We first like to thank the editor and reviewers for their time and consideration of the manuscript. Most importantly, the comments helped us communicate more clearly in the manuscript why this paper's impact goes beyond a translation/validation of a measure and provides important conceptual additions to the literature. For our specific responses to the reviewers' comments, see below.

Reviewer 1, Comment 1:

"There are many strengths associated with this manuscript. The study is generally competently executed. The preregistration is impressive. The data analyses fit the research questions well and the authors are open about deviations"

We wanted to specifically thank the reviewer for this kind comment, as such comments really increase our motivation to provide comprehensive preregistrations and openly communicate deviations in the future.

Reviewer 1, Comment 2:

"I think the authors need to bolster the rationale and significance of this research better. Typically, translations of measures into a new language is a rather modest contribution unless there is a major cross-cultural element. It would be helpful to get a better sense of how this study increments the broader literature on the DERS/S-DERS beyond examining a new language translation"

Response:

We thank the reviewer for this helpful comment! Before we show where and why we made changes in the manuscript addressing this comment, we want to briefly summarize we believe our study has impact beyond the translation/validation and opted to submit to broader English-language journals:

The main reason why we believe our study goes beyond a simple translation and conceptual replication of the original study are: The preregistered comparison between different important conceptual models. These models show that (1) there is no general "g-factor" of state-based emotion regulation difficulties, while (2) there is still overlap between the four factors in terms of cross-loadings, which (3) could be potentially accommodated within a network-framework, for which we additionally provided an exploratory solution.

Even already in the abstract, we see how this was not clearly presented in the paper. We therefore tried to make the distinction between the translation/validation aspects and these novel model comparisons more clear throughout the manuscript.

Old last sentence of the abstract Introduction:
"The present study aimed to translate the S-DERS into German, validate its psychometric properties, and explore the interrelations among its subcomponents using extended factor models and cross-sectional network analysis"

New version:
"The present study aimed to (a) translate the S-DERS into German, (b) validate its psychometric properties, and (c) provide novel preregistered examinations whether state-based emotion regulation difficulties share an underlying general factor or are interconnected but distinct sub-components, complemented with an exploratory network approach."

Rationale:

The a/b/c structure might make it more clear that the paper has three interconnected goals, which build on each other, leading from the basic work (a/b) to the novel work (c). We explicitly state that these model comparisons are novel to make it clearer to the reader this goes beyond a simple validation/replication of previous English results. Moreover, the word "explore" in the old version did not reflect well that these were preregistered substantive questions. Furthermore, we now state more directly what the conceptual gain of these models is.

Old last two sentences in the abstract methods section:

"Factor structure, reliability, and construct validity were examined using confirmatory factor analysis and exploratory structural equation modeling (ESEM) with cross-loadings. These were compared to bi-factor, higher-order, and network models. Additionally, network analysis was employed to generate hypotheses about pathways linking discrete emotion regulation difficulties"

Replacement of the last sentence:

"These were compared to bi-factor, higher-order, and network models"

Rationale:

Clearer focus and explicit mentioning of the relevant conceptual models we compared.

Further, we added to the abstract results and conclusion for more clarity (changes in italics):

"A four-factor correlated model outperformed both bifactor and higher-order models, suggesting that emotion regulation difficulties are best conceptualized as four distinct but interrelated constructs *without a shared general factor*"

"The *extended factor models* ~~findings~~ highlight the multidimensional nature of emotion regulation difficulties, with complex interrelations among distinct but related constructs.

In the introduction, we made many changes to address this comment. As for the abstract, they aim to make the distinction between (a) the basic translation/validation and (b) the novel models concerning a general factor more explicit. We also try to explain better why these analyses are both novel and important. The changes entail restructuring and new text segments, making it difficult to show changes clearly in the response letter. Therefore, we hope it is acceptable for the reviewers to view these changes in the manuscript version with tracked changes (which was provided additionally to the untracked version).

No changes were made to the methods section.

The results section after the manipulation check "Factor structure" was replaced with "Basic factor structure and extended higher-order models" to make it clearer that novel factor models can be found in this section.

