

Translation, Validation and Extended Factor Models of the German State Difficulties in Emotion Regulation Scale (S-DERS)

Sicorello^{1,2}, M., Elsaesser³, M., Kolar⁴, D. R.

¹Institute for Psychosomatic Medicine and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Heidelberg, Germany

²German Center for Mental Health (DZPG), Partner Site Mannheim, Mannheim, Germany

³Department of Psychiatry and Psychotherapy, Medical Center – University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany

⁴Department of Psychology, University of Regensburg, Regensburg, Germany

Author note

Correspondence concerning this article should be addressed to Dr. Maurizio Sicorello, Central Institute of Mental Health, J5, 68159 Mannheim, Germany.

Email: maurizio.sicorello@zi-mannheim.de

Acknowledgments

Maurizio Sicorello received funding for this study by the German Psychological Association Section: Biopsychology and Neuropsychology. We are thankful to all participants. We are further thankful to Wiebke Heyse, Nina Kölbl, Johanna Steinke und Elias Löffl for support during data collection.

Abstract

Difficulties in emotion regulation are an essential transdiagnostic phenomenon of mental health. While there is extensive literature focusing on emotion regulation difficulties as trait-like constructs, it is inherently a dynamic process that unfolds over time. This raises the need for state-like measures to capture these temporal dynamics in both laboratory and real-world settings, such as the State Difficulties in Emotion Regulation Scale (S-DERS). In the present study, we report the German translation and validation of the S-DERS and present a cross-sectional network analysis to stimulate novel hypotheses for pathways linking single discrete difficulties. A sample of 214 participants underwent a negative mood induction followed by an assessment of the 21-item S-DERS. The results confirm the four-factor structure, reliability, and construct validity of the original English version. Consistent with previous studies on the trait-DERS, we find that an exploratory structural equation modeling approach with extensive cross-loadings is necessary to reach good model fit. Moreover, a model with four correlated factors outperformed a bi-factor and high-order model. This indicates there is no single latent construct of emotion regulation difficulties, but rather four distinct but related constructs for (a) Non-acceptance of Current Emotions, (b) Limited Ability to Modulate Current Emotional and Behavioral Responses, (c) Lack of Awareness of Current Emotions, and (d) Lack of Clarity about Current Emotions. These associations may imply a complex causal interplay between these factors.

Keywords: emotion regulation, psychometrics, psychopathology, assessment, structural equation models

Introduction

Emotion regulation is a key concept for mental and physical well-being, comprising all implicit and explicit strategies affecting the occurrence, intensity, or duration of emotional experiences (Gross & Thompson, 2007). There are many stages at which these processes can fail, starting from the identification of emotions, over the selection of a strategy, its implementation, and the monitoring of success (Fernandez et al., 2016). One particularly influential clinical model outlined six categories of emotion regulation difficulties, which can negatively affect a person's well-being (Gratz & Roemer, 2004): A *Lack of Awareness* as the diminished tendency to attend and acknowledge one's emotions, a *Lack of Clarity* as the extent to which people understand their emotional experiences, the *Non-Acceptance* of primary emotions in the form of secondary meta-emotional responses (e.g., shame or anger for feeling a certain way), a limited access to helpful *Strategies*, difficulties engaging in *Goal-Directed Behavior* as well as *Impulsive Behavior* when confronted with negative emotions. This model gave rise to the six factor Difficulties in Emotion Regulation Scale (DERS), which has been used in thousands of empirical studies, repeatedly demonstrating the relevance of these dimensions for the etiology and maintenance of psychopathology across diagnostic boundaries (Clamor et al., 2024; Daros & Williams, 2019; Miola et al., 2022; Stellern et al., 2023).

Originally, most emotion regulation studies focused on trait-like constructs assessed with cross-sectional questionnaires—including the DERS—at a single time point, primarily measuring a person's general concept of their own emotional processes. Still, emotions and their regulation typically unfold over brief time periods (Gross, 2015). Studies in the laboratory or daily life can address this state-like dynamic nature of emotion regulation difficulties as they emerge, which is the fundamental building block of trait-like difficulties (Fernandez et al., 2016). Therefore, Lavender and colleagues (2017) used original items of the 36-item DERS to construct the 21-item State Difficulties in Emotion Regulation Scale (S-DERS) for momentary assessments. Based on an experimental induction of negative mood and exploratory factor analysis, they determined that state difficulties are best captured by a smaller item set comprising four factors instead of the original six DERS factors. This solution merged the factors for strategies, goal-directed behavior, and impulsiveness into a single factor reflecting the limited ability to *Modulate* current emotional and behavioral responses. It has been applied to a large range of clinical topics such as eating disorders, alcohol use, sleep disruption, and daily affective experiences (Emery & Simons, 2020; Forester et al., 2023; Pavlacic et al., 2022; Reid et al., 2022).

In the present study, we present a translation, preregistered validation, and factor modeling extension for a German version of the S-DERS. At the core, we aimed to perform a conceptual replication of the original S-DERS construction study, which included an experimental induction of negative mood to provide an appropriate affective context for emotion regulation difficulties to occur. We preregistered hypotheses that (H1) our mood induction increased negative affect based on the valence dimension, (H2) we replicate the original four-factor structure, (H3) internal consistency reliability will be satisfactory ($>.80$) for all scales except for the Clarity scale, which had limited internal consistency in the original study ($>.60$), (H4) there is convergent validity between the S-DERS and the trait-DERS scales, and (H5) concurrent criterion validity for other relevant constructs from the clinical and personality domain. Specifically, we expected positive associations of the global S-DERS score with psychopathology, neuroticism, and affective reactivity to the mood induction. Concerning S-DERS subscales, we expected positive associations between Non-Acceptance and experiential avoidance, Modulate and impulsivity, Awareness and mindfulness, as well as Clarity and mindfulness.

The basis of the original four-factor solution of the S-DERS was an exploratory factor analysis (EFA) with parallel analysis and main loadings above .40. We preregistered the same procedure as a criterion for successful structural validation, but also planned stricter models in terms of assumptions and decision criteria. These included a confirmatory factor analysis (CFA) without cross-loadings (simple structure), as the strictest possible model. Still, even for the trait-DERS models without cross-loadings fit measures are often insufficient (Gutzweiler & In-Albon, 2018). Exploratory structural equation modeling (ESEM) provides a framework to fit a model with cross-loadings for all items and has provided improved model fit for the DERS in the past as well as a reduction in factor correlations (Gutzweiler & In-Albon, 2018; Marsh et al., 2014; Sousa et al., 2023). Therefore, we also conducted an ESEM-based analysis with fit measures as the success criterion, which in terms of strictness is between inference based on CFA simple structures and EFA loading sizes.

