

Secondary Publication



Gnambs, Timo; Schroeders, Ulrich

Reliability and Factorial Validity of the Core Self-Evaluations Scale : A Meta-Analytic Investigation of Wording Effects

Date of secondary publication: 19.12.2024

Version of Record (Published Version), Article

Persistent identifier: urn:nbn:de:bvb:473-irb-1055106

Primary publication

Gnambs, Timo; Schroeders, Ulrich (2024): Reliability and Factorial Validity of the Core Self-Evaluations Scale : A Meta-Analytic Investigation of Wording Effects, in: European journal of psychological assessment : official organ of the European Association of Psychological Assessment, Göttingen ; Cambridge, Mass.: Hogrefe, Vol. 40, Nr. 5, pp. 343–359, doi: 10.1027/1015-5759/a000783.

Legal Notice

This work is protected by copyright and/or the indication of a licence. You are free to use this work in any way permitted by the copyright and/or the licence that applies to your usage. For other uses, you must obtain permission from the rights-holders.

This document is made available under a Creative Commons license.



The license information is available online:

<https://creativecommons.org/licenses/by/4.0/legalcode>



Reliability and Factorial Validity of the Core Self-Evaluations Scale

A Meta-Analytic Investigation of Wording Effects

Timo Gnams¹  and Ulrich Schroeders² 

¹Leibniz Institute for Educational Trajectories, Bamberg, Germany

²Institute of Psychology, University of Kassel, Germany

Abstract: The *Core Self-Evaluations Scale* (CSES) measures a broad personality trait reflecting individuals' self-appraisals of their worth, capabilities, and control of their lives. Although the CSES was designed to capture a single trait, factor analytic studies often found more complex measurement structures. These either referred to different content facets or methodological artifacts due to the item wording. The present random-effects meta-analysis summarized correlation matrices from 53 samples including 31,843 respondents. After accounting for acquiescent responding, meta-analytic confirmatory factor analyses revealed a single common factor for all items. The factor was highly reliable ($\omega = .87$) and demonstrated partial metric measurement invariance across English, German, and Spanish language versions as well as cultural tendencies of individualism and flexibility. However, Chinese and Romanian translations exhibited substantially lower factor loadings. These results corroborate the use of the CSES as a unidimensional measure, albeit systematic investigations of measurement invariance are recommended before its use in cross-cultural research.

Keywords: meta-analysis, structural equation modeling, wording effect, self-evaluation, cross-cultural psychology



The *Core Self-Evaluations Scale* (CSES; Judge et al., 2003) is a popular self-report measuring people's fundamental evaluations of themselves in terms of their psychological resources and self-worth. These core self-evaluations (CSE) characterize broad and enduring dispositions of individuals that not only shape more specific evaluations (e.g., job satisfaction) but also specific behaviors (e.g., job performance). It is associated with various organizational and work-related outcomes such as higher job commitment and salary, as well as lower psychological strain and turnover intentions (Chang et al., 2012). CSE is typically viewed as a meta-trait that captures the shared variance between several established personality traits (e.g., Johnson et al., 2008; Judge et al., 1998). Most often it is conceptualized as a higher-order construct indicated by global self-esteem, generalized self-efficacy, internal locus of control, and emotional stability (Judge et al., 2003). Accordingly, CSE represents a blend of various positively valenced dispositions that reflect persons' beliefs about their worth and self-regard, their abilities to achieve goals, their beliefs they are in

charge of their lives, and their ability to remain stable and even-tempered in the face of adversities. Indirect measurements often rely on established instruments for the four core traits to model CSE as a second-order factor (e.g., Gardner & Pierce, 2010; Johnson et al., 2008). However, this approach is rather cumbersome for applied research because it requires administering four separate (and often rather long) instruments to capture a single construct. Therefore, the CSES (Judge et al., 2003) was developed as an economic alternative. Because of its brevity and good criterion validity (e.g., Gardner & Pierce, 2010; Zenger et al., 2015), the CSES has become a widely used measure in work and organizational psychology (e.g., Zacher et al., 2021) but also other disciplines such as clinical and health psychology (e.g., Geuens et al., 2020) or quality of life research (e.g., Turska & Stępień-Lampa, 2021).

Although developed to represent a single latent trait, factor analytic investigations of the CSES often favored multi-as compared to unidimensional measurement models (e.g., Arias et al., 2022; Henderson & Gardiner, 2019; Mäkikangas et al., 2018; Schmalbach et al., 2021; Sun & Jiang, 2017; Zenger et al., 2015). These results have been interpreted either from a substantive point of view as reflecting different content facets of CSE (Mäkikangas et al., 2018; Zenger et al., 2015) or as the result of methodological artifacts stemming from the item wording (Arias et al., 2022;

Schmalbach et al., 2021). The present study contributes to this debate by presenting meta-analytic evidence on the psychometric properties of the CSES. To this end, we make use of recent methodological advancements in meta-analytic structural equation modeling (Cheung & Chan, 2005; Jak & Cheung, 2020) to clarify the dimensional structure of the CSES across diverse samples and settings. Furthermore, measurement invariance is examined across different language versions and cultural dimensions to evaluate the applicability of the CSES for comparative research in a cross-cultural context.

The Core Self-Evaluations Scale

The 12 items of the CSES were developed to cover the four individual core traits – global self-esteem, generalized self-efficacy, emotional stability, and internal locus of control (Judge et al., 2003). However, rather than being pure indicators of these traits, the items were written to optimally capture the paramount construct of CSE. Therefore, many items reflect a blend of two or more core traits. Despite the heterogeneity of these items, the ordinary sum score is typically used as a personal estimate of CSE. Extensive research on the CSES attested to its good reliability with an average coefficient alpha of .84 (Ock et al., 2021) and usefulness for predicting various outcomes such as psychological and physical health (Turska & Stępień-Lampa, 2021) or income and number of promotions (Stumpp et al., 2010). Despite the substantial evidence for the validity of the sum score, the internal structure of the CSES is still disputed (e.g., Arias et al., 2022; Gu et al., 2015; Henderson & Gardiner, 2019; Mäkikangas et al., 2018; Schmalbach et al., 2021; Sun & Jiang, 2017; Zenger et al., 2015). In line with its original conception, Judge and colleagues (2003) seem to provide evidence that a unidimensional confirmatory factor model had the best fit for the CSES as compared to more complex measurement models. However, a closer inspection of the reported degrees of freedom indicates that the authors seem to have included six undisclosed correlated error terms. Although such a model can attest to a common factor across all items, the correlated errors also indicate item dependencies that might arise from unmodeled additional traits or method variance.

The data-driven approach of additionally specifying residual correlations has also been taken up in other factor-analytic research that seemingly demonstrated the unidimensionality of the CSES in different languages (Heilmann & Jonas, 2010; Judge et al., 2004; Stumpp et al., 2010), although the number of residual correlations differed vastly from 4 to 10. Thus, the assumption of strict unidimensionality of the CSES has to be taken with a grain of salt and could only be maintained taking into account systematic covariations between some item pairs that do

not generalize across different samples. More rigorous tests of the CSES's dimensionality, however, often found more complex measurement structures that either distinguished different content facets to capture qualitatively different types of CSE or accounted for method artifacts caused by item wording or inattentive responding.

Substantive Content Facets of the Core Self-Evaluations Scale

The four different core traits of the CSES can be rarely empirically recovered in factor analytic research (Judge et al., 2003; Zenger et al., 2015). Confirmatory factor analyses that modeled four correlated factors typically do not show an improved fit as compared to more parsimonious models. However, Ferris and colleagues (2011) pointed out a conceptual fuzziness: the CSE has been either viewed as an indicator of high approach temperament or as an indicator of low avoidance temperament. Put differently, CSE has been simultaneously described as an individual's tendency to seek out positive outcomes (e.g., higher wages) and an individual's orientation towards averting negative outcomes (e.g., lay-offs). Because temperaments of approach and avoidance are assumed to independently influence personality traits (Elliot & Thrash, 2002), some traits can be seen as indicators of approach temperaments (e.g., general self-efficacy) and others as indicators of avoidance temperaments (e.g., neuroticism). Accordingly, Ferris and colleagues (2011) showed that both temperaments were associated with the sum score of the CSES and mediated the effect of CSE on job performance. Therefore, the CSES might reflect two qualitatively different facets of CSE that refer to these approaches and avoidance tendencies (Sun & Jiang, 2017; Zenger et al., 2015). Indeed, factor analytic studies often found substantial evidence for two factors underlying the CSES that have been interpreted as positive and negative CSE (see Model 2 in Figure 1). However, the loading structure did not always fully replicate. For example, Item 10 ("I do not feel in control of my success in my life.") of the CSES sometimes exhibited substantial cross-loadings on both factors and, thus, could not be clearly assigned to either factor (Mäkikangas et al., 2018). Moreover, the substantial factor correlations between positive and negative CSE ranging from .55 to .67 suggest that a second-order factor might be a reasonable representation of general CSE in line with the original conception of the CSES (Judge et al., 2003).

Method Artifacts and Wording Effects

Because the content facets of positive and negative CSE perfectly align with the wording of the items, it has been suggested that the CSES does not represent substantively

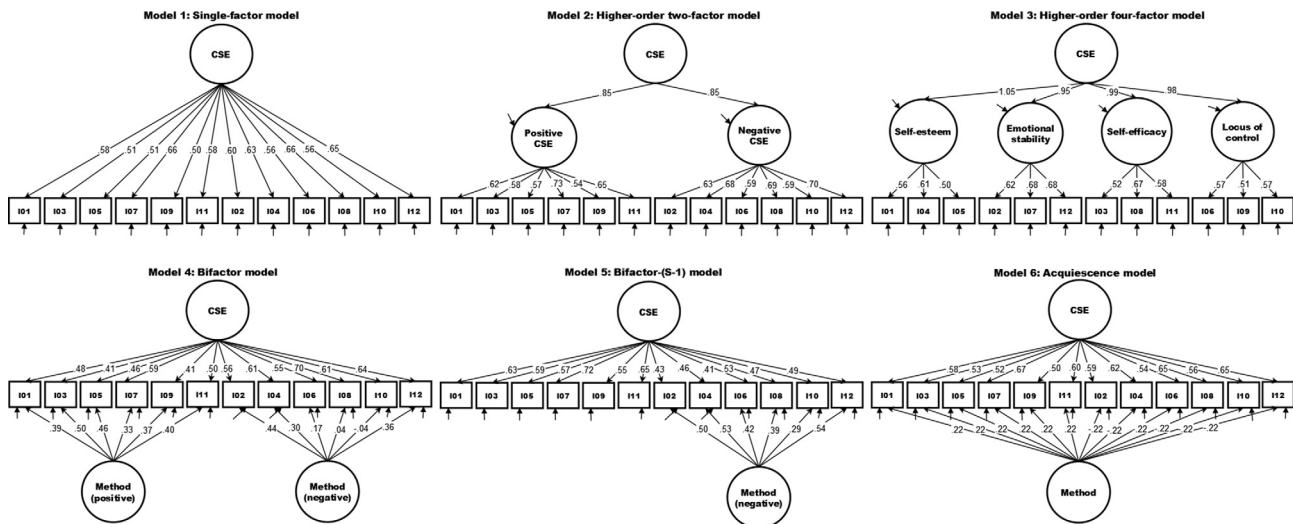


Figure 1. Factor models for the core self-evaluations scale. $N = 31,843$ from 53 samples. Presented are standardized factor loadings. CSE = core self-evaluations.

different traits. Rather, its dimensionality is often distorted by artifacts stemming from the use of positively and negatively worded items (e.g., Arias et al., 2022; Gu et al., 2015; Henderson & Gardiner, 2019; Schmalbach et al., 2021). A frequently observed phenomenon in self-report instruments is systematic variance captured by negatively worded items that can present itself as an additional factor beyond the focal construct (e.g., DiStefano & Motl, 2006; Gnamb et al., 2018; Koutsogiorgi et al., 2021). Prevalent explanations suggest that individual differences, for example, in reading competence or general cognitive abilities contribute to this structural ambiguity because negatively worded items require more complex cognitive processing (Gnamb & Schroeders, 2020; Michaelides, 2019). In line with this assumption, bifactor-(S-1) factor models (see Model 5 in Figure 1) often showed superior fit for the CSES as compared to unidimensional measurement models (e.g., Arias & Arias, 2017; Gu et al., 2015).