Similar to the introduction, we made comprehensive changes to the discussion. First, we now focus right at the beginning on the novel conceptual benefit which can be derived from our study, based on the extended factor models. We tried to make these points less technical and emphasize the added conceptual understanding gained from these models. Second, we removed a highly technical paragraph on the results of the parallel analysis, which was basically only a reiteration of statements in the results section, which readers might be more likely to look for there. Third, we moved detailed discussions on the meaning of high cross-loadings for two items to the limitations, as this part as well might be too detailed/technical for the conceptual discussion of results. Lastly, we changed the conclusion section to strongly focus on the conceptual knowledge gained concerning emotion regulation difficulties as a whole, rather than psychometric detail.

As these were many changes, we again hope it's acceptable and more accessible for the reviewers to check them as tracked-changes in the manuscript.

Reviewer 1, Comment 3:

"The authors suggest that n=200 is sufficient for most CFA models but I think that reflects a very simplistic read of the simulation study they cite. And there are several studies on this topic of power. I am not someone who believes massive sample sizes are needed for CFA as I agree that simulation studies do tend to support smaller n's than what general lore would suggest, but I think power considerations could be made specifically for the pre-registered models using Monte Carlo simulations rather than sweeping statements"

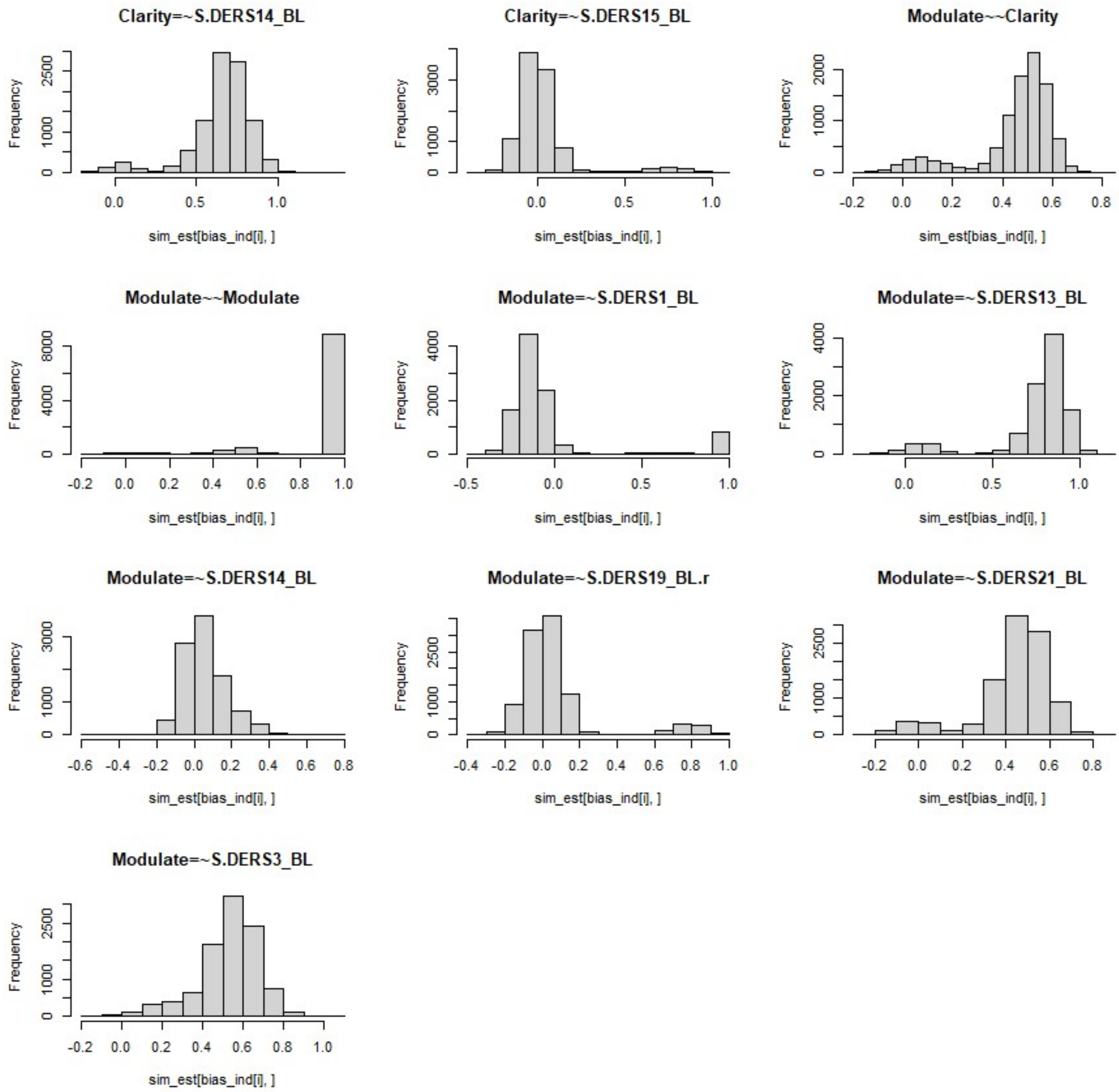
Response:
We thank the reviewer for this suggestion, which motivated us to learn how to perform Monte-Carlo sample size simulations for SEM. Using the criteria by Wolf et al. (which are also the standard for Mplus), our simulations showed that sample size is sufficient for CFA without cross-loadings as well as more complex models like the bi-factor model. We added the following to the methods section of the manuscript:

"Custom-simulations using the simsem package showed that a simple structure CFA based on the factor structure of Lavender and colleagues (2017) is unbiased for our sample size using the criteria of Wolf and colleagues (2013; i.e., relative bias in coefficients and standard errors below 0.05 and confidence interval coverage of the population parameter in at least 90% of 10,000 iterations). These criteria were also fulfilled for a more complex bi-factor model, simulated from a population model with general- and specific-factor loadings of .50, except that standard errors are slightly underestimated, albeit to a degree which maintained confidence interval coverage above the acceptable threshold (median relative standard error bias = .06, range -0.12-0.01)"

We also tried to perform these simulations for the ESEM model, but there were simulation-issues which we could not resolve. In case the reviewer is interested concerning details of these issues, we describe them below.

First, we believe the binary bias criteria (≤ 0.05) are not appropriate for loadings. The bias in factor loadings is calculated as: (population value – average simulated estimate)/population value. This value is supposed to be below .05 (sometimes also .10 is suggested), corresponding to less than 5% bias. When we implemented this, some biases were unreasonably high, reaching a value over 22000% for one specific loading. Inspecting these results, we saw this enormous relative bias was only due to a difference between population loading and observed loading of .055. This discrepancy in relative and absolute bias appeared, because the population value (and therefore denominator of relative bias) was already negligibly small (-0.000227386) with an average simulated estimate of 0.055. Another loading showed the same discrepancy between population value and average simulated estimate around 0.055, but only had a relative bias of 7%, because this loading was high (.70), as it connected an indicator with its intended factor and the population loading. Therefore, we believe the <5% criteria by Wolf et al. and Muthen & Muthen might not be appropriate for ESEM applications where population loadings are usually very small for many cross-loadings. We did not find a reference to solve this specific issue.

Another problem we could not solve was factor switching. While the average difference between population parameters and their estimates was negligibly small (0.002), we still observed larger bias (>.05) for some estimates. Plotting the estimates of all 10000 iterations for these "biased problem cases" showed this is likely also due to factor switching (see plot below). In contrast to CFA without cross-loadings, the ESEM model does not know "which factor is which". E.g., Factor 1 can be "Non-Acceptance" in one iteration and "Modulate" in another iteration. We tried to automatically assign the correct factor labels in each iteration using referent items (i.e., the items which load maximally on their intended factor and minimally on other factors according to the population model). Still, this issue continued to occur, even when removing improper SEM solutions and is apparent as the smaller second peak in the bimodal distributions shown below, which should be unimodal even for biased models. We already had similar issues in PCA simulations for a different project, which we could not resolve. As can be seen below, the most problematic cases were for the Modulate items, likely, because this factor had the highest correlation with other factors. Therefore, factor assignment remains incorrect in a substantial portion of simulated iterations, creating spurious bias and coverage estimates.