For the trait-DERS, it is still unclear whether subscales capture a common trait of emotion regulation difficulties. This question is relevant to decide whether the total score or subscale scores should be preferred in a given research application, but previous studies are inconsistent (Bardeen et al., 2016; Sousa et al., 2023; Weathers et al., 2013). Therefore, we also computed a higher-order model to test for a superordinate factor which influences the subscales. Conceptually, such a model implies that a general latent construct of emotion regulation difficulties causally influences the four subscales, which in turn causally influence

the participant responses to the items. In post-hoc exploratory analyses, we contrasted this model with a bifactor model, which implies a direct effect of global difficulties on indicators, which is distinct from the specific constructs underlying the subscale factors. All tested structural equation models are illustrated in Figure 1. Further, we provide an exploratory network analysis as a hypothesis-generating tool to support future research on the causal structure of emotion regulation difficulties (Borsboom et al., 2021).

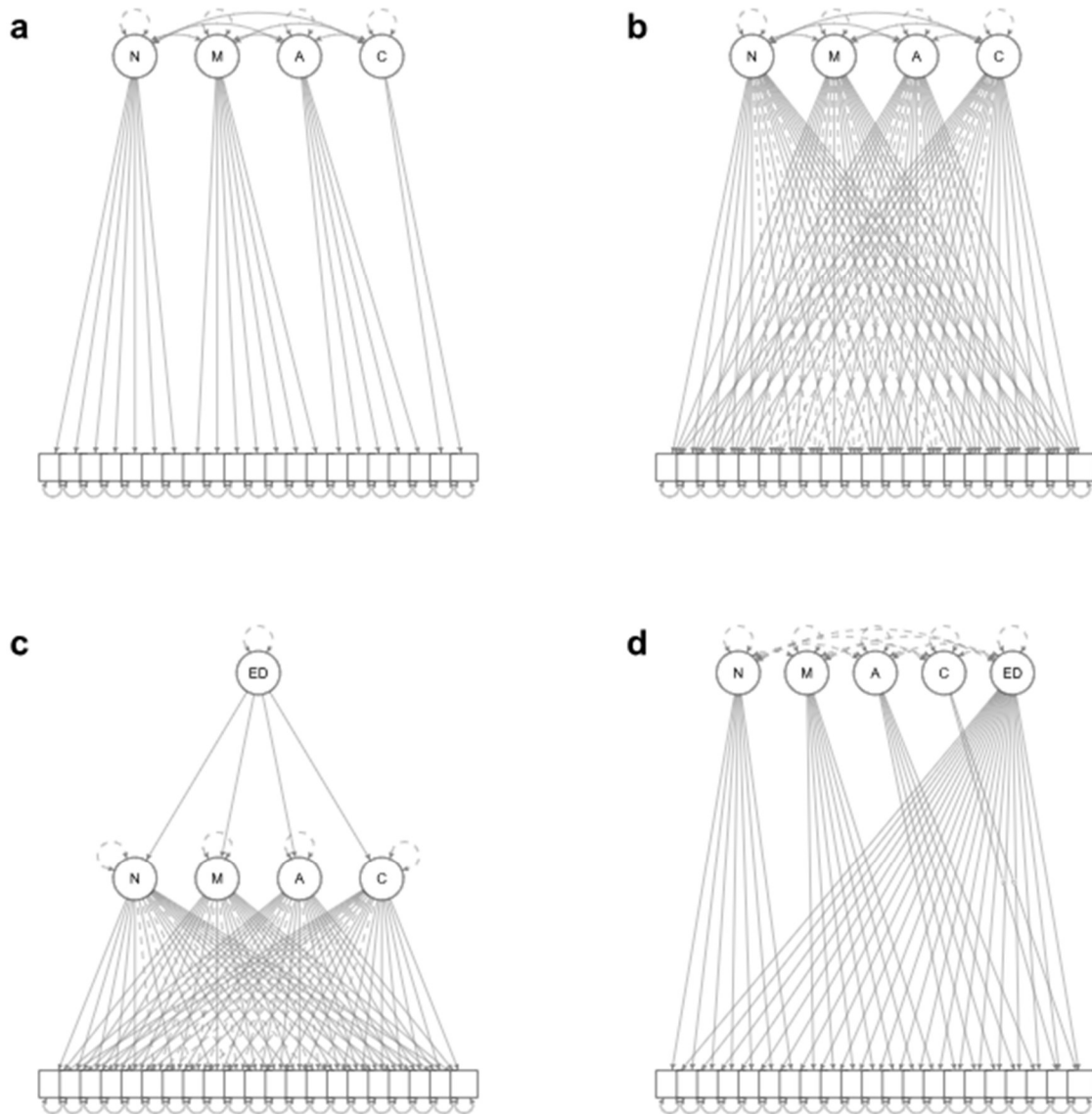


Figure 1. Factor models tested for the S-DERS. (a) CFA simple structure without cross-loadings. (b) ESEM with cross-loadings. (c) Model with a higher-order factor. (d) Bi-factor model. ED = Emotion regulation difficulties. N = Non-Acceptance. M = Modulate. A = Awareness. C = Clarity.

Methods

Participants and Procedures

The sample comprised 214 participants of ages 18-61, which were predominantly young (mean age = 22.5, $SD = 6.5$) female (89%, 10% male, one person gender-diverse) university students (93%). This composition is overall similar to the original S-DERS validation study, which only included young adult women between the age of 18-25. In our sample, 51 participants (24%) reported a diagnosis for a mental disorder in the past, with 26 (12%) having received treatment.

Following the literature on power simulations for confirmatory factor analysis, we aimed to recruit at least 200 participants (Wolf et al., 2013). An a priori stopping rule was set if recruiting plateaus after 200 participants (< 8 participants per week for 4 weeks in a row) or reaches a maximal sample size of 250. According to power calculations with G*Power (v3.1) for one-tailed zero-order correlations, a sample size of 200 allows a statistical power of 90% to detect a population correlation of .20 or higher. This can be viewed as sufficient, given the moderate-to-high correlations needed to attest validity.

Participants were recruited over mailing lists, social media, and internal advertisements at Heidelberg and Regensburg University. As a participation incentive, 25 online coupons á 40€ were provided after study completion to randomly drawn participants. The only inclusion criterion was age between 18-65. The study was performed contactless on participants' smartphones using the SEMA3 platform (O'Brien et al., 2024). This included a demographic section, followed by a section with psychological questionnaires, a mood induction, and the translated S-DERS. This procedure took on average 31.4 minutes ($SD = 20.6$). Psychology students could participate for course credit, which was taken by 88% of participants. The study was approved by the Ethical Board II of Heidelberg University, Germany (2021-519).

Translation

The S-DERS (21 items) was separately translated into German by M.S., M.E., and D.K. without reference to the German DERS version. Then, translations were compared for consistency and discussed, followed by a comparison to the German DERS version to resolve conflicts, settling on a consensual translation. The final German items were retranslated to English by two independent psychologists with no significant differences between the back-translation and the original English S-DERS items.