Another explanation for ostensible failures to corroborate unidimensionality is careless/insufficient effort responding (C/IER; Schroeders et al., 2022). Respondents that do not properly engage with the items might respond differently to positively and negatively worded items which can result in spurious secondary factors. Indeed, screening for and excluding inattentive respondents can substantially alleviate wording effects and lead to an essentially unidimensional CSES (Arias et al., 2022). Instead of excluding C/IER respondents, the influence of inattention or careless reading can also be directly modeled with a bifactor specification that acknowledges item-specific effects for both positively and negatively worded items (see Model 4 in Figure 1). Here, each item is assumed to be affected by additional variance components beyond the focal trait, but differently for the two item types. In practice, these models

often provide a very good fit to the data but tend to result in erratic factor loadings (e.g., sometimes close to zero or even negative) suggesting overparameterized models (see Arias & Arias, 2017; Henderson & Gardiner, 2019, for respective results on the CSES). To avoid anomalous results, Eid and colleagues (2017) proposed two alternative models – the bifactor-(S-1) model and the bifactor-(S-I-1) model – in which either a factor or a single item is set as a reference to instantiate the trait. Both modeling approaches are eligible to overcome irregular loading patterns caused by overparameterization.

Inconsistent responding can also be acknowledged by adding a constant person effect for all items, thus, estimating a second variance component in addition to the focal trait variance (Maydeu-Olivares & Coffman, 2006). This is supposed to capture the effect of systematic response styles such as acquiescence responding resulting from inattention or indifference towards the differently worded items (Aichholzer, 2014). Accordingly, Schmalbach and colleagues (2021) reported that the CSES is essentially unidimensional as soon as a rather small amount of variance related to a constant person effect is accounted for.

Present Meta-Analysis

Notwithstanding its popularity in applied research, the internal structure of the CSES is still an unresolved matter of discussion. Although the instrument supposedly captures a single latent construct (Judge et al., 2003), a strict interpretation of unidimensionality has been rarely found in factor analytic studies. Rather, various more complex measurement models have been proposed that either argued for different theoretical facets or suggested method artifacts clouding a single factor structure. Therefore, we present a

meta-analytic investigation of the psychometric properties of the CSES to evaluate competing measurement models described in the literature (see Figure 1). Moreover, the measurement precision is studied using model-based indicators of reliability (see Flora, 2020) to extend a recent reliability generalization on coefficient alphas for the CSES (Ock et al., 2021). Finally, we report exploratory analyses of measurement invariance across language versions and cultural dimensions to evaluate the potential limitations of the CSES for cross-cultural research.

Method

Auxiliary information including the code book, a summary of the used statistical software, and the results of supplemental analyses is provided in an open data repository (see Gnambs & Schroeders, 2023). Moreover, the online material also includes the coded data and annotated computer code for all analyses to reproduce the reported findings.

Meta-Analytic Database

Search Strategy

Primary studies and raw data reporting on the CSES were identified in October 2022 and April 2023 in major curated databases (PsycArticles, PsycINFO, PSYINDEX, ERIC, and ProQuest Dissertations & Theses), open data repositories (Open Science Framework, PsychArchives, Harvard Dataverse, Mendeley Data, Figshare, Kaggle, and Google Dataset Search), and journals sharing primary data (*Journal of Open Psychology Data*, *Scientific Data*, *Data in Brief*, *eLife*, *PLoS ONE*)¹ using the Boolean expression “core self-evaluations scale” OR “core self-evaluation scale”. Moreover, documents citing the original publication of the CSES (Judge et al., 2003) were identified in Google Scholar (limited to the initial 1,000 results). In November 2022 we also made an open social media call for unpublished studies including the CSES. Finally, six raw data sets including the CSES were obtained through personal contacts. This resulted in 2,146 potentially relevant sources. After reviewing the titles, abstracts, tables, or raw data, the full text (or raw data) was evaluated in detail for 86 sources. Only those sources were retained that met the following criteria:

(a) The original CSES with 12 items (or a translated version thereof) was administered. Thus, studies using

substantially modified items or excluding items were not considered.

- (b) The items were accompanied by their original 5-point (or more) response scales in order to conduct linear factor analyses for continuous indicators (see Rhemtulla et al., 2012).
- (c) The relevant item-level statistics were available or could be reproduced. This included the raw data for the CSES, the full correlation (or covariance) matrix between the 12 items, or the loading pattern from an exploratory (or confirmatory) factor analysis. Factor loading patterns from oblique factor rotations were only used if also the respective factor correlations were available. Moreover, factor pattern matrices were only considered if they reported at least half of the estimated factor loadings (i.e., studies with an excessive number of missing values were excluded).
- (d) The sample size was reported.
- (e) The sample included primarily healthy individuals without psychological disorders.

No further exclusions were made based on population characteristics, publication year, type of publication (e.g., peer-reviewed or not), or the language of publication. Authors of eligible studies reporting factor loading patterns of the CSES were contacted by email and asked to share the correlation matrix or raw data of their study. Three authors were responsive (Arias et al., 2022; Gu et al., 2015; Zenger et al., 2015). This literature search and screening process (see Supplement A in Gnambs & Schroeders, 2023) resulted in 49 sources with 53 samples that could be included in the meta-analytic database (see Table 1).

Coding Procedure

The relevant information to be collected from each source was described in a coding manual which specified all variables to be coded (see Gnambs & Schroeders, 2023). This included the correlations between the 12 items of the CSES or the respective loadings patterns and factor correlations from exploratory or confirmatory factor analyses. If several factor solutions were reported for the same sample, the factor pattern with the largest number of factors was chosen. In case results were available for the total sample and also different subsamples, only the total sample was considered. Additionally, we also coded information on the sample (i.e., sample size, language, mean age, percentage of women), the publication (i.e., publication year, type of publication), and the reported factor analysis (i.e., factor analytic

¹ Open Science Framework: <https://osf.io>, PsychArchives: <https://www.psycharchives.org>, Harvard Dataverse: <https://dataverse.harvard.edu>, Mendeley Data: <https://data.mendeley.com>, Figshare: <https://figshare.com>, Kaggle: <https://www.kaggle.com/datasets>, Google Dataset Search: <https://datasetsearch.research.google.com>, *Journal of Open Psychology Data*: <https://openpsychologydata.metajnl.com>, *Scientific Data*: <https://www.nature.com/sdata/>, *Data in Brief*: <https://www.sciencedirect.com/journal/data-in-brief>, *eLife*: <https://elifesciences.org>, *PLoS One*: <https://journals.plos.org/plosone/>.

Table 1. Overview of samples and coded data

Study, Year	Country	Lang.	N	Women	Age	Data	IDV	FLX	Std.Res.	Cook's <i>d</i>	Outlier	RoB
Aligner & Lorenz (2022)	Germany	German	612	99	37	RawD	102	46	71	1.62	No	3.50
Arias & Arias (2017)	Chile	Spanish	289	61	25	Load	-8	-153	45	0.78	No	5.00
Arias et al. (2022)	USA	English	694	37	36	RawD	33	11	141*	3.27	Yes	2.50
Arias et al. (2022)	Chile	Spanish	274	55	22	RawD	-8	-153	143*	2.56	Yes	2.50
Boyd (2006)	USA	English	204	43	36	Load	33	11	34	0.50	No	6.00
Coroiu et al. (2018)	Germany	German	2,389	53	50	RawD	102	46	43	1.45	No	1.50
Curşeu et al. (2020)	Romania	Romanian	269	81	21	RawD	-19	-64	58	0.95	No	3.00
Dieckmann & Hartman (2022)	USA	English	791	58	51	RawD	33	11	62	1.56	No	4.50
Ding & Yu (2022)	China	Chinese	238	46	36	RawD	-31	134	450*	7.94*	Yes	5.00
Farčić et al. (2020)	Croatia	Croatian	584	91	39	RawD			203*	4.53	Yes	3.00
Ferreira et al. (2013)	Brazil	Portuguese	393	47	34	Load	-56	-83	79	1.54	No	6.50
Förster et al. (2013)	Germany	German	369	56	38	Corr	102	46	34	0.62	No	4.50
Geuens et al. (2020)	Belgium	Dutch	212	83	39	RawD	110	63	31	0.48	No	5.00
Gu et al. (2015)	China	Chinese	449	50	16	Corr	-31	134	130*	2.64	No	6.00
Gurbuz et al. (2021)	Turkey	Turkish	216	45	22	Load	-18	-52	63	0.95	Yes	6.00
Heller et al. (2023)	Germany/Switzerland	German	122	52	39	RawD			72	0.88	No	4.00
Henderson & Gardiner (2019)	Australia	English	522	72	24	Load	83	41	28	0.59	No	6.00
Henderson & Gardiner (2019)	Australia	English	227	70	26	Load	83	41	16	0.27	No	7.00
Hou et al. (2022)	China	Chinese	117	79	20	RawD	-31	134	59	0.72	No	4.50
Jain & Nair (2019)	India	English	330	12		Load	-101	17	237*	3.45	Yes	7.50
Karasová & Očenášá (2014)	Slovakia	Slovak	182	92	22	Load			141*	2.06	No	7.00
Kim et al. (2015)	South Korea	Korean	471	14		Load	25	174	103*	2.17	No	6.50
Leonhardt et al. (2017)	Germany	German	204	36	45	RawD	102	46	48	0.75	No	5.50
Li et al. (2014)	China	Chinese	1,559	96	30	RawD	-31	134	170*	7.14	Yes	4.00
Littrell et al. (2021)	USA/Canada	English	317	43	38	RawD			52	0.83	No	4.00
Liu et al. (2023)	China	Chinese	265	43	15	RawD	-31	134	93	1.47	No	2.00
Liu et al. (2023)	China	Chinese	518	63	43	RawD	-31	134	113*	2.41	No	4.00
Love (2016)	USA	English	4,908	83	43	Load	33	11	67	2.48	No	5.00
Mäkikangas et al. (2018)	Finland	Finnish	2,137	67	44	Load			74	3.04	No	6.00
Mussel et al. (2023)	Germany	German	214	39	39	RawD	102	46	43	0.67	No	5.00
Nastasa et al. (2021)	Romania	Romanian	186	85	44	RawD	-19	-64	79	1.03	No	4.00
Nurmohamed et al. (2021)	USA	English	220	56	39	RawD	33	11	64	0.86	No	6.00
Pătraşcu-Lungu & Iliescu (2022)	USA	English	297	47	32	RawD	33	11	39	0.65	No	5.00
Pitt-Catsoupes & Smyer (2013)	USA	English	1,774	59	41	RawD	33	11	28	0.81	No	5.50
Rosenbloom & Ash (2013)	USA	English	754	42	39	RawD	33	11	23	0.52	No	5.00
Scheurer (2013)	USA	English	225			Corr	33	11	102*	1.60	Yes	7.00
Smedema et al. (2016)	USA	English	247	50	42	Corr	33	11	28	0.45	No	5.00
Sprung (2021)			412		51	RawD			55	1.07	No	6.00
Sulaiman et al. (2021)	Jordan	Arabic	214	60		RawD			138*	2.16	No	5.00
Sun & Jiang (2017)	China	Chinese	233	58	27	Load	-31	134	174*	2.49	No	7.50
Swab (2021)		English	320	45	22	RawD			50	0.89	No	5.00
Thielmann & Hilbig (2019)	Germany	German	578	65	25	RawD	102	46	59	1.35	No	3.50
Tims & Akkermans (2017)	Netherlands		403	60	26	RawD	182	87	46	0.89	No	5.00
Tims & Akkermans (2017)	Germany		303	63	24	RawD	102	46	44	0.78	No	5.00
Tisu et al. (2020)	Romania	Romanian	353	65	33	RawD	-19	-64	37	0.70	No	5.00
Turska & Stępień-Lampa (2021)	Poland	Polish	1,000	65	22	RawD	-15	9	86	2.38	No	5.00
Virga et al. (2017)	Romania	Romanian	272	39	36	RawD	-19	-64	50	0.86	No	5.00
Virga & Rusu (2018)	Romania	Romanian	216	69	40	RawD	-19	-64	221*	3.20	Yes	5.00
Yaras (2022)	Turkey	Turkish	310			Load	-18	-52	525*	7.01	Yes	8.00

