**Reviewers’ comments and suggestions**

**Reviewer #1’s comment**

We thank the reviewer for their time and effort in reviewing our manuscript, and the positive critique. We have responded to all of the reviewer’s comments below.

**(RC#1.1):**

Please define "ecological momentary assessment" when it is first introduced in the introduction.

**Authors’s reponse (AR#1.1):** We added the definition of EMA in the Introduction, in the first sentence in which this term was used:

*Through Ecological Momentary Assessments (EMA), they explored the relationship between state self-compassion and well-being, distinguishing between the CS and UCS components, where EMA involves repeated, real-time data collection in participants’ natural environments.*

**RC#1.2:** I found the statement that the current study improves on existing EMA studies in that prior study durations are shorter (< 7 days) to be not particularly compelling, especially since the studies presented here only include 10 and 16 days (albeit over longer periods). Why does having 3-9 more days make a big difference in "capturing the full complexity of state self-compassion dynamics"? More broadly, the sampling design of the current study is unusual (1 day per week over several months); a rationale should be provided, along with consideration of the strengths and limitations of this approach.

**AR#1.2:** We appreciate the reviewer’s comment. We have addressed these points by adding the following text to the Introduction of the revised manuscript.

*Despite advancements in understanding the temporal dynamics and variability of self-compassion, significant methodological challenges persist. Many previous studies have relied on ad hoc measures, raising questions about the validity of state-level assessments. Moreover, these studies often focus on short time frames, such as seven days (e.g., Mey et al., 2023; Sahdra et al., 2023), which may fail to capture the full complexity of state self-compassion. The use of randomly selected time windows also risks overlooking critical life events that could substantially influence self-compassion dynamics.*

*To address these limitations, our study employs a three-month EMA protocol using the validated State Self-Compassion Scale (Neff, 2022). This approach represents the first investigation of the Bipolar Continuum Hypothesis within an extended EMA framework. By collecting data across multiple levels—moments, days, and individuals—our design provides a nuanced and naturalistic examination of state self-compassion. Additionally, we explore how significant life events, such as academic exams, shape self-compassion dynamics among university students.*

*Our study design, which includes one day of notifications per week with five prompts per day, contrasts with the more intensive protocols often used, such as five daily notifications over a single week. The extended time frame allows for greater variability in state self-compassion, capturing shifts associated with major life events. Unlike a randomly chosen week, our design intentionally incorporates two academic exams, enabling a comparison of self-compassion dynamics during periods proximal and distal to these stressors. Furthermore, the reduced frequency of notifications helps mitigate participant fatigue (Shiffman, Stone, & Hufford, 2008), thereby enhancing both data quality and participant engagement.*

**RC#1.3:** The first hypothesis on p. 6 was very vague ("...expected to display pronounced temporal dynamics”) -- what would this mean, and how would it be operationalized? Isn't this already known from prior EMA studies?

**AR#1.3:** We agree that, in the previous submission, the hypotheses presented in the Introduction should have been clearer. In the revised manuscript, we have changed the relevant part of the Introduction as follows:

*Building on Ferrari et al.'s (2022) conceptualization of self-compassion as a dynamic process, we formulate four specific hypotheses to test the idea of a bipolar continuum in self-compassion using an EMA protocol.*

***H1:*** *A purely cross-sectional psychometric analysis conducted at a single time point may be insufficient to determine the dimensionality of self-compassion, as multiple CFA models with distinct theoretical implications can yield comparable goodness-of-fit indices (e.g., Bifactor ESEM vs. Correlated Two-Bifactor ESEM; Neff et al., 2019). In contrast, state-level CS and UCS are expected to exhibit distinct temporal dynamics, reflecting moment-to-moment regulatory processes in which increases in one component correspond to decreases in the other, depending on the context. These temporal fluctuations provide a richer and more nuanced understanding of the underlying structure of self-compassion. To rigorously test the Bipolar Continuum Hypothesis, we will examine the reliability and the factor structure of state self-compassion using EMA data analyzed through multilevel CFA models, capitalizing on the within-person temporal dynamics captured by this approach.*