The translated items can be found in Table S1 (ordered by factor) and Table S2 (administration order with English items for comparison). Item responses ranged from 1 ("Not

at all”/“Überhaupt nicht”) to 5 (“Completely”/“Völlig”). The items of the Awareness scale were reversed and then sum scores for the total scale and the subscales calculated. Notably, while this scoring is consistent with the previous procedure, it differently weighs the four subscales, as they consist of a different number of items. We present exploratory analyses on the differences between this and an unweighted procedure.

Further Psychological Measures

Other psychological measures used for validation included the trait-DERS (36 items, Gutzweiler & In-Albon, 2018), the Neuroticism scale of the NEO Five Factor Inventory (12 items, Borkenau & Ostendorf, 2007), the Depression, Anxiety, and Stress Scale (DASS-21, 21 items, Nilges & Essau, 2021), the Barrett Impulsivity Scale (BIS-15, 15 items, Meule et al., 2011), the Acceptance and Action Questionnaire II (AAQ-II, 7 items, Hoyer & Gloster, 2013), and Awareness and Describe Subscales of the Five Facet Mindfulness Questionnaire Short (FFMQ, 16 items, Michalak et al., 2016). Lastly, the Interpersonal Emotion Regulation Questionnaire (20 items; Pruessner et al., 2020) and the Childhood Trauma Questionnaire (Klinitzke et al., 2012) were administered for a different research project and are therefore not reported in the present study.

Mood Induction

Before the assessment of the S-DERS, all participants underwent a negative mood induction, as negative emotions are likely a prerequisite for the meaningful assessment of momentary emotion regulation difficulties (Lavender et al., 2017). The mood induction consisted of a combination of sad music and autobiographical recall (Werthmann et al., 2014). Participants listened to 6 minutes of the ‘Adagio for Strings’ by Samuel Barber while writing about a recent negative experience. This negative experience was instructed to be a 5-6 on a scale from 0 (not a negative experience) to 10 (extremely negative experience). Immediately before and after the mood induction, participants indicated their momentary affective state based on a two-dimensional affective grid with the dimensions *valence* (from 0 = unpleasant emotion to 10 = pleasant emotion) and *arousal* (from 0 = low energy to 10 = high energy; Tiede, 2020). Lastly, as another manipulation check, participants were asked how well they could immerse themselves in the recalled negative experience (from 1 = not at all to 5 = very well).

Data Analysis

The success of the mood induction was tested as the difference between affective valence before versus after the manipulation using a paired t-test. An exploratory test on the affective arousal dimension is also reported without previously specified expectations.

All factor models tested for the S-DERS are shown in Figure 1. First, we conducted a confirmatory factor analysis for four factors without cross-loadings (i.e., simple structure; Figure 1a) using maximum likelihood estimation with the R package Lavaan (Rosseel, 2012). We preregistered that if model fit measures did not meet the criteria of $RMSEA \leq 0.06$; $CFI \geq 0.95$; $TLI \geq 0.95$; $SRMR \leq 0.08$ (Hu & Bentler, 1999), we test an ESEM model with cross-loadings for all items (Figure 1b) using the esem package (Prokofieva et al., 2023). In this procedure, first, an EFA was fit with the psych package using principal axis factoring and promax rotation, which was also used in the original S-DERS validation. Referent items were identified using the automated procedure of the esem package. Then, this model was refit as a CFA in Lavaan with the loadings of the referent items fixed to their EFA loadings, while the remaining EFA loadings were used as starting values. If this model also did not meet the fit measure criteria above, we preregistered to perform EFA with the number of factors based on parallel analysis at the 99% quantile, with loadings $>.40$ considered as substantial loadings on the main factors and loadings $>.30$ considered as substantial cross-loadings. This EFA procedure is identical to that used in the original validation study. Internal consistency (composite reliability) was calculated as McDonald's ω from the final model (Flora, 2020). Additionally, the higher-order (Figure 1c) and bifactor models (Figure 1d) were tested against the winning model using the Vuong test (Vuong, 1989). Lastly, we conducted an exploratory network analysis on the partial correlation matrix (Figure 2). Pruning of edges was performed based on the Extended Bayesian Information Criterion with $\lambda = 0.5$ and graphical lasso using the bootNet package (Epskamp et al., 2018).

For associations between the S-DERS and other variables, bivariate correlations were computed with a preregistered one-tailed $\alpha = .05$. All analyses were performed in R (v4.3.2).

Open Science Practices

Preregistration, data, and analysis code can be found at: <https://osf.io/ebj6u/>. For reproducibility or other reuse of data and code, we recommend directly accessing the github repository, which also includes a documentation of files, version control, and steps to reproduce analyses: https://github.com/MaurizioSicorello/SDERSvalid_Analysis. A print version of the questionnaire can be found in both repositories as well as the supplements.

There were two notable changes to the preregistration. First, we originally preregistered to conduct EFA with Geomin rotation and maximum likelihood estimation. This is in contrast to the analyses performed by Lavender et al. (2017) and was considered as the *esem* package did not exist at the time of preregistration and the only validated procedure was performed with this estimation procedure. Using the same procedure as Lavender et al. (2017) is an improvement not only due to consistency, but also because principal axis factoring is better suited to identify factors with few items, such as the Clarity scale which only consists of two items. Secondly, we preregistered likelihood ratio tests for model comparisons, but the net procedure indicated models resulting from CFA and ESEM were not nested. Possibly, this was due to the selective fixing of values in ESEM, thus making the Vuong test for non-nested models the safer option (Vuong, 1989). Both procedures led to the same inferential decision.

Results

Manipulation check (H1)

Affective valence decreased after the mood induction, indicating a successful manipulation, albeit with a small-to-moderate effect size ($t(213) = 4.58, p < .001, d_c = .31; M_{pre} = 5.9, SD_{pre} = 1.9; M_{post} = 5.3, SD_{post} = 2.1$). Exploratory analyses indicated that affective arousal also decreased with a slightly larger effect size ($t(213) = 5.9, p < .001, d_c = .40; M_{pre} = 5.6, SD_{pre} = 1.9; M_{post} = 4.9, SD_{post} = 1.9$), further confirming the successful induction of sadness, which is generally classified as a low arousal emotion (Barrett & Russell, 1999). Overall, most participants stated they were successfully immersed in the experience ($M = 3.8, SD = 0.9$).

Factor structure (H2)

The correlations between items visually indicated a solution with four factors and considerable cross-loadings (Figure 2a). For the four-factor simple structure CFA without cross-loadings, none of the fit measures crossed the preregistered threshold (Table 1). The four-factor ESEM structure with cross-loadings improved model fit significantly, while only SRMR crossed the threshold. This model also outperformed the higher-order and bifactor models (Table 1) and no modification indices were statistically significant after Bonferroni-Holm correction. A parallel analysis also indicated a four-factor solution. Overall, the loading pattern conformed to the original S-DERS validation except for two cross-loadings (Table 2): Item 17 of the Modulate scale (“My emotions feel out of control”) had a substantial cross-loading of

.46 on the Clarity factor; item 7 of the Clarity scale had a substantial cross-loading of .37 on the Awareness factor (see discussion).