(Continued on next page)

Table 1. (Continued)

Study, Year	Country	Lang.	N	Women	Age	Data	IDV	FLX	Std.Res.	Cook's <i>d</i>	Outlier	RoB
Zacher et al. (2021)	Germany	German	591	37	45	RawD	102	46	128*	2.93	No	3.50
Zaniboni et al. (2021)	Spain	Spanish	190	53	55	RawD	58	2	35	0.54	No	6.00
Zenger et al. (2015)	Germany	German	2,510	53	49	Corr	102	46	56	1.44	No	1.50
Zheng et al. (2021)	China	Chinese	159	38		RawD	-31	134	72	0.94	No	5.50

Note. Lang. = Language version of CSES; Women = Percentage of women; Age = Mean age of sample; Data = Data type for meta-analysis; RawD = Raw data; Corr = Correlations; Load = Loadings; IDV = Score on individualism versus collectivism dimension (Minkov et al., 2017); FLX = Score on flexibility versus monumentalism dimension (Minkov et al., 2018); Std.Res. = Multivariate standardized residual (Viechtbauer & Cheung, 2010); Cook's *d* = Cook's (1977) distance; Outlier = Sample was classified as an outlier in sensitivity analyses; RoB = Risk of bias score with higher values indicating more risk. * $p < .05$ corrected for multiple comparisons (Benjamini & Hochberg, 1995).

method, type of rotation, number of extracted factors). All studies were initially coded by the first author. If raw data were available, the respective information was calculated from the data. To evaluate the quality of the coding process, the second author independently coded all study characteristics a second time. Krippendorff's (2013) alphas for the two codings fell between .92 and 1.00. As values greater than .80 are customarily considered satisfactory, the intercoder agreement in the present study can be considered excellent. Moreover, factor loading patterns and correlation matrices of about a third of the primary studies were also coded twice, yielding a perfect intercoder agreement of 1.00.

The meta-analytic database was expanded with country-level information to study measurement invariance. Information on the cultural background of the examined samples was taken from Minkov and colleagues (2017, 2018). For each sample, we coded the relative standing of each country on the cultural dimensions of individualism versus collectivism and flexibility versus monumentalism. Whereas the first dimension describes the cultural tendency favoring autonomy and freedom versus restrictiveness and conformism, the second dimension describes the tendency towards modesty and adaptability versus grandiosity and stability. These scores were given on a scale with a mean of 0 and a standard deviation of 100.

Evaluation of Risk of Bias

The quality of the available studies was evaluated with eight items from a (slightly adapted) risk of bias scale (Nudelman & Otto, 2020). This quality appraisal instrument was specifically designed for observational studies without interventions. Among others, these items referred to the participant recruitment (i.e., were appropriate methods used to sample respondents), sample size, and data management procedures (i.e., were data cleaned, e.g., regarding invalid responses or outliers). The specific items (including amendments) are given in Supplement B (Gnambs & Schroeders, 2023). The risk of bias was given by the sum score across the eight items with higher values indicating a larger risk. Again, all studies were rated by both authors and yielded

a good interrater reliability of Krippendorff's (2013) alpha of .88, which is why we used the mean scores across both ratings for our primary analyses.

Meta-Analytic Procedure

Effect Size

The effect sizes for the current meta-analysis were the zero-order product-moment correlations between the 12 items of the CSES. If these were not available, we calculated them from the available raw data or reproduced the implied correlations from the reported factor pattern matrices (see Supplement D, Gnambs & Schroeders, 2023). Cross-loadings in the factor pattern matrices that were omitted by the authors (e.g., below .30) were imputed with a value of 0 which leads to appropriate recovery of the correlations between items (Gnambs & Staufenbiel, 2016). Five samples provided full correlation matrices, whereas factor loading patterns from exploratory or confirmatory factor analyses were available from ten and three samples, respectively. All but two-factor analyses (Karasová & Očenášá, 2014; Sun & Jiang, 2017) reported full-factor loading patterns without missing values. The remaining 35 samples provided raw data.

Meta-Analytic Factor Analyses

The factor structure of the CSES was examined using meta-analytic structural equation modeling (MASEM). Following the two-stage structural equation modeling approach (TSSEM; Cheung & Chan, 2005), we first conducted a multivariate random-effects meta-analysis to pool the correlation matrices reported in the individual studies with a maximum likelihood estimator (see Cheung, 2013). Because correlations are estimated more precisely in larger samples, each correlation matrix was weighted by the inverse of its asymptotic sampling (co)variances which were derived following Cheung and Chan (2004). Then, we determined the optimal number of factors based on the pooled correlation matrix. In line with prevalent recommendations (Auerswald & Moshagen, 2019), multiple

criteria were used to decide on the optimal number of factors to retain. These included the empirical Kaiser criterion (Braeken & Van Assen, 2017), Velicer's (1976) minimum average partial (MAP) test, the Hull method (Lorenzo-Seva et al., 2011), and Horn's (1965) parallel analysis.

In the second step, the pooled correlation matrix was subjected to weighted least square factor analyses. Because the precision of the pooled correlations can vary (e.g., depending on their homogeneity in the primary studies), we used the asymptotic sampling variance-covariance matrix from the first step as weights in our factor analyses. First, we conducted exploratory factor analyses with oblimin rotation ($\delta = 0$). Then we compared different theoretically derived models with confirmatory factor analyses (see Figure 1). Model fit was considered acceptable for a comparative fit index (CFI) $\geq .95$, non-normed fit index (NNFI; also known as Tucker-Lewis Index) $\geq .95$, root mean square error of approximation (RMSEA) $\leq .08$, and a standardized root mean square residual (SRMR) $\leq .10$. Values of CFI $\geq .97$, NNFI $\geq .97$, RMSEA $\leq .05$, and SRMR $\leq .05$ were considered indicators for good model fit (Schermele-Engel et al., 2003). Following Bader and Moshagen (2022), we used the NNFI and RMSEA for model comparisons. In contrast to fit indices that do not take the model complexity into account or information criteria that are strongly affected by sample size, parsimony-adjusted goodness of fit indices more accurately identify the best relative fit for model selection. The measurement precision of the latent factors was quantified using model-based reliabilities with different variants of ω (see Flora, 2020).

Although the adopted MASEM approach indirectly accounts for between-study heterogeneity by estimating random effects for the pooled correlations, it cannot directly estimate the heterogeneity of the structural parameters. Therefore, we adopted a simulation approach to create credibility intervals for the factor loadings (see Yu et al., 2016). To this end, 1,000 correlation matrices were randomly drawn from a multivariate normal distribution that used the pooled correlations and their random variances as distributional parameters. Because the goodness of fit indices from these simulated samples can be severely biased (Cheung, 2018), the thus generated samples were only used to create 95% credibility intervals, but not for model selection.

Analysis of Measurement Invariance

Measurement invariance across different context variables (i.e., language versions and cultural dimensions) was examined using moderated factor analysis (Bauer, 2017) extended to the meta-analytic context (Jak & Cheung, 2020). In contrast to the more popular multi-group approach to measurement invariance that focuses on global fit indices to infer (non-)invariance for the entire factor

model (Schroeders & Gnamb, 2020), moderated factor analyses estimate moderating effects for each parameter individually, thus, giving access to more fine-grained information on (non-)invariance. To this end, one-stage meta-analytic structural equation models (OSMASEM) were estimated in a single step by constraining the implied covariance matrix of the fitted factor model to reflect the pooled correlations (Jak & Cheung, 2020). OSMASEM and TSSEM without moderators typically result in highly comparable point estimates and standard errors for the SEM parameters (e.g., Gnamb & Sengewald, 2023; Jak & Cheung, 2022) and, thus, can be used interchangeably. However, TSSEM is computationally more efficient and faster. Moreover, meta-analytic exploratory factor analyses currently require the pooled correlation matrix, similar to TSSEM. However, moderation analyses including analyses of measurement invariance in TSSEM are limited to subgroup comparisons for categorical moderators (Jak & Cheung, 2018). In contrast, OSMASEM is more versatile and can accommodate categorical as well as metric moderators in the context of moderated factor analysis (Bauer, 2017). In this approach, factor loadings can be modeled conditional on a moderator to gauge measurement invariance. To guard against an inflated alpha error rate, we used *p*-values that were corrected for multiple comparisons following Benjamini and Hochberg (1995). Moreover, as a threshold for practically relevant non-invariance, we consider standardized moderating effects (i.e., differences in factor loadings) of more than .10 as noteworthy, thus potentially problematic for fair comparisons along the studied variable. This threshold corresponds to a small to moderate effect according to a review of factor loading differences in empirical research (Nye et al., 2019).