***H2:*** *The Bipolar Continuum Hypothesis posits that the valence of contextual situations should have equal, opposite, and symmetric effects on the two components of self-compassion (CS and UCS) when examined separately. Each component can be modeled independently using a hierarchical Bayesian regression framework, with predictors representing contextual valence dimensions. If the hypothesis holds, we expect the regression coefficients for CS and UCS to be of comparable magnitude but opposite in sign, reflecting their inherent interdependence within the bipolar continuum.*

***H3:*** *Contextual stressors or heightened negative affect may intensify the bipolar relationship between CS and UCS by activating self-regulatory mechanisms that sharpen the emotional distinction between these two components (Dejonckheere et al., 2021). Research suggests that emotionally salient events, which bring central personal concerns to the forefront, can amplify affective polarization. In such instances, positive and negative emotional states become more mutually exclusive, functioning as an adaptive mechanism to emphasize the event's importance and direct attention toward appropriate responses. This increased bipolarity helps allocate cognitive and emotional resources to evaluate success or failure concerning the concern, streamlining emotional processing to support effective behavioral responses. Building on this phenomenon, we will examine the association between CS and UCS before and after participants encounter a salient event, with varying levels of emotional arousal (high vs. low). According to the Bipolar Continuum Hypothesis, the relationship between CS and UCS should remain stable regardless of fluctuations in emotional arousal, reflecting the inherent balance within the continuum.*

***H4:*** *The Bipolar Continuum Hypothesis predicts a consistently negative relationship between CS and UCS at the individual level. Any deviations, such as a zero or positive correlation between these components for certain individuals, would directly challenge the validity of the Bipolar Continuum Hypothesis, calling its assumptions into question.*

**RC#1.4:** While I think the additional analyses of structure and idionomic analyses strengthen the paper, I found their presentation to be lacking in detail and presented ineffectively. They were simply tacked on to the end of the paper, after the Study 2 discussion, without even an indication of which sample was analyzed and no foreshadowing in the study set up, which was quite odd and confusing. Instead, they should be integrated into the relevant study from the beginning, set up in the introduction, included in hypotheses, described in data analysis, etc., just like any other analysis would be. In addition, there is so little detail provided in the main text as to be mostly uninformative, especially for the multilevel CFA. It is essential to describe and consider factor loadings (particularly in the bifactor solution, which are notable), specific values of fit indices, factor correlations (i.e., the two factors are so strongly correlated so as to suggest redundancy, even if they give better fit), and consideration of indices such as omega hierarchical in the bifactor analysis. Right now, important information is overlooked and the reader cannot understand the main points without reading all of the supplemental materials. For the idionomic analyses, there is not a single statistic reported in the main text to support the conclusions. While I understand there are space limitations, I think it is essential to include sufficient detail in the main text for the reader to understand and evaluate each analysis, and I think other sections could likely be reduced.

**AR#1.4:** We agree that the manuscript would benefit from a revision according to the above suggestions. We have addressed these points as follows. (1) We integrated the multilevel CFA and the idionomic analysis in the Introduction:

In previous research, hypotheses like those described above are often evaluated using multilevel models. However, these models have significant limitations that may obscure the temporal dynamics and individual variability inherent in EMA data. One key limitation is their reliance on the assumption of psychological homogeneity, which often treats individual differences as statistical noise or unexplained variance rather than as meaningful information (Sahdra et al., 2024). Additionally, multilevel models tend to aggregate data toward group-level trends, effectively "smoothing out" individual-level heterogeneity and potentially misrepresenting the nuanced within-person associations that unfold over time. This issue becomes particularly problematic when the ergodic assumption—that group-level effects accurately reflect individual-level processes—is violated, leading to a distorted understanding of the underlying dynamics.