The four factors were moderately to highly correlated, with the lowest correlation between Modulate and Awareness ($r = .18$) and the largest between Modulate and Non-Acceptance ($r = .58$) the latter being visible as a large block in Figure 2a. Notably, this correlation was even higher in the validation of the original S-DERS. In-between were the correlations for Non-Acceptance and Awareness ($r = .30$), Non-Acceptance and Clarity ($r = .43$), Awareness and Clarity ($r = .35$), and Modulate and Clarity ($r = .41$). Overall, this pattern of factor correlations resembled those of the original S-DERS validation, except for larger correlations between the Awareness scale and the remaining three scales.

In contrast, the network model based on partial correlations emphasizes the proximity of the Awareness and Clarity scales (Figure 2b). Also, it indicates that items 5 (“I am feeling very bad about myself”) and 15 (“I believe that I am going to end up feeling very depressed”) might serve as connectors between the Non-Acceptance and the Modulate clusters (see discussion).

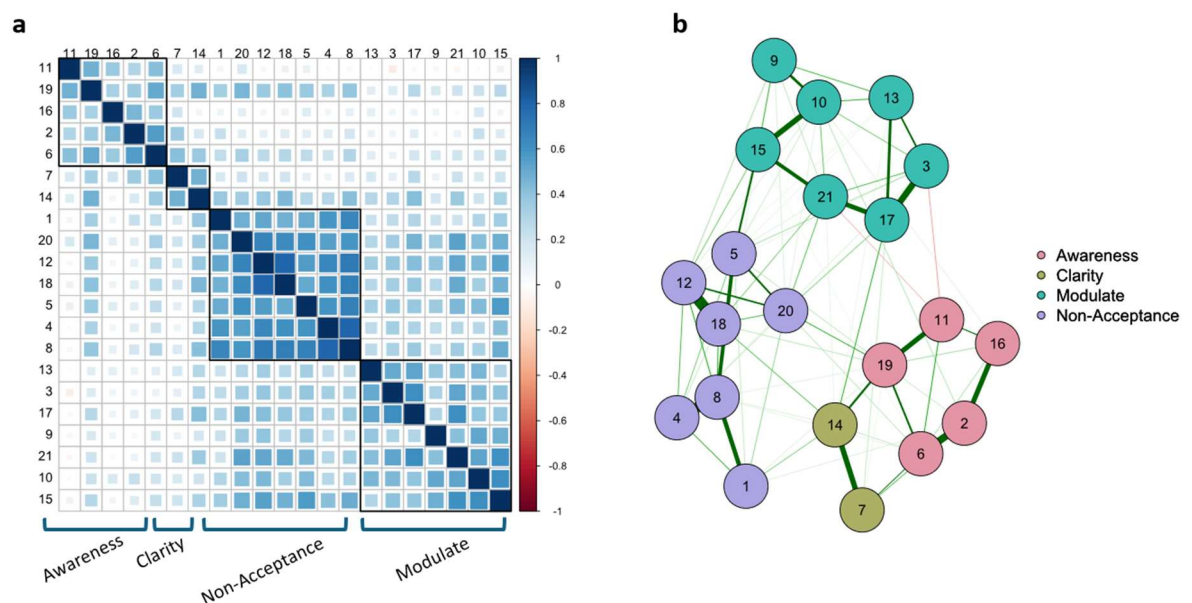


Figure 2. (a) Correlations between S-DERS items. Clusters were determined using hierarchical agglomerative clustering with complete linkage. (b) Network analysis based on the partial correlation matrix.

Table 1*Fit measures and inferential statistics for competing structural equation models*

	SRMR	RMSEA	CFI	TLI	AIC	BIC	χ^2	df	Model Comparison	p-Value
CFA										
Simple Structure	0.078	0.094	0.850	0.828	10,849	11,011	527.6	183		
ESEM										
Cross-loadings	0.039	0.078	0.923	0.881	10,729	11,049	313.3	136	Model 2 > Model 1	0
ESEM										
Higher-order	0.076	0.086	0.904	0.854	10,770	11,083	357.9	138	Model 2 > Model 3	0.001
Bi-factor	0.066	0.082	0.893	0.867	10,764	10,973	414.3	169	Model 2 > Model 4	0.008

Note. Based on $N = 214$ participants. SRMR = standardized root mean square residual, RMSEA = root mean square error of approximation, CFI = comparative fit index, TLI = Tucker–Lewis index, CFA = confirmatory factor analysis, ESEM = exploratory structural equation model, EFA = exploratory factor analysis, AIC = Akaike information criterion, BIC = Bayesian information criterion.

Table 2*Factor loadings and correlations*

Item	Factor			
	1	2	3	4
Factor 1: Non-acceptance of Current Emotions				
Item 8	0.97 (0.91)	-0.12 (0.48)	0.00 (0.29)	0.03 (0.38)
Item 4	0.91 (0.82)	-0.07 (0.43)	-0.06 (0.2)	-0.06 (0.28)
Item 12	0.71 (0.83)	0.19 (0.62)	-0.06 (0.23)	0.04 (0.41)
Item 18	0.69 (0.78)	0.05 (0.53)	-0.01 (0.26)	0.17 (0.47)
Item 1	0.68 (0.68)	-0.08 (0.37)	0.06 (0.27)	0.06 (0.33)
Item 5	0.59 (0.73)	0.29 (0.61)	0.06 (0.28)	-0.13 (0.26)
Item 20	0.53 (0.72)	0.24 (0.6)	0.09 (0.33)	0.05 (0.4)
Factor 2: Limited Ability to Modulate Current Emotional and Behavioral Responses				
Item 10	-0.06 (0.42)	0.87 (0.77)	0.14 (0.26)	-0.22 (0.16)
Item 21	-0.02 (0.47)	0.71 (0.77)	-0.09 (0.11)	0.21 (0.48)
Item 9	0.02 (0.35)	0.64 (0.59)	0.05 (0.16)	-0.15 (0.14)
Item 3	-0.03 (0.38)	0.62 (0.66)	-0.15 (0.03)	0.21 (0.43)
Item 13	-0.12 (0.31)	0.61 (0.6)	-0.03 (0.11)	0.17 (0.37)
Item 15	0.23 (0.58)	0.6 (0.72)	0.07 (0.25)	-0.1 (0.27)
Item 17	-0.08 (0.42)	0.54 (0.67)	-0.07 (0.14)	0.46 (0.64)
Factor 3: Lack of Awareness of Current Emotions				
Item 6 (r)	0.03 (0.32)	0.00 (0.24)	0.70 (0.75)	0.16 (0.35)
Item 2 (r)	-0.09 (0.17)	0.06 (0.16)	0.67 (0.66)	0.02 (0.18)
Item 11 (r)	-0.04 (0.09)	-0.09 (0.01)	0.62 (0.58)	-0.02 (0.08)
Item 16 (r)	-0.13 (0.09)	0.13 (0.13)	0.61 (0.56)	-0.11 (0.04)
Item 19 (r)	0.25 (0.46)	-0.09 (0.27)	0.54 (0.66)	0.22 (0.42)
Factor 4: Lack of Clarity about Current Emotions				
Item 14	0.14 (0.44)	-0.03 (0.37)	0.13 (0.34)	0.66 (0.74)
Item 7	-0.16 (0.17)	0.04 (0.22)	0.37 (0.44)	0.47 (0.51)

Note. Structure coefficients are in parentheses. Bold coefficients mark absolute loadings larger than .30. (r) indicated reversed items. The structure of this table is based on Lavender et al. (2017) for ease of comparison.