Sensitivity Analyses

Correlation matrices representing potential outliers were identified using standardized residuals and Cook's (1977) distance (see Viechtbauer & Cheung, 2010). The impact of these samples on the factor analytic results was examined by excluding these outliers and repeating the focal analyses. Because there are no established methods for the examination of publication bias in MASEM, we compared the factor structure of the CSES between samples published in peer-reviewed articles and those from other sources (e.g., theses, unpublished datasets). Finally, we considered the study quality as another biasing influence that might affect the factor analytic results. Therefore, we weighted each correlation matrix by the inverse of the risk of bias score and estimated the MASEM for a set of hypothetical samples of the highest quality. Detailed information on these analyses is given in Supplement G (Gnamb & Schroeders, 2023).

Results

Study Characteristics

The meta-analytic database included 53 independent samples (see Table 1 for an overview) that administered the CSES between 2006 and 2022 (*Mdn* = 2019). These samples included a total of $N = 31,843$ participants with a median sample size of $N = 310$ participants (Min = 117, Max = 4,908). The share of women ranged from 12% to 99% (*Mdn* = 56%), while the average ages varied from 15 to 55 years (*Mdn* = 36). Most samples came from the United States (19%), Germany (19%), and China (15%). Consequently, the most frequent language versions (with the number of samples in parenthesis) were English (15), German (9), Chinese (8), Romanian (5), and Spanish (3). The remaining samples were administered in various languages (see Table 1). The cultural values of the included samples spanned a broad range. The dimension of individualism versus collectivism had a median of 33 (Min = -101, Max = 182), whereas values for the dimension of flexibility versus monumentalism fell between -153 and 174 (*Mdn* = 29). Most samples were retrieved from peer-reviewed articles (81%), whereas the rest were available from theses (6%) or unpublished data (13%). Finally, the study quality varied greatly with risk of bias scores ranging from 1.5 to 8.0 with a median of 5.

Exploratory Factor Analyses

The pooled correlations between the 12 items of the CSES (see Supplement E, Gnamb & Schroeders, 2023) showed moderate item correlations between .22 and .53 (*Mdn* = .33) that were slightly larger within positively or negatively worded items (*Mdn* = .38/.40) as compared to items with different wording (*Mdn* = .28). The criteria used to decide on the number of underlying factors for the pooled correlation matrix came to different conclusions: Whereas the Hull method and the minimum average partial test suggested 1 factor, the empirical Kaiser criterion and parallel analysis indicated 2 factors. Therefore, we estimated two exploratory factor models with either 1 or 2 factors (see Table 2). The unidimensional model exhibited strong factor loadings for all items that fell between .54 and .71 (*Mdn* = .59), corroborating the assumption of a common underlying trait for the CSES. In contrast, the two-dimensional model split the positively and negatively worded items into distinct dimensions. The salient factor loadings fell between .42 and .73 (*Mdn* = .59) and, thus, were substantially larger than the cross-loadings (*Mdn* = .05, Min = .00, Max = .21). The two factors were substantially correlated at $r = .64$, advocating for a common second-order factor for the CSES.

Confirmatory Factor Analyses

We fitted three theory-driven models (see Figure 1) to the pooled correlation matrix. The first model followed the theoretical conceptualization of the CSES as a unidimensional measure and specified a single general factor for all items. Although all items had substantial loadings on the latent factor (*Mdn* = .58, Min = .50, Max = .66), the model fit was not acceptable concerning the above-mentioned cutoff values (see Table 3). In contrast, a second-order model with two theoretically motivated factors that correspond to the item key provided a substantially better fit. Because a second-order factor with only two indicators is not identified, the respective factor loadings were constrained to be equal. Please note that this higher-order model is mathematically identical to a two-dimensional correlated factor model. All items had acceptable loadings on the first-order factors (see Model 2 in Figure 1). Moreover, the standardized factor loadings of .85 on the second-order factor indicated that, despite the two (wording) facets, a strong general factor could explain most of the covariations between the items. Finally, a second-order model that operationalized the four content facets as first-order factors (i.e., self-esteem, emotional stability, self-efficacy, and locus of control) showed a worse model fit. Moreover, the loadings on the second-order factor of all facets approached or even slightly exceeded 1 (also known as an ultra-Heywood case), indicating that the facets cannot be properly distinguished. Taken together, these analyses show that the CSES is not completely unidimensional but might subsume two content facets of positive and negative CSE.

In contrast to these content-driven factor models, we also explored three method artifact models that assumed a single content trait for all items but additionally acknowledged method effects related to the item wording (see Figure 1). The bifactor specification assumed two orthogonal factors capturing method effects for the positively or negatively worded items. Although this model exhibited an excellent fit (see Table 3), two items showed rather low loadings on the negative method factor ($< .10$), thus, questioning the assumption of a homogeneous method effect for all negatively worded items. A bifactor-(S-1) model, in which only a single method factor for the negatively worded is specified, also yielded an excellent model fit. In contrast to the complete bifactor model, the method factor in this reference model provided substantial factor loadings and factor saturation ($\omega_s = .39$). Finally, we explored an acquiescence model that assumed homogeneous wording effects for all items. Because of software constraints, we constrained all factor loadings to be equal while fixing the variance to 1 rather than following the typical approach of fixing the factor loadings and freely estimating the factor variance. As a result, the constrained factor loading reflects the standard

Table 2. Meta-analytic exploratory factor loading pattern for Core Self-Evaluations Scale

Item	Single factor model		Oblique two-factor model		
	Factor 1	h^2	Factor 1	Factor 2	h^2
I01: I am confident I get the success I deserve in life.	.62	.38	.04	.59	.39
I02: Sometimes I feel depressed. [#]	.56	.32	.69	-.04	.44
I03: When I try, I generally succeed.	.56	.31	-.07	.68	.41
I04: Sometimes when I fail, I feel worthless. [#]	.59	.35	.70	-.02	.47
I05: I complete tasks successfully.	.54	.29	-.05	.64	.37
I06: Sometimes, I do not feel in control of my work. [#]	.54	.29	.59	.01	.35
I07: Overall, I am satisfied with myself.	.71	.50	.21	.51	.45
I08: I am filled with doubts about my competence. [#]	.66	.44	.57	.14	.44
I09: I determine what will happen in my life.	.54	.29	.00	.56	.32
I10: I do not feel in control of my success in my career. [#]	.60	.35	.42	.20	.32
I11: I am capable of coping with most of my problems.	.64	.41	.07	.59	.41
I12: There are times when things look pretty bleak and hopeless. [#]	.63	.39	.73	-.01	.52
Eigenvalue	4.33		2.53	2.35	
Proportion of variance	36%		21%	20%	
Proportion of explained variance	100%		52%	48%	

Note. $N = 31,843$ from 53 samples. Exploratory weighted least square factor analysis with direct oblimin rotation based on the pooled correlation matrix. The factor correlation for the oblique model was $r = .64$. All items were recoded in such a way that higher values indicate higher core self-evaluations. Salient factor loadings with absolute values greater than .40 are in bold. h^2 = Communality. [#]Negatively keyed item.

Table 3. Goodness of Fit Statistics for different meta-analytic confirmatory factor models

Model	χ^2 (df)	CFI	NNFI	SRMR	RMSEA	AIC	BIC	ω_g	ω_s
Substantive factor models									
1. Single factor model	1,637 (54)*	.91	.89	.07	.03	1,529	1,078	.86	
2. Higher-order two-factor model	476 (53)*	.98	.97	.03	.02	370	-74	.73	.79/.81
3. Higher-order four-factor model	1,601 (50)*	.91	.89	.06	.03	1,501	1,083	.86	.58/.70/.61/.56
Artifact models with wording effects									
4. Bifactor model	175 (42)*	.99	.99	.02	.01	91	-260	.74	.34/.09
5. Bifactor-(S-1) model	328 (48)*	.98	.98	.03	.01	232	-169	.75	.39
6. Acquiescence model	438 (53)*	.98	.97	.03	.02	332	-112	.87	.00

Note. $N = 31,843$ from 53 samples. CFI = Comparative Fit Index; NNFI = Non-normed Fit Index; SRMR = Standardized Root Mean Residual; RMSEA = Root Mean Square Error of Approximation; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; ω_g = Omega reliability for general factor; ω_s = Omega reliability for specific factors. Model evaluation should also take into account the factor loadings of the different models (see Figure 1). * $p < .05$.

deviation of the latent factor. Again, the goodness of fit was good, only slightly inferior to the bifactor specification despite requiring substantially fewer free parameters. However, the variance estimated for the method factor was rather small falling below .06.

Despite the differences in global model fit, the estimated loadings for the general factor were highly comparable for the three models. The loadings in the substantive single-factor model correlated at .87 and .98 with the respective loadings of the bifactor and acquiescence models, respectively. In contrast, the factor loadings of the bifactor-(S-1) artifact model were substantially different because they acknowledged wording effects for only a subset of items. As a result, the loadings of the negatively worded items were substantially smaller as compared to the loadings of the positively

worded items (see Model 5 in Figure 1). Consequently, the factor loadings of the bifactor-(S-1) factor model correlated with the loadings of the other three between $-.32$ and $.21$. Finally, the credibility intervals (see Supplement F, Gnamb & Schroeders, 2023) showed pronounced between-sample heterogeneity in the factor loadings. On average, the width of the intervals was .20, thus, suggesting that potential moderators might affect the measurement of CSES across samples.