To address these limitations, we will complement multilevel model analyses with an idionomic approach (Hayes et al., 2022). This approach prioritizes the modeling of idiographic patterns – those unique to individuals – before attempting to generalize findings to nomothetic (group-level) patterns. Importantly, only nomothetic generalizations that provide incremental value to our understanding of idiographic insights are retained (Ciarrochi et al., 2024; Ferrari et al., 2022; Sahdra et al., 2024; Hayes & Hofmann, 2021; Ciarrochi et al., 2022).

The multilevel CFA and the idionomic analysis have been added to the hypotheses of the study as indicated in the previous point. We have revised the Data Analysis section by adding the following text:

Hypothesis H4 was tested using a two-stage idionomic analysis. First, an idiographic approach employed a hierarchical Bayesian model in Stan to estimate within-person relationships between UCS and CS for each participant. Second, a nomothetic approach used a hierarchical model in the brms package (R) to identify group-level patterns and quantify variability in UCS–CS associations (e.g., Sahdra et al., 2024). While traditional multilevel modeling separates within- and between-person variance, it can underestimate individual-level heterogeneity due to shrinkage toward the group mean. In contrast, the idionomic framework integrates idiographic and nomothetic perspectives, preserving individual variability while identifying shared patterns. This approach provides a nuanced understanding of individual differences and contextual dynamics in EMA studies.

A greater number of details of the results of the multilevel CFA analysis are now present in the revised manuscript. Specifically, we have added Table 1 with the loadings of the multilevel bifactor model and Table 2, with the Goodness-of-Fit Indices for CFA models. In the revision, we have included the inter-factor correlation for the two-factor model:

Factor intercorrelations in the Two-Factor Model were 0.437 (SE = 0.009) at the within level and 0.720 (SE = 0.059) at the between level.

In the revision, we have also added the hierarchical omega for the multilevel bifactor model:

The hierarchical ω values revealed that the general factor explained more variance than the specific factors at both levels. Specifically, at the within level: ωh,gen,w = 0.26, ωh,cs,w = 0.10, ωh,ucs,w = 0.04; and at the between level: ωh,gen,b = 0.51, ωh,cs,b = 0.07, ωh,ucs,b = 0.21.

In the revision, more clear information about the sample size is provided in the Data Analysis section:

Hypothesis H1 was tested using multilevel structural equation modeling (SEM) with maximum likelihood estimation in Mplus, employing a two-level (within and between) approach. Data from both studies were combined to enhance sample size and improve parameter precision in confirmatory factor analyses.

This information is also provided in the introduction to the section “Multilevel Dimensionality Analysis”:

Building on prior evidence for the Bipolar Continuum Hypothesis in trait self-compassion, we assessed the dimensionality of state self-compassion using multilevel confirmatory factor analysis (H1) applied to the combined datasets from both studies.

Concerning the idionomic statistical analyses, in the revised manuscript, information about the results of the statistical tests is provided in the section “Results of the Idionomic Analysis”.

**RC#1.5:** Related to point #4, it is striking that there are no tables reporting results and just a single figure in the main text. While I appreciate the detail in the online supplement, I think too much has been offloaded there and that the authors should report results for the main analyses in tables or figures in the main text.

**AR#1.5:** We agree with the reviewer. In the revised manuscript, we have added two more tables related to the multilevel CFA analysis.

***Table 1***

*Standardized Factor Loadings for Multilevel Bifactor Model.*

|  |  |  |
| --- | --- | --- |
| *Item* | *Within-level Loadings* | *Between-level Loadings* |
| *CS1* | *0.614* | *0.86* |
| *UCS1* | *0.579* | *0.774* |
| *CS2* | *0.327* | *0.41* |
| *UCS2* | *0.49* | *0.697* |
| *UCS3* | *0.585* | *0.819* |
| *CS3* | *0.583* | *0.781* |
| *CS4* | *0.623* | *0.935* |
| *UCS4* | *0.21* | *0.143* |
| *CS1 (CS)* | *0.219* | *0.215* |
| *CS2 (CS)* | *-0.313* | *-0.329* |
| *CS3 (CS)* | *-0.009* | *-0.073* |
| *CS4 (CS)* | *0.249* | *0.322* |
| *UCS1 (UCS)* | *0.473* | *0.626* |
| *UCS2 (UCS)* | *0.291* | *0.378* |
| *UCS3 (UCS)* | *0.31* | *0.422* |
| *UCS4 (UCS)* | *0.208* | *0.288* |