Reliability (H3)

As expected, reliabilities were sufficient for the total scale and all subscales except the Clarity scale, converging with the pattern of the original validation study (in parentheses): Total Score: $\omega = .94$ (.86), Non-Acceptance: $\omega = .92$ (.92), Modulate: $\omega = .79$ (.85), Clarity: $\omega = .63$ (.65).

Construct and Criterion validity (H4-H5)

The S-DERS subscales all had the highest correlations with their analogue trait-DERS subscales (Table 3). Still, notably, almost all subscales from both questionnaires had substantial intercorrelations. Together with ESEM-like structure of the four factors with considerable cross-loadings, this indicates a moderate specificity of subscales, inherent to all English and German versions of the trait and state DERS (Gratz & Roemer, 2004; Gutzweiler & In-Albon, 2018; Lavender et al., 2017). A similar pattern emerged for the remaining external constructs: Experiential avoidance most strongly correlated with Non-Acceptance, but also very strongly with the Modulate subscale. Mindful Description appeared to be more specific to the Awareness subscale, as hypothesized, while Mindful Acting with Awareness was strongly correlated to all subscales, with the strongest correlation for the Modulate scale.

There were two surprising findings, which deviated from our expectations. First, the correlations between the BIS-15 scale for impulsivity and the S-DERS subscales were much smaller than expected, albeit still statistically significant for the preregistered association with the Modulate scale, as well as the non-preregistered association with the Clarity scale. Post-hoc analyses indicated this is because only the BIS-15 subscale for attentional impulsivity was substantially related to the S-DERS scales. For this subscale, the correlations conformed to the expectations, with the strongest effect for the Modulate scale at $r = .32$ (Non-Acceptance: $r = .16$, Awareness: $r = .12$, Clarity: $r = .23$).

The second surprising finding was the absence of the expected correlation between affective reactivity to the mood induction and the global S-DERS. In fact, the association went in the opposite direction, with higher state emotion regulation difficulties showing a small association with a stronger decrease in negative affect. These reactivity scores were calculated from two variables: affective valence before minus after the mood induction. Therefore, we inspected the correlations with the two separate variables (pre and post measurement). This showed that the global S-DERS score was substantially correlated with lower (i.e., more

negative) affect before but not after the mood induction (before: $r = -.22, p = .001$; after: $r = -.07, p = .260$). To test whether this observation is specific to the S-DERS, we repeated these correlation analyses with the neuroticism scores instead of S-DERS scores as a trait-based construct for negative affective reactivity, showing the same pattern (reactivity: $r = -.18, p = .008$; before: $r = -.28, p < .001$; after: $r = -.08, p = .231$). In sum, both the S-DERS and neuroticism were more strongly correlated to affect before than after the mood induction, albeit with the plausible effect direction of more negative affect (smaller values) being associated with more emotion regulation difficulties and neuroticism (higher values). To understand how two correlations in a psychologically plausible direction can lead to an unexpected effect direction for a difference score, consider the covariance decomposition:

$$\begin{aligned}
 & \text{cov}(\text{S-DERS}, \text{Reactivity}) \\
 = & \text{cov}(\text{S-DERS}, \text{Affect}_{\text{pre}} - \text{Affect}_{\text{post}}) \\
 = & \text{cov}(\text{S-DERS}, \text{Affect}_{\text{pre}}) - \text{cov}(\text{S-DERS}, \text{Affect}_{\text{post}})
 \end{aligned}$$

Inserting different correlation sizes in the last line shows, if the covariance between the S-DERS and affect *before* the manipulation is more negative than after the manipulation, the observed effect will have a negative sign (i.e., the “wrong” direction). Consequently, if the covariance between the S-DERS and affect *after* the manipulation is more negative than before the manipulation, the observed effect will have a positive sign (i.e., the “right” direction). We will expand on potential psychological reasons in the discussion section.

Table 3*Correlations between S-DERS scales and other constructs*

	S-DERS				
	Total	Non-Acceptance	Modulate	Awareness	Clarity
DERS	0.78	0.64	0.63	0.55	0.48
DERS: Nonacceptance	0.64	0.69	0.44	0.32	0.30
DERS: Awareness	0.44	0.27	0.16	0.68	0.35
DERS: Clarity	0.47	0.31	0.24	0.53	0.55
DERS: Strategies	0.75	0.63	0.68	0.40	0.39
DERS: Impulse	0.58	0.42	0.57	0.32	0.36
DERS: Goals	0.46	0.36	0.50	0.20	0.19
DASS-21	0.58	0.45	0.60	0.30	0.29
Neuroticism	0.58	0.46	0.57	0.33	0.24
Affective Reactivity: Valence	-0.12	-0.07	-0.15	-0.08	-0.07
AAQ-II	0.64	0.51	0.58	0.37	0.39
BIS	0.10	0.01	0.12	0.08	0.15
FFMQ: Describe	-0.29	-0.14	-0.12	-0.48	-0.26
FFMQ: Awareness	-0.44	-0.27	-0.46	-0.29	-0.33

Note. (S-)DERS = (State) Difficulties in Emotion Regulation Scale, DASS-21 = Depression, Anxiety, and Stress Scale, BIS-15 = Barrett Impulsivity Scale, AAQ-II = Acceptance and Action Questionnaire II, FFMQ = Five Facet Mindfulness Questionnaire (short form).

Statistically significant correlations (one-sided, $\alpha = .05$) are shown in bold. The cutoffs for one-sided tests at different common statistical thresholds were:

$p < .05$ (one-sided): 0.11

$p < .01$ (one-sided): 0.16

$p < .001$ (one-sided): 0.21

Weighted versus Unweighted Subscales (Post-hoc)

The fact that S-DERS subscales consist of a different number of items implies a differential weighting of these subscales when a global score is calculated. Therefore, we tested whether equally weighting all subscales increases correlations with psychopathology-related outcomes. We calculated mean scores instead of sum scores for each subscale. Then, we calculated a global scale score as the average over those four means, leading to an equal weight for all subscales independent of the number of items. Overall, we found that the previously used (implicit) differential weighting of scales in the S-DERS is superior to equally weighting all scales for the DASS ($\Delta r = .10$), neuroticism ($\Delta r = .11$), and having been diagnosed with a mental disorder ($\Delta d = .14$). One possible explanation is that the overrepresented subscales (i.e., Non-Acceptance and Modulate) have stronger relations to the three outcomes, as indicated in Table 3. Still, this might also be a result of their higher reliability (i.e., equally representing Clarity in the total score lowers the overall reliability). Therefore, we also computed correlations between the latent (measurement error-free) S-DERS factors and the three outcomes within the ESEM framework (Figure 3). This provided evidence for both explanations as latent effect sizes were much more similar than manifest effect sizes for the four subscales, but the Modulate scale still appeared to outperform the other measures in terms of cumulative effect sizes across all three outcomes. This was driven by strong relations to DASS and neuroticism.