In summary, based on model fit alone neither the content-driven factor models nor the artifact models seemed clearly superior. However, when also taking parsimony and interpretability into account, we prefer the acquiescence model (Model 6 in Figure 1). In contrast to the bifactor artifact models, it accounted for a single response style

Table 4. Meta-Analytic Measurement Invariance across language versions and cultural scores

Item	Measurement invariance for language versions					Measurement invariance for cultural scores		
	Loadings for	Difference in factor loadings for				Loadings for	Difference in factor loadings for	
	English (k = 15)	Chinese (k = 8)	German (k = 9)	Romanian (k = 5)	Spanish (k = 3)	United States (k = 10)	Individualism	Flexibility
I01	.62 (.01)	-.16 (.03)	.00 (.02) ^{ns}	-.12 (.03)	.02 (.04) ^{ns}	.59 (.01)	.04 (.01)	-.06 (.01)
I02 [#]	.62 (.01)	-.21 (.03)	-.03 (.02) ^{ns}	-.08 (.03)	-.08 (.04) ^{ns}	.59 (.01)	.05 (.01)	-.06 (.01)
I03	.55 (.01)	-.11 (.03)	.00 (.02) ^{ns}	-.04 (.03) ^{ns}	-.02 (.03) ^{ns}	.54 (.01)	.01 (.01) ^{ns}	-.03 (.01)
I04 [#]	.64 (.01)	-.09 (.03)	-.02 (.02) ^{ns}	-.07 (.03)	-.01 (.03) ^{ns}	.63 (.01)	.00 (.01) ^{ns}	-.05 (.01)
I05	.52 (.01)	-.02 (.03) ^{ns}	.05 (.02)	.00 (.03) ^{ns}	-.03 (.04) ^{ns}	.53 (.01)	.01 (.02) ^{ns}	-.01 (.01) ^{ns}
I06 [#]	.57 (.01)	-.10 (.03)	-.01 (.02) ^{ns}	-.04 (.03) ^{ns}	-.05 (.03) ^{ns}	.56 (.01)	-.01 (.01) ^{ns}	-.05 (.01)
I07	.70 (.01)	-.15 (.03)	.03 (.02) ^{ns}	-.05 (.03) ^{ns}	-.06 (.03) ^{ns}	.68 (.01)	.08 (.01)	-.03 (.01)
I08 [#]	.67 (.01)	-.07 (.03)	.02 (.02) ^{ns}	-.09 (.03)	-.02 (.03) ^{ns}	.66 (.01)	.02 (.01) ^{ns}	-.03 (.01)
I09	.53 (.01)	-.19 (.03)	.04 (.02) ^{ns}	-.16 (.03)	-.02 (.04) ^{ns}	.51 (.01)	.05 (.01)	-.02 (.01) ^{ns}
I10 [#]	.64 (.01)	-.17 (.03)	-.11 (.02)	-.13 (.03)	-.03 (.04) ^{ns}	.58 (.01)	-.01 (.01) ^{ns}	-.06 (.01)
I11	.60 (.01)	-.18 (.02)	.07 (.02)	.03 (.03) ^{ns}	-.11 (.03)	.60 (.01)	.04 (.01)	-.05 (.01)
I12 [#]	.68 (.01)	-.18 (.02)	-.03 (.02) ^{ns}	-.03 (.03) ^{ns}	-.09 (.03)	.65 (.01)	.05 (.01)	-.06 (.01)
SD _w	.20 (.01)	.06 (.01)	.08 (.01)	.14 (.01)	.01 (.02)	.24 (.00)	.01 (.01)	.01 (.01)

Note. $N = 25,006$ from 40 samples and $N = 27,677$ from 46 samples for measurement invariance on language versions and cultural scores, respectively. $k =$ Number of samples. $SD_w =$ Standard deviation of latent wording factor. Presented are (differences in) factor loadings with standard errors in parentheses of moderated factor analyses. The English language version was used as the reference for the analyses presented left, while moderators were centered on the values of the United States and standardized for the analyses presented right. All items were recoded in such a way that higher values indicate higher core self-evaluations. All parameters are significant at $p < .05$, except those marked with ^{ns} p -values for each moderator were corrected for multiple comparisons (Benjamini & Hochberg, 1995). [#]Negatively keyed item.

that equally affected all items and, in line with the CSES's original conceptualization, reflected a technically one-dimensional construct.

Reliability

For all examined models, the general or second-order factor showed good omega reliabilities between .73 and .87 (see Table 3). Thus, independent of the chosen modeling approach the CSES captured a common construct rather precisely. In contrast, the specific method factors reflected only rather small variance components which is quite common for nested factors. The respective reliability estimates reached .39 for the bifactor-(S-1) factor model but were often substantially lower. For the acquiescence model, the respective omega reliability was close to .00.

Analyses of Measurement Invariance

Because our meta-analytic database included samples of different cultural backgrounds that were often administered translated versions of the CSES, we examined metric measurement invariance of the acquiescence model across language versions and cultural dimensions.

We selected a subgroup of 40 samples administering the English, Chinese, German, Romanian, or Spanish language versions because multiple independent samples were available for these languages. As a global test of measurement invariance, we contrasted a multi-group factor model with

out invariance constraints (*configural invariance*) with a model that constrained the general factor loadings across the five groups (*metric invariance*). Because the latter exhibited a worse fit, $\chi^2(313) = 2106$, CFI = .93, NNFI = .93, SRMR = .06, RMSEA = .03, than the configural model, $\chi^2(265) = 1267$, CFI = .96, NNFI = .95, SRMR = .04, RMSEA = .03, the assumption of comparable measurement models for all five translations was not supported. However, constraining only the factor loadings for English, German, and Spanish language versions while freely estimating the loadings in Chinese and Romanian translations exhibited a fit that was comparable to the configural model, $\chi^2(289) = 1477$, CFI = .96, NNFI = .95, SRMR = .05, RMSEA = .03. These results indicate that the CSES exhibited metric invariance for three language versions.

An evaluation of measurement invariance at the item level with moderated factor analyses supported this conclusion. Using four dummy-coded variables (with English as a reference) as moderators of the factor structure highlighted partial metric measurement invariance for the English, German, and Spanish versions (see Table 4). Although some items showed noticeably different factor loadings for the German and Spanish translations in comparison to the English version, only for one item the difference in factor loadings was considered substantial ($\Delta = -.11$). In contrast, half of the items in the Romanian translation exhibited significant non-invariance, with three items showing substantially lower loadings as compared to the English version (Δ s between $-.16$ and $-.12$). Also, the factor loadings of

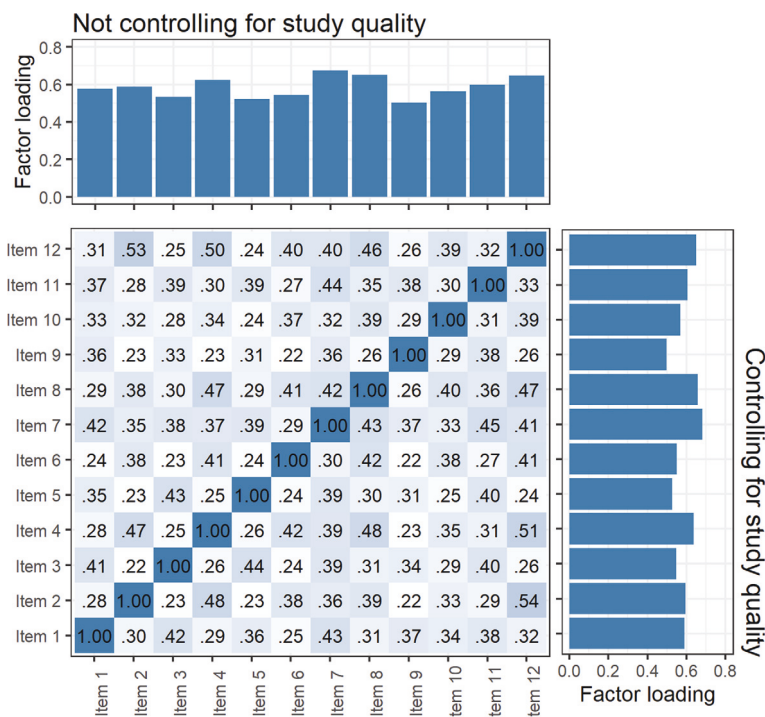


Figure 2. Pooled correlations and factor loadings for the core self-evaluations scale with and without controlling for study quality. Presented are pooled correlations between the items of the CSES and the factor loadings for meta-analytic confirmatory factor analysis of the acquiescence model. Upper-diagonal results on the top do not control for study quality, while lower diagonal results on the right control for study quality.

Chinese language versions were systematically lower as compared to the original version, with a median difference in factor loadings of $-.16$ (Max = $-.20$). Moreover, for all but one item there were significant moderating effects. Thus, cross-cultural comparisons between different language versions are most likely infeasible for the Chinese translation of the CSES.

Measurement invariance across cultural scores was limited to 46 samples for which respective information was available. To facilitate interpretations, the cultural scores were centered at the mean scores for the United States and standardized to yield standard deviations of 1. The moderating effects summarized in Table 4 show that the cultural dimension of flexibility yielded more pronounced effects than individualism. Although six items had significantly ($p < .05$) larger factor loadings with increasing individualism scores, the respective differences in factor loadings were rather modest and did not exceed $.08$. In contrast, flexibility showed significant moderating effects for all but two items, thus, exhibiting a more consistent pattern. But again, the differences in factor loadings were rather moderate and did not exceed $.06$.

Sensitivity Analyses

The factor analytic results were rather robust and hardly distorted by ten samples that were classified as outliers (see Table S4 in Supplement G, Gnambs & Schroeders,

2023). Also, samples pooling correlation matrices reproduced from factor loading patterns yielded highly comparable results to MASEMs based on raw data or correlation matrices (see Table S5). However, unpublished studies showed slightly larger factor loadings for three items as compared to published studies (Table S7); thus, it cannot be ruled out that publication bias impacted the reported results to some degree. Finally, although the risk of bias for the included studies varied substantially (see Table 1), controlling for the study quality hardly affected the factor analytic results. Figure 2 shows that the pooled correlations and factor loadings were rather similar, regardless of whether we controlled for the study quality or not. The maximum difference in factor loadings between the two analyses was $.01$, indicating that differences in the quality of scientific reporting did not affect the statistics that underlie the results of the present meta-analysis.

Discussion

In contrast to the deficit orientation often adopted in psychological research, CSE research takes a more positive perspective on people’s strengths and resources (Judge et al., 1998; Johnson et al., 2008). In this context, the brief CSES has become an established self-report instrument transcending its original use in work and organizational

psychology. Its demonstrated validity, for example, in predicting psychopathological symptoms and physical health (e.g., Turska & Stepień-Lampa, 2021; Zenger et al., 2015) also made it a useful measure to study people's quality of life or evaluate psychosocial interventions. Despite its popularity, the internal structure remained an open point for debate. Because prior research cast doubts on the single-factor conceptualization of the CSES (e.g., Henderson & Gardiner, 2019; Mäkikangas et al., 2018; Sun & Jiang, 2017), the present study evaluated its dimensionality from a meta-analytic perspective. These analyses confirmed that the CSES is not strictly unidimensional, but secondary factors confound the measurement of CSE.

