satisfaction and job performance: A meta-analysis. *Journal of Applied Psychology, 86*, 80–92. https://doi.org/10.1037/0021-9010.86.1.80

Judge, T. A., Locke, E. A., & Durham, C. C. (1997). The dispositional causes of job satisfaction: A core evaluations approach. *Research in Organizational Behavior, 19*, 151–188.

Judge, T. A., Locke, E. A., Durham, C. C., & Kluger, A. N. (1998). Dispositional effects on job and life satisfaction: The role of core evaluations. *Journal of Applied Psychology, 83*(1), 17–34. https://doi.org/10.1037/0021-9010.83.1.17

Judge, T. A., Erez, A., Bono, J. E., & Thoresen, C. J. (2002). Are measures of self-esteem, neuroticism, locus of control, and generalized self-efficacy indicators of a common core construct? *Journal of Personality and Social Psychology, 83*(3), 693–710. https://doi.org/10.1037/0022-3514.83.3.693

Katsantonis, I., McLellan, R., & Marquez, J. (2023). Development of subjective well-being and its relationship with self-esteem in

early adolescence. *British Journal of Developmental Psychology, 41*(2), 157–171. https://doi.org/10.1111/bjdp.12436

Kotov, R., Gamez, W., Schmidt, F., & Watson, D. (2010). Linking "big" personality traits to anxiety, depressive, and substance use disorders: A meta-analysis. *Psychological Bulletin, 136*(5), 768–821. https://doi.org/10.1037/a0020327

Kuhl, J. (1981). Motivational and functional helplessness: The moderating effect of state versus action orientation. *Journal of Personality and Social Psychology, 40*(1), 155–170. https://doi.org/10.1037/0022-3514.40.1.155

Lei, P.-W., & Shiverdecker, L. K. (2020). Performance of estimators for confirmatory factor analysis of ordinal variables with missing data. *Structural Equation Modeling: A Multidisciplinary Journal, 27*(4), 584–601. https://doi.org/10.1080/10705511.2019.1680292

Litalien, D., Morin, A. J. S., Gagné, M., Vallerand, R. J., Losier, G. F., & Ryan, R. M. (2017). Evidence of a continuum structure of academic self-determination: A two-study test using a bifactor-ESEM representation of academic motivation. *Contemporary Educational Psychology, 51*, 67–82. https://doi.org/10.1016/j.cedpsych.2017.06.010

Luszczynska, A., Scholz, U., & Schwarzer, R. (2005). The general self-efficacy scale: Multicultural validation studies. *The Journal of Psychology: Interdisciplinary and Applied, 139*(5), 439–457. https://doi.org/10.3200/JRLP.139.5.439-457

Marcionetti, J., & Rossier, J. (2016). Global life satisfaction in adolescence: The role of personality traits, self-esteem, and self-efficacy. *Journal of Individual Differences, 37*(3), 135–144. https://doi.org/10.1027/1614-0001/a000198

Marcionetti, J., & Rossier, J. (2021). A longitudinal study of relations among adolescents' self-esteem, general self-efficacy, career adaptability, and life satisfaction. *Journal of Career Development, 48*(4), 475–490. https://doi.org/10.1177/0894845319861691

Marsh, H. W., & Hau, K.-T. (1996). Assessing goodness of fit: Is parsimony always desirable? *Journal of Experimental Education, 64*(4), 364–390. https://doi.org/10.1080/00220973.1996.10806604

Marsh, H. W., Liem, G. A. D., Martin, A. J., Morin, A. J. S., & Nagengast, B. (2011). Methodological measurement fruitfulness of exploratory structural equation modeling (ESEM): New approaches to key substantive issues in motivation and engagement. *Journal of Psychoeducational Assessment, 29*(4), 322–346. https://doi.org/10.1177/0734282911406657

Marsh, H. W., Morin, A. J. S., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology, 10*, 85–110. https://doi.org/10.1146/annurev-clinpsy-032813-153700

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Public Health* Resources. *556*. https://digitalcommons.unl.edu/publichealthresources/556

McCrae, R. R., & Costa, P. T., Jr. (1997). Personality trait structure as a human universal. *American Psychologist, 52*(5), 509–516. https://doi.org/10.1037/0003-066x.52.5.509

Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. Routledge.

Morin, A. J. S., Arens, A. K., & Marsh, H. W. (2016). A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Structural Equation Modeling: A Multidisciplinary Journal, 23*(1), 116–139. https://doi.org/10.1080/10705511.2014.961800

Morin, A. J. S., Myers, N. D., & Lee, S. (2020). Modern factor analytic techniques: Bifactor models, exploratory structural equation modeling (ESEM) and bifactor-ESEM. In G. Tenenbaum & R. C. Eklund (Eds.), *Handbook of Sport Psychology* (4th ed., pp. 1044–1073). Wiley.

Muris, P., & Otgaar, H. (2023). Self-esteem and self-compassion: A narrative review and meta-analysis on their links to psychological problems and well-being. *Psychology Research and Behavior Management, 16*, 2961–2975. https://doi.org/10.2147/PRBM.S402455

Muthén, L. K., & Muthén, B. O. (1998–2019). *Mplus User's Guide. Eighth Edition*. Muthén & Muthén.