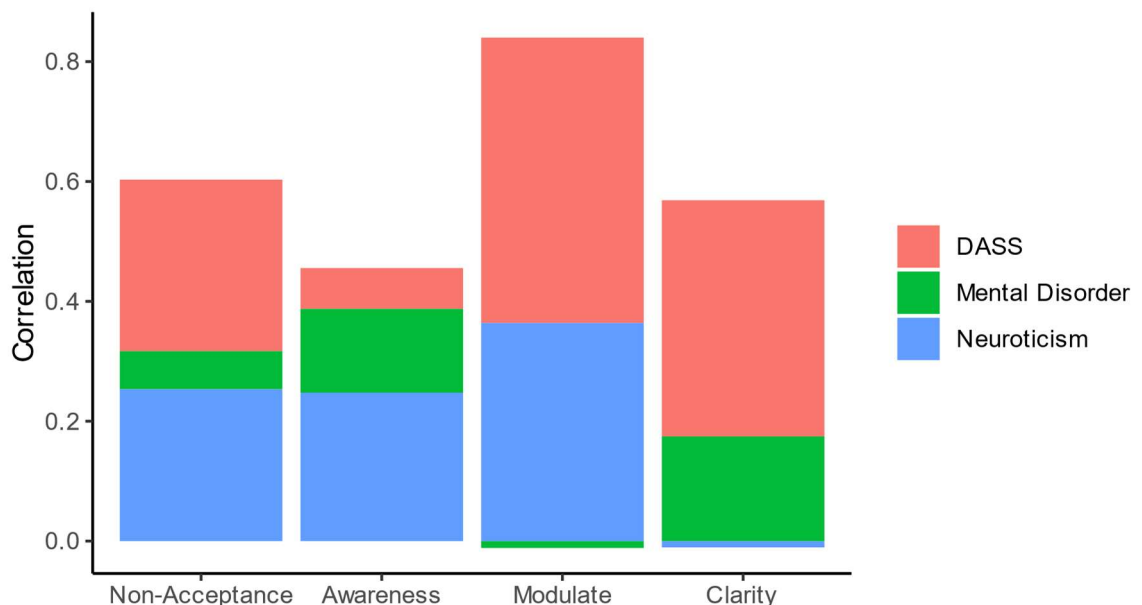


Figure 3. Latent correlations between S-DERS subscales and psychopathology-related outcomes.

Discussion

We presented a German translation and validation of the S-DERS, which might serve future studies on the temporal dynamics of emotion regulation difficulties in the laboratory or daily life. We found that our version largely conforms to the standards of the original validation study based on factor structure, reliability, and associations with related constructs.

Based on parallel analysis and item loadings, we confirmed a four-factor solution with moderately correlated factors and considerable cross-loadings. The importance of cross-loadings for the S-DERS was also confirmed by a comparison to a simple structure model, which had considerably worse fit. This four-factor solution was overall also discernible from the raw correlation matrix and the network analysis.

While all items had sufficiently large loadings on their targeted factor, two items showed stronger cross-loadings than expected. First, item 17 of the Modulate factor “My emotions feel out of control” had a substantial cross-loading on the Clarity factor. In the original trait-DERS, the item wording was “I experience my emotions as overwhelming and out of control”, which might have benefitted a stronger alignment with the Modulate versus the Clarity factor. For example, if someone is confused about their feelings (item 14, Clarity), this might be closer to experiencing emotions as out of control than them being overwhelming. Still, given that we confirmed previous patterns of only moderate discriminant validity for the S-DERS subscales, it might be advantageous to retain this item translation for consistency with the English S-DERS. Also, the trait-DERS wording combines two different items into one, “out of control” and “overwhelming”, which are both already present in the trait-DERS as separate items. This might artificially improve factor solutions by effective doubling of content, which could even favor the removal of this item from all DERS versions entirely.

The second item with substantial cross-loadings was item 7 of the Clarity scale (“I have no idea how I am feeling”), which also loaded on the Awareness scale. We would argue it is almost trivial that the awareness someone has of their current emotions influences whether they have an idea how they are feeling. In fact, a well-powered study on the trait-DERS found that the Awareness and Clarity scales were not distinct (Bardeen et al., 2016). This highlights general questions on the causal structure of the (S-)DERS.

As most previous studies, we started by assuming a common factor model, implying that latent causes (Non-Acceptance, Modulate, Awareness, and Clarity) influence the responses on the individual items. Here, we found that generally no overall factor of “emotion regulation difficulties” can be assumed, based on comparisons to a higher-order and bi-factor model. Hence, a global S-DERS score would be simply an amalgam of four correlated but distinct

factors. Network models are an alternative to such common-cause factor models, which are difficult to compare on the basis of fit measures (Fried, 2020). They represent a (mostly) hypothesis-generating tool to consider potentially causal links between single items. For example, outside its target scale, item 7 of the Clarity scale (“I have no idea how I am feeling”) is most strongly connected to item 6 Awareness scale (“I am acknowledging my emotions”). As mentioned above, recognizing one’s emotions appears as a prerequisite to acknowledging them at face-value, showing a potential causal interpretation of substantial cross-loadings for item 7. Similarly, we found that item 5 might serve as a connector between the Modulate and the Non-Acceptance scale. A hypothesis-generating causal interpretation would be, for example, that feeling very bad about oneself (item 5) may make someone expect increasingly depressed mood (item 15), which makes one feel overwhelmed (item 21) and believing one will continue to feel this way for a long time (item 10).

There are several limitations in our study. First, while our sample composition resembled that of the original S-DERS validation and had a substantial proportion of people with a history of diagnosed mental disorders, this is still a highly educated, mostly young, female sample, imposing constraints on generalizability. Secondly, while our mood induction did increase negative affect, there was no association between affective reactivity and the S-DERS. In turn, there was a substantial association between affect *before* but not after the mood induction and the S-DERS, despite the latter being administered *after* the mood induction. One potential explanation might be that despite the mood induction being successful, its only small-to-moderate effect size was not sufficient to observe an association with reactivity, as participants might fastly revert to their affective set point before the mood induction. Put simply, our mood induction might have been insufficiently strong to overwrite the emotion regulation difficulties participants experienced before the mood induction.