The interpretation of these additional factors can be challenging because the second-order model with two first-order factors that specify qualitatively different types of CSE (i.e., positive and negative CSE) and models focusing on methodological artifacts fitted the data equally well. Given that the CSES was constructed with a single factor in mind (Judge et al., 2003), we believe that different content facets should only be considered after methodological explanations can be ruled out. Following this reasoning, evidence for a facet structure of the CSES is scarce as it was largely unidimensional after accounting for C/IER in the form of an acquiescence factor. Also, prior studies discussing different content facets of the CSES only derived post hoc explanations for these facets after failing to corroborate the single factor structure (e.g., Mäkikangas et al., 2018; Zenger et al., 2015), thus, making the theoretical underpinning of different CSES facets rather weak. Therefore, it is more likely that the unmodeled method effects bias the measurement structure of the CSES to some degree. In our opinion, the decision of how to appropriately account for these biases is best guided by matters of parsimony and interpretability (see also Preacher & Merkle, 2012). In this respect, the acquiescence model (Model 6 in Figure 1) seems to represent a good compromise between both criteria. First, it requires only one additional parameter as compared to the single-factor model and, thus, does not suffer from overparameterization. Second, the loading structure conformed to the original conception of the CSES with comparable loadings for positively and negatively worded items. In contrast, the bifactor specification (with two specific factors) resulted in anomalous factor loadings for selected items while the bifactor-(S-1) model showed systematically lower general factor loadings for negatively worded items. Both patterns would not be expected from homogenous method effects or CSE theory.

From an applied perspective, the exact nature of the multidimensionality might not be as important as the knowledge that a general common factor accounts for most of the item variance. This seems to be the case for the CSES as demonstrated by the reliability estimates for the different

models. Regardless of whether one adheres to the view of substantive content facets or methodological artifacts, a common factor accounted for 74%–87% of the score variance. These results fall in line with a recent reliability generalization of coefficient alphas that attested to a high measurement precision to the CSES (Ock et al., 2021).

Implications for Practice

The meta-analytic results indicate that using the English version of the CSES as an essentially unidimensional measure is warranted. In line with its theoretical understanding (Judge et al., 2003), the 12 items predominately capture a single latent factor. Although acknowledging method factors in addition to the focal CSE trait improves model fit from a psychometric perspective, the informational gain from the more complex modeling is rather small. Our results also provided some support for the simple sum score that is typically used for the CSES because the general factor loadings were rather similar for most items. This justifies ordinary sum scoring over more complex scoring schemes that incorporate different item weights or try to separate systematic method and trait variance (see McNeish & Wolf, 2020) because the latter is unlikely to improve person estimates of CSE. We further discourage the use of subscales for positive and negative CSE because these subscales not only lack a substantive theoretical underpinning but also cannot be properly distinguished from response styles associated with the item wording.

As a caveat, we think that the dimensionality of self-reports should not be considered independently of sample characteristics or the context. For example, in online research C/IER responses can bias assessments (Woods, 2006). Thus, empirical research is well-advised to adopt appropriate countermeasures, for example, by excluding conspicuous respondents (e.g., Arias et al., 2022; Schroeders et al., 2022) or explicitly modeling response styles (e.g., Aichholzer, 2014; Scharl & Gnamb, 2022). Finally, although the reported results do not exempt researchers from conducting systematic analyses of measurement invariance before addressing substantive research questions, our findings provide preliminary evidence of comparable measurement structures in English, German, and Spanish versions of the CSES.

Limitations and Future Research

The meta-analytic results can be extended in several ways. First, our analyses of measurement invariance were necessarily brief and restricted by the available data. We could only examine metric invariance for five language versions of the CSES. Therefore, cross-cultural research would be

well advised to extend these analyses to additional languages and, furthermore, to investigations of scalar invariance which is required for mean-level comparisons (see Schroeders & Gnambs, 2020). Second, our results point to potential problems with some existing translations. Whereas the German and Spanish language versions of the CSES were largely invariant to the original English version, Chinese and Romanian translations were more problematic and did not allow for cross-cultural comparisons. However, we have to concede that we were unable to validate the exact Chinese language version administered in the available primary studies (i.e., Mandarin or Cantonese) because this information was rarely reported. Moreover, our analyses of measurement invariance were limited to differences in *standardized* factor loadings. However, if the latent factor variances differed substantially between groups, measurement invariance of unstandardized versus standardized factor loadings can yield different results. Indeed, in our meta-analytic database, the samples providing raw data showed restricted variances in the Chinese and Romanian samples as compared to the English-speaking samples; that is, the median variance ratios of the item scores were 0.76 and 0.87. In contrast, for the German and Spanish samples, no variance restriction was observed. Thus, it is conceivable that the observed non-invariance was a result of differences in sample characteristics (i.e., trait distributions), rather than differences in item characteristics. Therefore, we encourage further attempts to improve adaptations of the CSES in other languages by evaluating unstandardized measurement structures or using properly matched samples. Finally, an important extension of the present findings would be meta-analytic research on the criterion validity of the CSES. In this context, it might also be worthwhile to evaluate whether these validity correlations are susceptible to the choice of a specific psychometric model or how response styles are taken into account (see Scharl & Gnambs, 2022, for related findings).

Conclusion

Prior factor analytic studies often failed to substantiate strict unidimensionality for the CSES (Judge et al., 2003). The present meta-analytic investigation of its measurement structure showed that a rather simple extension that accounted for acquiescence responding resulting from careless or insufficient effort responding towards differently worded items substantially improved the measurement of CSE. In line with its original conception, the CSES is dominated by a strong and reliable general factor for all items that can be measured rather precisely across different language versions and cultural dimensions. On a more general note, we agree with Schmalbach and colleagues (2021) that

response biases should be controlled when evaluating psychological instruments before more complex measurement structures are considered that were not originally hypothesized.

References

- References marked with * were included in the meta-analysis.
- Aichholzer, J. (2014). Random intercept EFA of personality scales. *Journal of Research in Personality, 53*, 1–4. <https://doi.org/10.1016/j.jrp.2014.07.001>
- *Algner, M., & Lorenz, T. (2022). You're prettier when you smile: Construction and validation of a questionnaire to assess microaggressions against women in the workplace. *Frontiers in Psychology, 13*, Article 809862. <https://doi.org/10.3389/fpsyg.2022.809862>
- *Arias, V. B., & Arias, B. (2017). The negative wording factor of Core Self-Evaluations Scale (CSES): Methodological artifact, or substantive specific variance? *Personality and Individual Differences, 109*, 28–34. <https://doi.org/10.1016/j.paid.2016.12.038>
- *Arias, V. B., Ponce, F. P., & Martinez-Molina, A. (2022). How a few inconsistent respondents can confound the structure of personality survey data: An example with the Core-Self Evaluations Scale. *European Journal of Psychological Assessment*. Advance online publication. <https://doi.org/10.1027/1015-5759/a000719>
- Auerswald, M., & Moshagen, M. (2019). How to determine the number of factors to retain in exploratory factor analysis: A comparison of extraction methods under realistic conditions. *Psychological Methods, 24*(4), 468–491. <https://doi.org/10.1037/met0000200>
- Bader, M., & Moshagen, M. (2022). Assessing the fitting propensity of factor models. *Psychological Methods*. Advance online publication. <https://doi.org/10.1037/met0000529>
- Bauer, D. J. (2017). A more general model for testing measurement invariance and differential item functioning. *Psychological Methods, 22*, 507–526. <https://doi.org/10.1037/met0000077>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B, 57*, 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- *Boyd, S. L. (2006). *Core self-evaluations as a moderator of the job stress-burnout relationship* (Publication No. 304912080) [Doctoral dissertation, Alliant International University]. ProQuest Dissertations & Theses Global. <https://www.proquest.com/dissertations-theses/core-self-evaluations-as-moderator-job-stress/docview/304912080/se-2>
- Braeken, J., & Van Assen, M. A. (2017). An empirical Kaiser criterion. *Psychological Methods, 22*(3), 450–466. <https://doi.org/10.1037/met0000074>
- Chang, C. H., Ferris, D. L., Johnson, R. E., Rosen, C. C., & Tan, J. A. (2012). Core self-evaluations: A review and evaluation of the literature. *Journal of Management, 38*(1), 81–128. <https://doi.org/10.1177/0149206311419661>
- Cheung, M. W.-L. (2013). Multivariate meta-analysis as structural equation models. *Structural Equation Modeling, 20*(3), 429–454. <https://doi.org/10.1080/10705511.2013.797827>
- Cheung, M. W. L. (2018). Issues in solving the problem of effect size heterogeneity in meta-analytic structural equation modeling: A commentary and simulation study on Yu, Downes, Carter, and O'Boyle (2016). *Journal of Applied Psychology, 103*(7), 787–803. <https://doi.org/10.1037/apl0000284>
- Cheung, M. W.-L., & Chan, W. (2004). Testing dependent correlation coefficients via structural equation modeling.