Olsson, C. A., Bond, L., Burns, J. M., Vella-Brodrick, D. A., & Sawyer, S. M. (2003). Adolescent resilience: A concept analysis. *Journal of Adolescence, 26*(1), 1–11. https://doi.org/10.1016/S0140-1971(02)00118-5

Orth, U., & Robins, R. W. (2022). Is high self-esteem beneficial? *Revisiting a Classic Question. American Psychologist, 77*(1), 5–17. https://doi.org/10.1037/amp0000922

Orth, U., Robins, R. W., & Widaman, K. F. (2012). Life-span development of self-esteem and its effects on important life outcomes. *Journal of Personality and Social Psychology, 102*(6), 1271–1288. https://doi.org/10.1037/a0025558

Paulhus, D. L., & Vazire, S. (2007). The self-report method. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 224–239). The Guilford Press.

Pavot, W., & Diener, E. (2008). The Satisfaction With Life Scale and the emerging construct of life satisfaction. *The Journal of Positive Psychology, 3*(2), 137–152. https://doi.org/10.1080/17439760701756946

Preckel, F. & Schmidt, I. (2014). *Projektbericht 2013: Wissenschaftliche Begleitung der 5. bis 8. Klassen der gymnasialen Hochbegabtenförderung in Rheinland-Pfalz, Zeitraum 2005-2013*. Unveröffentlichter Bericht. Universität Trier. http://www.unitrier.de/fileadmin/fb1/prof/PSY/HBF/Sachbericht_2013_Klassen5-8_final_12-11-2014.pdf

Rammstedt, B., & John, O. P. (2005). Kurzversion des Big Five Inventory (BFI-K): Entwicklung und Validierung eines ökonomischen Inventars zur Erfassung der fünf Faktoren der Persönlichkeit [Short version of the Big Five Inventory (BFI-K): Development and validation of an economic inventory for assessment of the five factors of personality]. *Diagnostica, 51*(4), 195–206. https://doi.org/10.1026/0012-1924.51.4.195

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment, 92*(6), 544–559. https://doi.org/10.1080/00223891.2010.496477

Rindskopf, D., & Rose, T. (1988). Some theory and applications of confirmatory second-order factor analysis. *Multivariate Behavioral Research, 23*(1), 51–67. https://doi.org/10.1207/s15327906mbr2301_3

Robins, R. W., Tracy, J. L., Trzesniewski, K., Potter, J., & Gosling, S. D. (2001). Personality correlates of self-esteem. *Journal of Research in Personality, 35*(4), 463–482. https://doi.org/10.1006/jrpe.2001.2324

Rosenberg, M. (1979). *Conceiving the self*. Krieger.

Schwarzer, R., & Jerusalem, M. (Eds.). (1999). *Skalen zur Erfassung von Lehrer- und Schülermerkmalen. Dokumentation der psychometrischen Verfahren im Rahmen der Wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen*. Freie Universität Berlin.

Steel, P., Schmidt, J., & Shultz, J. (2008). Refining the relationship between personality and subjective well-being. *Psychological Bulletin, 134*(1), 138–161. https://doi.org/10.1037/0033-2909.134.1.138

Strobel, M., Tumasjan, A., & Spörrle, M. (2011). Be yourself, believe in yourself, and be happy: Self-efficacy as a mediator between

personality factors and subjective well-being. *Scandinavian Journal of Psychology, 52*(1), 43–48. https://doi.org/10.1111/j.1467-9450.2010.00826.x

Suldo, S. M., Riley, K. N., & Shaffer, E. J. (2006). Academic correlates of children and adolescents' life satisfaction. *School Psychology International, 27*(5), 567–582. https://doi.org/10.1177/0143034306073411

Swami, V., Maïano, C., & Morin, A. J. S. (2023). A guide to exploratory structural equation modelling (ESEM) and bifactor-ESEM in body image research. *Body Image, 41*, 101641. https://doi.org/10.1016/j.bodyim.2023.101641

Von Collani, G., & Herzberg, P. Y. (2003). Eine revidierte Fassung der deutschsprachigen Skala zum Selbstwertgefühl von Rosenberg [A revised version of the German adaptation of Rosenberg's Self-Esteem Scale]. *Zeitschrift Für Differentielle und Diagnostische Psychologie, 24*(1), 3–7. https://doi.org/10.1024/0170-1789.24.1.3

Weidmann, R., Ledermann, T., Robins, R. W., Gomez, V., & Grob, A. (2018). The reciprocal link between the Big Five traits and self-esteem: Longitudinal associations within and between parents and their offspring. *Journal of Research in Personality, 74*, 166–181. https://doi.org/10.1016/j.jrp.2018.04.003

Wood, R., & Bandura, A. (1989). Impact of conceptions of ability on self-regulatory mechanisms and complex decision making. *Journal of Personality and Social Psychology, 56*(3), 407–415. https://doi.org/10.1037/0022-3514.56.3.407

Zullig, K. J., Huebner, E. S., & Patton, J. M. (2011). Relationships among school climate domains and school satisfaction. *Psychology in the Schools, 48*(2), 133–145. https://doi.org/10.1002/pits.20532