Lastly, while the EFA models largely conformed to the criteria for conceptual replication of the original S-DERS study (with the exception of two slightly larger cross-loadings discussed above), only the SRMR index crossed the preregistered threshold for a stricter test of model fit. These thresholds were determined based on default standards in the field, which has been increasingly shown to have major shortcomings. Fit measures other than SRMR have been shown to lead to over-extraction of factors, underestimation of model fit for small sample sizes, and particularly bad diagnostic performance for determining the number for factors of RMSEA based on simulations (Haslbeck & van Bork, 2024; Hu & Bentler, 1999; Montoya & Edwards, 2021). For our sample size, more recent studies have shown that flexible cutoffs on the SRMR based on simulations have the best performance for model diagnostics

(Mai et al., 2021, Groskurth et al., 2024; McNeish & Wolf, 2023; Niemand & Mai, 2018). When performing such analyses with the FCO package (Niemand & Mai, 2018), we found a cutoff of SRMR = .055, which our ESEM model passes. Generally, we believe that for DERS-like structures, psychometric decisions based on fixed absolute literature-based thresholds might not be a fruitful path. In items and scales with such complex interrelations, it is highly unlikely that a CFA simple structure without any cross-loadings or an ESEM structure with cross-loadings for all items can reflect the true data generating model. In the long run, such formal models of emotion regulation difficulties might be improved based on clinical theory and evidence targeting causal relations.

In conclusion, the translated S-DERS showed a sufficient four-factor solution similar to the original S-DERS version. The S-DERS further showed good reliability and association with related constructs. We therefore recommend using the S-DERS when investigating temporal dynamics of emotion regulation difficulties in laboratory or daily life settings.

References

- Bardeen, J. R., Fergus, T. A., Hannan, S. M., & Orcutt, H. K. (2016). Addressing Psychometric Limitations of the Difficulties in Emotion Regulation Scale Through Item Modification. *Journal of Personality Assessment*, 98(3), 298–309. <https://doi.org/10.1080/00223891.2015.1091774>
- Barrett, L. F., & Russell, J. A. (1999). The structure of current affect: Controversies and emerging consensus. *Current Directions in Psychological Science*, 8(1), 10–14. <https://doi.org/10.1111/1467-8721.00003>
- Borkenau, P., & Ostendorf, F. (2007). NEO-Fünf-Faktoren-Inventar nach Costa und McCrae/NEO-FFI; 2. Neu normierte und vollständig überarbeitete Auflage. *Hogrefe*, 2007.
- Borsboom, D., Deserno, M. K., Rhemtulla, M., Epskamp, S., Fried, E. I., McNally, R. J., Robinaugh, D. J., Perugini, M., Dalege, J., Costantini, G., Isvoranu, A.-M., Wysocki, A. C., van Borkulo, C. D., van Bork, R., & Waldorp, L. J. (2021). Network analysis of multivariate data in psychological science. *Nature Reviews Methods Primers*, 1(1), 1–18. <https://doi.org/10.1038/s43586-021-00055-w>
- Clamor, A., Lincoln, T. M., & Schulze, L. (2024). *Emotion Regulation Difficulties in Mental Disorders: A Systematic Review and Multilevel Meta-Analysis of 25 Years of Questionnaire Research*. OSF. <https://doi.org/10.31234/osf.io/yzuk8>
- Daros, A. R., & Williams, G. E. (2019). A Meta-analysis and Systematic Review of Emotion-Regulation Strategies in Borderline Personality Disorder. *Harvard Review of Psychiatry*, 27(4), 217–232. <https://doi.org/10.1097/HRP.0000000000000212>
- Emery, N. N., & Simons, J. S. (2020). The role of affect, emotion management, and attentional bias in young adult drinking: An experience sampling study. *Psychopharmacology*, 237(5), 1557–1575. <https://doi.org/10.1007/s00213-020-05480-5>

- Epskamp, S., Borsboom, D., & Fried, E. I. (2018). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods*, 50(1), 195–212.
<https://doi.org/10.3758/s13428-017-0862-1>
- Fernandez, K. C., Jazaieri, H., & Gross, J. J. (2016). Emotion Regulation: A Transdiagnostic Perspective on a New RDoC Domain. *Cognitive Therapy and Research*, 40(3), 426–440. <https://doi.org/10.1007/s10608-016-9772-2>
- 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*, 251524592095174. <https://doi.org/10.1177/2515245920951747>
- Forester, G., Schaefer, L. M., Dodd, D. R., Burr, E. K., Bartholomay, J., Berner, L. A., Crosby, R. D., Peterson, C. B., Crow, S. J., Engel, S. G., & Wonderlich, S. A. (2023). Time-of-day and day-of-week patterns of binge eating and relevant psychological vulnerabilities in binge-eating disorder. *International Journal of Eating Disorders*, 56(9), 1694–1702. <https://doi.org/10.1002/eat.23995>
- Fried, E. I. (2020). Lack of Theory Building and Testing Impedes Progress in The Factor and Network Literature. *Psychological Inquiry*, 31(4), 271–288.
<https://doi.org/10.1080/1047840X.2020.1853461>
- Gratz, K. L., & Roemer, L. (2004). Multidimensional Assessment of Emotion Regulation and Dysregulation: Development, Factor Structure, and Initial Validation of the Difficulties in Emotion Regulation Scale. *Journal of Psychopathology and Behavioral Assessment*, 26(1), 41–54. <https://doi.org/10.1023/B:JOBA.0000007455.08539.94>
- Groskurth, K., Bluemke, M., & Lechner, C. M. (2024). Why we need to abandon fixed cutoffs for goodness-of-fit indices: An extensive simulation and possible solutions. *Behavior Research Methods*, 56(4), 3891–3914. <https://doi.org/10.3758/s13428-023->

- Gross, J. J. (2015). The Extended Process Model of Emotion Regulation: Elaborations, Applications, and Future Directions. *Psychological Inquiry*, 26(1), 130–137.
<https://doi.org/10.1080/1047840X.2015.989751>
- Gross, J. J., & Thompson, R. A. (2007). Emotion Regulation: Conceptual Foundations. In *Handbook of emotion regulation* (pp. 3–24). The Guilford Press.
- Gutzweiler, R., & In-Albon, T. (2018). Überprüfung der Gütekriterien der deutschen Version der Difficulties in Emotion Regulation Scale in einer klinischen und einer Schülerstichprobe Jugendlicher. *Zeitschrift Für Klinische Psychologie Und Psychotherapie*, 47(4), 274–286. <https://doi.org/10.1026/1616-3443/a000506>
- Haslbeck, J. M. B., & van Bork, R. (2024). Estimating the number of factors in exploratory factor analysis via out-of-sample prediction errors. *Psychological Methods*, 29(1), 48–64. <https://doi.org/10.1037/met0000528>
- Hoyer, J., & Gloster, A. T. (2013). Psychologische Flexibilität messen: Der Fragebogen zu Akzeptanz und Handeln II. *Verhaltenstherapie*, 23(1), 42–44.
<https://doi.org/10.1159/000347040>
- Hu L.-T. & Bentler P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(July 2012), 1–55.
- Klinitzke, G., Romppel, M., Häuser, W., Brähler, E., & Glaesmer, H. (2012). Die deutsche Version des Childhood Trauma Questionnaire (CTQ) – psychometrische Eigenschaften in einer bevölkerungsrepräsentativen Stichprobe. *PPmP - Psychotherapie · Psychosomatik · Medizinische Psychologie*, 62(02), 47–51.
<https://doi.org/10.1055/s-0031-1295495>
- Lavender, J. M., Tull, M. T., DiLillo, D., Messman-Moore, T., & Gratz, K. L. (2017).