- Organizational Research Methods*, 7(2), 206–223. <https://doi.org/10.1177/1094428104264024>
- Cheung, M. W.-L., & Chan, W. (2005). Meta-analytic structural equation modeling: A two-stage approach. *Psychological Methods*, 10(1), 40–64. <https://doi.org/10.1037/1082-989X.10.1.40>
- Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, 19(1), 15–18. <https://doi.org/10.2307/1268249>
- *Coroiu, A., Kwakkenbos, L., Moran, C., Thombs, B., Albani, C., Bourkas, S., Zenger, M., Braehler, E., & Körner, A. (2018). Structural validation of the Self-Compassion Scale with a German general population sample. *PLoS One*, 13(2), Article e0190771. <https://doi.org/10.1371/journal.pone.0190771>
- *Curșeu, P. L., Rusu, A., Maricuțoiu, L. P., Virgă, D., & Măgurean, S. (2020). Identified and engaged: A multi-level dynamic model of identification with the group and performance in collaborative learning. *Learning and Individual Differences*, 78, Article 101838. <https://doi.org/10.1016/j.lindif.2020.101838>
- *Dieckmann, N. F., & Hartman, R. O. (2022). Conspiracist and paranormal beliefs: A typology of non-reductive ideation. *International Journal of Personality Psychology*, 8, 47–57. <https://doi.org/10.21827/ijpp.8.38006>
- *Ding, H., & Yu, E. (2022). How and when does follower's strengths-based leadership contribute to follower work engagement? The roles of strengths use and core self-evaluation. *German Journal of Human Resource Management*, 36(2), 180–196. <https://doi.org/10.1177/23970022211053284>
- DiStefano, C., & Motl, R. W. (2006). Further investigating method effects associated with negatively worded items on self-report surveys. *Structural Equation Modeling*, 13, 440–464. https://doi.org/10.1207/s15328007sem1303_6
- Eid, M., Geiser, C., Koch, T., & Heene, M. (2017). Anomalous results in G-factor models: Explanations and alternatives. *Psychological Methods*, 22(3), 541–562. <https://doi.org/10.1037/met0000083>
- Elliot, A. J., & Thrash, T. M. (2002). Approach-avoidance motivation in personality: Approach and avoidance temperaments and goals. *Journal of Personality and Social Psychology*, 82, 804–818. <https://doi.org/10.1037/0022-3514.82.5.804>
- *Farčić, N., Barac, I., Plužarić, J., Ilakovac, V., Pačarić, S., Gvozdanić, Z., & Lovrić, R. (2020). Personality traits of core self-evaluation as predictors on clinical decision-making in nursing profession. *PLoS One*, 15(5), Article e0233435. <https://doi.org/10.1371/journal.pone.0233435>
- *Ferreira, M. C., Thadeu, S. H., da Costa Masagão, V., da Silva Gottardo, L. F., Gabardo, L. M. D., Sousa, S. A. A., & Mana, T. C. T. (2013). Escala de avaliações autorreferentes: características psicométricas em amostras brasileiras [Core Self-Evaluation Scale: Psychometric properties in Brazilian samples]. *Avaliação Psicológica: Interamerican Journal of Psychological Assessment*, 12(2), 227–232. <http://pepsic.bvsalud.org/pdf/avp/v12n2/v12n2a13.pdf>
- Ferris, D. L., Rosen, C. R., Johnson, R. E., Brown, D. J., Risavy, S. D., & Heller, D. (2011). Approach or avoidance (or both?): Integrating core self-evaluations within an approach/avoidance framework. *Personnel Psychology*, 64(1), 137–161. <https://doi.org/10.1111/j.1744-6570.2010.01204.x>
- Flora, D. B. (2020). Your coefficient alpha is probably wrong, but which coefficient omega is right? A tutorial on using R to obtain better reliability estimates. *Advances in Methods and Practices in Psychological Science*, 3(4), 484–501. <https://doi.org/10.1177/2515245920951747>
- *Förster, P., Brähler, E., Stöbel-Richter, Y., & Berth, H. (2013). *Saxonian longitudinal study – Wave 25, 2011* (ZA6243; Version 1.0.0) [Data set]. GESIS Data Archive. <https://doi.org/10.4232/1.11511>
- Gardner, D. G., & Pierce, J. L. (2010). The Core Self-Evaluation Scale: Further construct validation evidence. *Educational and Psychological Measurement*, 70(2), 291–304. <https://doi.org/10.1177/0013164409344505>
- *Geuens, N., Verheyen, H., Vlerick, P., Van Bogaert, P., & Franck, E. (2020). Exploring the influence of core-self evaluations, situational factors, and coping on nurse burnout: A cross-sectional survey study. *PLoS One*, 15(4), Article e0230883. <https://doi.org/10.1371/journal.pone.0230883>
- Gnamb, T., Scharl, A., & Schroeders, U. (2018). The structure of the Rosenberg Self-Esteem Scale: A cross-cultural meta-analysis. *Zeitschrift für Psychologie*, 226(1), 14–29. <https://doi.org/10.1027/2151-2604/a000317>
- Gnamb, T., & Schroeders, U. (2020). Cognitive abilities explain wording effects in the Rosenberg Self-Esteem Scale. *Assessment*, 27(2), 404–418. <https://doi.org/10.1177/1073191117746503>
- Gnamb, T., & Schroeders, U. (2023). *Supplement Material for "Reliability and factorial validity of the Core Self-Evaluations Scale: A meta-analytic investigation of wording effects"*. <https://osf.io/zjvwg/>
- Gnamb, T., & Sengewald, M.-A. (2023). Meta-analytic structural equation modeling with fallible measurements. *Zeitschrift für Psychologie*, 231(1), 39–52. <https://doi.org/10.1027/2151-2604/a000511>
- Gnamb, T., & Staufenbiel, T. (2016). Parameter accuracy in meta-analyses of factor structures. *Research Synthesis Methods*, 7(2), 168–186. <https://doi.org/10.1002/jrsm.1190>
- *Gu, H., Wen, Z., & Fan, X. (2015). The impact of wording effect on reliability and validity of the Core Self-Evaluation Scale (CSES): A bi-factor perspective. *Personality and Individual Differences*, 83, 142–147. <https://doi.org/10.1016/j.paid.2015.04.006>
- *Gurbuz, S., Costigan, R., & Teke, K. (2021). Does being positive work in a Mediterranean collectivist culture? Relationship of core self-evaluations to job satisfaction, life satisfaction, and commitment. *Current Psychology*, 40(1), 226–241. <https://doi.org/10.1007/s12144-018-9923-6>
- Heilmann, T., & Jonas, K. (2010). Validation of a German-language Core Self-Evaluations Scale. *Social Behavior and Personality: An International Journal*, 38(2), 209–225. <https://doi.org/10.2224/sbp.2010.38.2.209>
- *Heller, S., Ullrich, J., & Mast, M. S. (2023). Power at work: Linking objective power to psychological power. *Journal of Applied Social Psychology*, 53(1), 5–20. <https://doi.org/10.1111/jasp.12922>
- *Henderson, T., & Gardiner, E. (2019). The Core Self-Evaluation Scale: A replication of bi-factor dimensionality, reliability, and criterion validity. *Personality and Individual Differences*, 138, 312–320. <https://doi.org/10.1016/j.paid.2018.10.015>
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2), 179–185. <https://doi.org/10.1007/BF02289447>
- *Hou, L., Gu, X., & Ding, G. (2022). From an identity process theory perspective: A daily investigation of why and when ostracism triggers ingratiation. *Journal of Social Psychology*. Advance online publication. <https://doi.org/10.1080/00224545.2022.2139215>
- *Jain, S., & Nair, S. K. (2019). Exploring the moderating role of core self-evaluation in the relationship between demands and work-family enrichment. *Journal of Indian Business Research*, 12(2), 249–270. <https://doi.org/10.1108/JIBR-08-2017-0125>
- Jak, S., & Cheung, M. W. L. (2018). Testing moderator hypotheses in meta-analytic structural equation modeling using subgroup analysis. *Behavior Research Methods*, 50(4), 1359–1373. <https://doi.org/10.3758/s13428-018-1046-3>
- Jak, S., & Cheung, M. W. L. (2020). Meta-analytic structural equation modeling with moderating effects on SEM parameters. *Psychological Methods*, 25(4), 430–455. <https://doi.org/10.1037/met0000245>
- Jak, S., & Cheung, M. W. L. (2022). *Can findings from meta-analytic structural equation modeling in management and*

- organizational psychology be trusted?* PsyArxiv. <https://doi.org/10.31234/osf.io/b3qvn>
- Johnson, R. E., Rosen, C. C., & Levy, P. E. (2008). Getting to the core of core self-evaluation: A review and recommendations. *Journal of Organizational Behavior*, 29(3), 391–413. <https://doi.org/10.1002/job.514>
- Judge, T. A., Erez, A., Bono, J. E., & Thoresen, C. J. (2003). The Core Self-Evaluations Scale: Development of a measure. *Personnel Psychology*, 56(2), 303–331. <https://doi.org/10.1111/j.1744-6570.2003.tb00152.x>
- 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., Van Vianen, A. E., & De Pater, I. E. (2004). Emotional stability, core self-evaluations, and job outcomes: A review of the evidence and an agenda for future research. *Human Performance*, 17(3), 325–346. https://doi.org/10.1207/s15327043hup1703_4
- *Karasová, J., & Očenášá, L. (2014). Testing the psychometric properties of the questionnaire the Core Self-Evaluations Scale: Pilot study. In E. Maierová, R. Procházka, M. Dolejš, & O. Skopal (Eds.), *Proceedings of the Czech and Slovak psychological conference (not only) for postgraduates and about postgraduates* (pp. 231–238). Palacký University Olomouc.
- *Kim, J., Lee, S., & You, C. Y. (2015). Analysis of the factor structure of core self-evaluations through exploratory structural equation modeling. *Korean Journal of Industrial and Organizational Psychology*, 28(3), 355–384. <https://doi.org/10.24230/kjiop.v28i3.355-384>
- Koutsogiorgi, C. C., Lordos, A., Fanti, K. A., & Michaelides, M. P. (2021). Factorial structure and nomological network of the inventory of callous-unemotional traits accounting for item keying variance. *Journal of Personality Assessment*, 103(3), 312–323. <https://doi.org/10.1080/00223891.2020.1769112>
- Krippendorff, K. (2013). *Content analysis: An introduction to its methodology*. Sage.
- *Leonhardt, M., Bechtoldt, M. N., & Rohrmann, S. (2017). All impostors aren't alike: Differentiating the impostor phenomenon. *Frontiers in Psychology*, 8, Article 1505. <https://doi.org/10.3389/fpsyg.2017.01505>
- *Li, X., Guan, L., Chang, H., & Zhang, B. (2014). Core self-evaluation and burnout among nurses: The mediating role of coping styles. *PLoS ONE*, 9(12), Article e115799. <https://doi.org/10.1371/journal.pone.0115799>
- *Littrell, S., Risko, E. F., & Fugelsang, J. A. (2021). The bullshitting frequency scale: Development and psychometric properties. *British Journal of Social Psychology*, 60(1), 248–270. <https://doi.org/10.1111/bjso.12379>
- *Liu, Z., Sun, X., Guo, Y., & Luo, Y. (2023). Mindful parenting inhibits adolescents from being greedy: The mediating role of adolescent core self-evaluations. *Current Psychology*, 42, 15991–16000. <https://doi.org/10.1007/s12144-019-00577-3>
- Lorenzo-Seva, U., Timmerman, M. E., & Kiers, H. A. (2011). The Hull method for selecting the number of common factors. *Multivariate Behavioral Research*, 46(2), 340–364. <https://doi.org/10.1080/00273171.2011.564527>
- *Love, Z. M. (2016). *Rules of (employee) engagement: A comprehensive model* (Publication No. 10307459) [Doctoral dissertation, East Carolina University]. ProQuest Dissertations & Theses Global. <https://www.proquest.com/dissertations-theses/rules-employee-engagement-comprehensive-model/docview/1872716901/se-2>
- *Mäkikangas, A., Kinnunen, U., Mauno, S., & Selenko, E. (2018). Factor structure and longitudinal factorial validity of the Core Self-Evaluation scale. *European Journal of Psychological Assessment*, 34(6), 444–449. <https://doi.org/10.1027/1015-5759/a000357>
- Maydeu-Olivares, A., & Coffman, D. L. (2006). Random intercept item factor analysis. *Psychological Methods*, 11, 344–362. <https://doi.org/10.1037/1082-989X.11.4.344>
- McNeish, D., & Wolf, M. G. (2020). Thinking twice about sum scores. *Behavior Research Methods*, 52, 2287–2305. <https://doi.org/10.3758/s13428-020-01398-0>
- Michaelides, M. P. (2019). Negative keying effects in the factor structure of TIMSS 2011 motivation scales and associations with reading achievement. *Applied Measurement in Education*, 32(4), 365–378. <https://doi.org/10.1080/08957347.2019.1660349>
- Minkov, M., Bond, M. H., Dutt, P., Schachner, M., Morales, O., Sanchez, C., Jandosova, J., Khassenbekov, Y., & Mudd, B. (2018). A reconsideration of Hofstede's fifth dimension: New flexibility versus monumentalism data from 54 countries. *Cross-Cultural Research*, 52(3), 309–333. <https://doi.org/10.1177/1069397117727488>
- Minkov, M., Dutt, P., Schachner, M., Morales, O., Sanchez, C., Jandosova, J., Khassenbekov, Y., & Mudd, B. (2017). A revision of Hofstede's individualism-collectivism dimension: A new national index from a 56-country study. *Cross Cultural & Strategic Management*, 24(3), 386–404. <https://doi.org/10.1108/CCSM-11-2016-0197>
- *Mussel, P., De Vries, J., Spengler, M., Frintrup, A., Ziegler, M., & Hewig, J. (2023). The development of trait greed during young adulthood: A simultaneous investigation of environmental effects and negative core beliefs. *European Journal of Personality*, 37(3), 352–371. <https://doi.org/10.1177/08902070221090101>
- *Nastasa, M., Golu, F., Buruiana, D., & Oprea, B. (2021). Teachers' work-home interaction and satisfaction with life: The moderating role of core self-evaluations. *Educational Psychology*, 41(6), 806–820. <https://doi.org/10.1080/01443410.2020.1852182>
- Nudelmann, G., & Otto, K. (2020). The development of a new generic risk-of-bias measure for systematic reviews of surveys. *Methodology*, 16(4), 278–298. <https://doi.org/10.5964/meth.4329>
- *Nurmohamed, S., Kundro, T. G., & Myers, C. G. (2021). Against the odds: Developing underdog versus favorite narratives to offset prior experiences of discrimination. *Organizational Behavior and Human Decision Processes*, 167, 206–221. <https://doi.org/10.1016/j.obhdp.2021.04.008>
- Nye, C. D., Bradburn, J., Olenick, J., Bialko, C., & Drasgow, F. (2019). How big are my effects? Examining the magnitude of effect sizes in studies of measurement equivalence. *Organizational Research Methods*, 22(3), 678–709. <https://doi.org/10.1177/1094428118761122>
- Ock, J., McAbee, S. T., Ercan, S., Shaw, A., & Oswald, F. L. (2021). Reliability generalization analysis of the Core Self-Evaluations Scale. *Practical Assessment, Research & Evaluation*, 26, Article 6. <https://doi.org/10.7275/zsc7-jw58>
- *Pătrașcu-Lungu, A., & Iliescu, D. (2022). Goal self-concordance mediates the relation of core self-evaluations with organizational citizenship behavior but not with environmental organizational citizenship behavior. *Psihologia Resurselor Umane*, 20(2), 112–130. <https://doi.org/10.24837/pru.v20i2.514>
- *Pitt-Catsouphes, M., & Smyer, M. (2013). *Age and generations study 2007–2008* (ICPSR 34837; Version V1) [Data set]. ICPSR. <http://doi.org/10.3886/ICPSR34837.v1>
- Preacher, K. J., & Merkle, E. C. (2012). The problem of model selection uncertainty in structural equation modeling. *Psychological Methods*, 17(1), 1–14. <https://doi.org/10.1037/a0026804>
- Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation

- methods under suboptimal conditions. *Psychological Methods*, 17(3), 354–373. <https://doi.org/10.1037/a0029315>
- *Rosenbloom, J. L., & Ash, R. A. (2013). *Professional worker career experience survey, United States, 2003–2004* (ICPSR 26782; Version V1) [Data set]. ICPSR. <https://doi.org/10.3886/ICPSR26782.v1>
- Scharl, A., & Gnambs, T. (2022). The impact of different methods to correct for response styles on the external validity of self-reports. *European Journal of Psychological Assessment*. Advance online publication. <https://doi.org/10.1027/1015-5759/a000731>
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, 8(2), 23–74.
- *Scheurer, A. J. (2013). *Antecedents of informal learning: A study of core self-evaluations and work-family conflict and their effects on informal learning* (Publication No. 10307459) [Master's thesis, Ohio State University]. OhioLINK Electronic Theses and Dissertations Center. http://rave.ohiolink.edu/etdc/view?acc_num=osu1366270012
- Schmalbach, B., Zenger, M., Michaelides, M. P., Schermelleh-Engel, K., Hinz, A., Körner, A., Beutel, M. E., Decker, O., Kliem, S., & Brähler, E. (2021). From bi-dimensionality to unidimensionality in self-report questionnaires: Applying the random intercept factor analysis model to six psychological tests. *European Journal of Psychological Assessment*, 37(2), 135–148. <https://doi.org/10.1027/1015-5759/a000583>
- Schroeders, U., & Gnambs, T. (2020). Degrees of freedom in multi-group confirmatory factor analysis: Are models of measurement invariance testing correctly specified? *European Journal of Psychological Assessment*, 36(1), 105–113. <https://doi.org/10.1027/1015-5759/a000500>
- Schroeders, U., Schmidt, C., & Gnambs, T. (2022). Detecting careless responding in survey data using stochastic gradient boosting. *Educational and Psychological Measurement*, 82(1), 29–56. <https://doi.org/10.1177/00131644211004708>
- *Smedema, S. M., Morrison, B., Yaghmaian, R. A., Deangelis, J., & Aldrich, H. (2016). Psychometric validation of the Core Self-Evaluations Scale in people with spinal cord injury. *Disability and Rehabilitation*, 38(9), 889–896. <https://doi.org/10.3109/09638288.2015.1065012>
- *Sprung, J. (2021). *Self-employed & WFC* [Data set]. Figshare. https://figshare.com/articles/dataset/Self-Employed_WFC/14128583
- Stumpp, T., Muck, P. M., Hülshager, U. R., Judge, T. A., & Maier, G. W. (2010). Core self-evaluations in Germany: Validation of a German measure and its relationships with career success. *Applied Psychology*, 59(4), 674–700. <https://doi.org/10.1111/j.1464-0597.2010.00422.x>
- *Sulaiman, A. M., Alfuqaha, O. A., Shaath, T. A., Alkurdi, R. I., & Almomani, R. B. (2021). Relationships between core self-evaluation, leader empowering behavior, and job security among Jordan University Hospital nurses. *PLoS One*, 16(11), Article e0260064. <https://doi.org/10.1371/journal.pone.0260064>
- *Sun, P., & Jiang, H. (2017). Psychometric properties of the Chinese version of core self-evaluations scale. *Current Psychology*, 36(2), 297–303. <https://doi.org/10.1007/s12144-016-9418-2>
- *Swab, R. (2021). *CSECompCleanData* [Data set]. Figshare. https://figshare.com/articles/dataset/CSECompCleanData_xlsx/13708198
- *Thielmann, I., & Hilbig, B. E. (2019). Nomological consistency: A comprehensive test of the equivalence of different trait indicators for the same constructs. *Journal of Personality*, 87(3), 715–730. <https://doi.org/10.1111/jopy.12428>
- *Tims, M., & Akkermans, J. (2017). Core self-evaluations and work engagement: Testing a perception, action, and development path. *PLoS One*, 12(8), Article e0182745. <https://doi.org/10.1371/journal.pone.0182745>
- *Tisu, L., Lupşa, D., Virgă, D., & Rusu, A. (2020). Personality characteristics, job performance and mental health: The mediating role of work engagement. *Personality and Individual Differences*, 153, Article 109644. <https://doi.org/10.1016/j.paid.2019.109644>
- *Turska, E., & Stępień-Lampa, N. (2021). Well-being of Polish university students after the first year of the coronavirus pandemic: The role of core self-evaluations, social support and fear of COVID-19. *PLoS One*, 16(11), Article e0259296. <https://doi.org/10.1371/journal.pone.0259296>
- Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, 41(3), 321–327. <https://doi.org/10.1007/BF02293557>
- Viechtbauer, W., & Cheung, M. W. L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, 1(2), 112–125. <https://doi.org/10.1002/jrsm.11>
- *Virga, D., De Witte, H., & Cifre, E. (2017). The role of perceived employability, core self-evaluations, and job resources on health and turnover intentions. *Journal of Psychology*, 151(7), 632–645. <https://doi.org/10.1080/00223980.2017.1372346>
- *Virga, D., & Rusu, A. (2018). Core self-evaluations, job search behaviour and health complaints: The mediating role of job search self-efficacy. *Career Development International*, 23(3), 261–273. <https://doi.org/10.1108/CDI-11-2017-0208>
- Woods, C. M. (2006). Careless responding to reverse-worded items: Implications for confirmatory factor analysis. *Journal of Psychopathology and Behavioral Assessment*, 28(3), 186–191. <https://doi.org/10.1007/s10862-005-9004-7>
- *Yaras, Z. (2022). The mediating role of personality traits on head of school's core self-evaluation presenteeism behaviors. *International Online Journal of Educational Sciences*, 14(1), 157–176.
- Yu, J. J., Downes, P. E., Carter, K. M., & O'Boyle, E. H. (2016). The problem of effect size heterogeneity in meta-analytic structural equation modeling. *Journal of Applied Psychology*, 101(10), 1457–1473. <https://doi.org/10.1037/apl0000141>
- *Zacher, H., Rudolph, C. W., & Posch, M. (2021). Individual differences and changes in self-reported work performance during the early stages of the COVID-19 pandemic. *Zeitschrift für Arbeits- und Organisationspsychologie*, 65(4), 188–201. <https://doi.org/10.1026/0932-4089/a000365>
- *Zaniboni, S., Topa, G., & Balducci, C. (2021). Core self-evaluations affecting retirement-related outcomes. *International Journal of Environmental Research and Public Health*, 18(1), Article 174. <https://doi.org/10.3390/ijerph18010174>
- *Zenger, M., Körner, A., Maier, G. W., Hinz, A., Stöbel-Richter, Y., Brähler, E., & Hilbert, A. (2015). The Core Self-Evaluation Scale: Psychometric properties of the German version in a representative sample. *Journal of Personality Assessment*, 97(3), 310–318. <https://doi.org/10.1080/00223891.2014.989367>
- *Zheng, X., Wu, B., Li, C. S., Zhang, P., & Tang, N. (2021). Reversing the Pollyanna effect: The curvilinear relationship between core self-evaluation and perceived social acceptance. *Journal of Business and Psychology*, 36(1), 103–115. <https://doi.org/10.1007/s10869-019-09666-3>

History

Received December 16, 2022
 Revision received April 27, 2023
 Accepted May 11, 2023
 Published online February 5, 2024
 EJPA Section / Category I/O Psychology

Conflict of Interest

We have no conflicts of interest to disclose.

Open Science

Open Data: We confirm that there is sufficient information for an independent researcher to reproduce all of the reported results. The coded data and analysis code is available in the supplement material (Gnams & Schroeders, 2023).

Open Materials: We confirm that there is sufficient information for an independent researcher to reproduce the reported methodology (see Gnams & Schroeders, 2023).

Preregistration of Studies and Analysis Plans: This study was not preregistered.

Funding

Open access publication enabled by University of Bamberg.

ORCID

Timo Gnams

 <https://orcid.org/0000-0002-6984-1276>

Ulrich Schroeders

 <https://orcid.org/0000-0002-5225-1122>

Timo Gnams

Leibniz Institute for Educational Trajectories

Wilhelmsplatz 3

96047 Bamberg

Germany

timo.gnams@lifbi.de