Dziak, J. J., Dierker, L. C., & Abar, B. (2020). The interpretation of statistical power after the data have been gathered. *Current Psychology*, *39*, 870-877.

Reviewer 2, comment 1:

“One major concern I have with this study is its contribution to the literature. It does not really provide meaningfully new information. It is essentially demonstrating that a German translation of the SDERS obtains closely similar results for the SDERS reported in English language studies. It is not uncommon for researchers to demonstrate that a translation of a measure performs equally well but that has been so frequently demonstrated with so many measures that it’s not really clear that these translation validation studies are really that necessary or important. It would be useful for the authors to demonstrate why there might be some issues or concerns with respect to a German translation; otherwise it just seems obvious that it will work well.”

Response:

We do believe the reviewer raises a very important point here, which was also the main comment of reviewer 1. Through these comments, we realized that we did not convey well why we believe our paper adds to the current literature beyond a translated version, which we do believe and was the reason for submitting this paper to an English-language journal. In short, the extended factor models are what provides additional conceptual value to the literature and we made extensive changes to the manuscript (mainly introduction and discussion) to convey why this is the case. Due to the overlap with the criticism of reviewer 1, we would like to refer to our response to “Reviewer 1, Comment 2” for details on our general argument and changes made to the manuscript.

Reviewer 2, comment 2:

“The sample is also disappointing, being a college student sample with few clinical issues: 88% had no history of any clinical treatment and 76% had not history of a clinical diagnosis.”

Response:
We certainly acknowledge that a more diverse sample would have been desirable, especially concerning age, gender, ethnicity, and socioeconomic status. As stated in the manuscript, we believed this sample composition to be acceptable, as it strongly resembled the original S-DERS construction sample. We were happy to generally find compatible results with Lavender et al. (2017) and believe such a first validation is an important basic step on which the other analyses rely. Still, if our results would not have matched theirs, there would have been two possible explanations for this: (1) Lavender et al. generally does not replicate, which would be reasonable given they used only one sample for questionnaire construction (from a larger item pool) and validation or (2) our translation is insufficient. If we had used a clinical sample, there would have been another explanation: (3) Lavender et al. does not generalize to other (e.g. clinical) populations (which has been observed previously for the trait DERS in some cases, e.g., Gutzweiler & In-Albon, 2018). Therefore, we wanted to avoid a situation in which we find non-correspondence between our version and Lavender et al. that can simply be explained by a lack of generalizability to clinical samples, potentially inherent to any S-DERS version, but making it seem as if the translation failed. As a result, we focused on a sample composition similar to theirs with the clear disadvantage of limitations on generalizability, as we state in the manuscript.

Overall, concerning the proportion of participants with a clinical history, we believe that it is reasonable to first perform such analyses on a mixed sample, before focusing on specific clinical populations, and that 25% with a history of clinical diagnosis might be a meaningful proportion for a dimensional perspective on mental well-being, being similar to life time prevalence of mental disorders in the general population (Kessler et al., 2007; especially as age is lower).

Kessler, R. C., Angermeyer, M., Anthony, J. C., De Graaf, R. O. N., Demyttenaere, K., Gasquet, I., ... & Üstün, T. B. (2007). Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization's World Mental Health Survey Initiative. *World psychiatry*, *6*(3), 168.

Reviewer 2, comment 3:

“A novel aspect of the study was a negative mood induction prior to completion, but this could also be suggesting that the negative mood was not actually characteristic of the participants, instead being an artifactual induction”

Response:

We agree with the reviewer. The S-DERS, as validated here, was originally constructed and validated for experimental mood inductions by Lavender et al. (2017), with two characteristics: (1) state-based emotion regulation difficulties need an appropriate context to occur and the induction of negative emotions makes this more likely. (2) the questionnaire was only validated for experimental settings. We do believe that validations across different naturalistic settings is important and include this point more explicitly now in the conclusion:

“Taken together, these findings underscore the importance of moving beyond static trait assessments to study the fluid nature of emotion regulation difficulties. Future research should further explore these dynamics beyond experimental settings, using intensive longitudinal designs in daily life, integrating both factor-analytic and network approaches to capture the full complexity of state-based emotion regulation.”