- Development and Validation of a State-Based Measure of Emotion Dysregulation. *Assessment*, 24(2), 197–209. <https://doi.org/10.1177/1073191115601218>
- Mai, R., Niemand, T., & Kraus, S. (2021). A tailored-fit model evaluation strategy for better decisions about structural equation models. *Technological Forecasting and Social Change*, 173, 121142. <https://doi.org/10.1016/j.techfore.2021.121142>
- 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(Mimic), 85–110. <https://doi.org/10.1146/annurev-clinpsy-032813-153700>
- McNeish, D., & Wolf, M. G. (2023). Dynamic fit index cutoffs for confirmatory factor analysis models. *Psychological Methods*, 28(1), 61–88. <https://doi.org/10.1037/met0000425>
- Meule, A., Vögele, C., & Kübler, A. (2011). Psychometrische Evaluation der deutschen Barratt Impulsiveness Scale – Kurzversion (BIS-15). *Diagnostica*, 57(3), 126–133. <https://doi.org/10.1026/0012-1924/a000042>
- Michalak, J., Zarbock, G., Drews, M., Otto, D., Mertens, D., Ströhle, G., Schwinger, M., Dahme, B., & Heidenreich, T. (2016). Erfassung von Achtsamkeit mit der deutschen Version des Five Facet Mindfulness Questionnaires (FFMQ-D). *Zeitschrift Für Gesundheitspsychologie*, 24(1), 1–12. <https://doi.org/10.1026/0943-8149/a000149>
- Miola, A., Cattarinussi, G., Antiga, G., Caiolo, S., Solmi, M., & Sambataro, F. (2022). Difficulties in emotion regulation in bipolar disorder: A systematic review and meta-analysis. *Journal of Affective Disorders*, 302, 352–360. <https://doi.org/10.1016/j.jad.2022.01.102>
- Montoya, A. K., & Edwards, M. C. (2021). The Poor Fit of Model Fit for Selecting Number of Factors in Exploratory Factor Analysis for Scale Evaluation. *Educational and*

Psychological Measurement, 81(3), 413–440.

<https://doi.org/10.1177/0013164420942899>

Niemand, T., & Mai, R. (2018). Flexible cutoff values for fit indices in the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, 46(6), 1148–1172. <https://doi.org/10.1007/s11747-018-0602-9>

Nilges, P., & Essau, C. (2021). *DASS. Depressions-Angst-Stress-Skalen—Deutschsprachige Kurzfassung*. <https://www.psycharchives.org/en/item/5bcda8ff-b672-43e3-a6e8-6cbc598bf37e>

O'Brien, S. T., Dozo, N., Hinton, J. D. X., Moeck, E. K., Susanto, R., Jayaputera, G. T., Sinnott, R. O., Vu, D., Alvarez-Jimenez, M., Gleeson, J., & Koval, P. (2024). SEMA3: A free smartphone platform for daily life surveys. *Behavior Research Methods*, 56(7), 7691–7706. <https://doi.org/10.3758/s13428-024-02445-w>

Pavlicic, J. M., Witcraft, S. M., Dixon, L. J., Buchanan, E. M., & Schulenberg, S. E. (2022). Daily changes in state emotion regulation abilities predict positive and negative affect. *Journal of Contextual Behavioral Science*, 26, 63–68. <https://doi.org/10.1016/j.jcbs.2022.08.006>

Prokofieva, M., Zarate, D., Parker, A., Palikara, O., & Stavropoulos, V. (2023). Exploratory structural equation modeling: A streamlined step by step approach using the R Project software. *BMC Psychiatry*, 23(1), 546. <https://doi.org/10.1186/s12888-023-05028-9>

Pruessner, L., Holt, D. V., Gözl, R., Sevcenko, N., Hofmann, S. G., & Backenstrass, M. (2020). Psychometrische Eigenschaften der deutschsprachigen Version des Interpersonal Emotion Regulation Questionnaire. *Diagnostica*, 66(1), 62–73. <https://doi.org/10.1026/0012-1924/a000239>

Reid, Mj., Omlin, X., Espie, Ca., Sharman, R., Tamm, S., & Kyle, Sd. (2022). *The effect of sleep continuity disruption on multimodal emotion processing and regulation: A*

laboratory-based, randomized, controlled experiment in good sleepers.

<https://doi.org/10.1101/2022.04.22.489209>

- Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48, 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Sousa, R., Linharelhos, M., Silva, D. R. D., & Rijo, D. (2023). The difficulties in emotion regulation scale: Dimensionality, measurement invariance, and physiological correlates in community adolescents. *Journal of Clinical Psychology*, 79(10), 2351–2363. <https://doi.org/10.1002/jclp.23553>
- Stellern, J., Xiao, K. B., Grennell, E., Sanches, M., Gowin, J. L., & Sloan, M. E. (2023). Emotion regulation in substance use disorders: A systematic review and meta-analysis. *Addiction*, 118(1), 30–47. <https://doi.org/10.1111/add.16001>
- Tiede, T. (2020, July 14). *How did I feel? Recalling reported core affect in light of its fluctuation, the present state, and individual degrees of neuroticism* [Info:eu-repo/semantics/masterThesis]. University of Twente. <https://essay.utwente.nl/82569/>
- Vuong, Q. H. (1989). Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica*, 57(2), 307–333. <https://doi.org/10.2307/1912557>
- Weathers, F. W., Blake, D. D., Schnurr, P. P., Kaloupek, D. G., Marx, B. P., & Keane, T. M. (2013). *The Clinician-Administered PTSD Scale for DSM-5 (CAPS-5)*. www.ptsd.va.gov.
- Werthmann, J., Renner, F., Roefs, A., Huibers, M. J. H., Plumanns, L., Krott, N., & Jansen, A. (2014). Looking at food in sad mood: Do attention biases lead emotional eaters into overeating after a negative mood induction? *Eating Behaviors*, 15(2), 230–236. <https://doi.org/10.1016/j.eatbeh.2014.02.001>
- Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample Size Requirements for Structural Equation Models: An Evaluation of Power, Bias, and

Solution Propriety. *Educational and Psychological Measurement*, 76(6), 913–934.

<https://doi.org/10.1177/0013164413495237>