***Table 2***

*Goodness-of-Fit Indices for CFA Models.*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| *Model* | *CFI* | *TLI* | *RMSEA* | *SRMR (Within)* | *SRMR (Between)* | *AIC* | *BIC* |
| *One-factor* | *0.935* | *0.91* | *0.05* | *0.033* | *0.058* | *475881.5* | *476195.3* |
| *Two-factor* | *0.972* | *0.959* | *0.034* | *0.024* | *0.05* | *474810.2* | *475139.6* |
| *Bifactor* | *0.987* | *0.971* | *0.029* | *0.016* | *0.031* | *474376.7* | *474816.0* |

For space constraints, we were not able to add more figures. However, we have thoroughly revised the Supplementary Material, and several new figures have been added there.

**RC#1.6:** For the finding that a subset of people had positive associations between UCS and CS, one methodological explanation could be careless responding/response styles (e.g., selecting the same response for all or nearly all items would lead to a positive correlation that is not substantively meaningful). Could the authors evaluate whether this may explain the finding? Please also describe the procedures they used to detect any careless or otherwise invalid responses, which are particularly likely in EMA assessments and require identification and removal.

**AR#1.6:** To ensure data quality, we conducted a thorough assessment to identify inattentive or insufficient effort responses. Metrics included compliance rates, survey completion times, and response variability indices (e.g., Longstring Index, Intra-Individual Response Variability). Participants with compliance rates below 50% were excluded. Momentary lapses in engagement were identified using occasion-level analyses, which flagged <2% of occasions for potential inattention. Reanalysis of flagged data showed no substantial impact on primary outcomes, confirming the robustness of the dataset.

Data quality, and the presentation of the results of the statistical analyses that have been conducted, has been discussed in the sections “Data Quality Management and Subject Selection”, for both Study 1 and Study 2, provided in the revised Supplementary Material.

In the revised manuscript, in the Data Analysis section, we have added the following text:

To ensure data quality, we conducted a thorough assessment to identify inattentive or insufficient effort responses. Metrics included compliance rates, survey completion times, and response variability indices (e.g., Longstring Index, Intra-Individual Response Variability). Participants with compliance rates below 50% were excluded. Momentary lapses in engagement were identified using occasion-level analyses, which flagged <2% of occasions for potential inattention. Reanalysis of flagged data showed no substantial impact on primary outcomes, confirming the robustness of the dataset (for details, see SI).

We thank the reviewer for raising the important point concerning a possible alternative explanation for the positive associations between UCS and CS, which could be due to careless responding/response styles. To address this point, we have added the following text in the revised manuscript:

***Potential Response Bias.*** *To address whether response bias could explain the positive associations between CS and UCS observed in some participants, we analyzed indices of careless responding, including the Longstring Index, Intra-Individual Response Variability (IRV), Even-Odd Inconsistency Index, Mahalanobis Distance, and time to completion. These metrics were compared across participants with positive UCS-CS associations and those with neutral or negative associations, using data combined from both studies.*

*The analysis found no credible evidence that response biases accounted for the positive UCS-CS associations. Bayesian multilevel models accounted for the nested data structure, and posterior estimates indicated no reliable differences across indices between the groups (see SI for details). This suggests that the unexpected patterns are unlikely to be due to careless or insufficient effort responding.*

The revised Supplementary Material provides the complete details of the statistical analyses that have been performed.

Thank you.

**Reviewer #2:**

We thank the reviewer for their time and effort in reviewing our manuscript, and the positive critique. We have responded to all of the reviewer’s comments below.

Overall, this revision seems good.  There is one thing that is problematic, however.  In several places the authors describe Neff as having a unidimensional view of self-compassion.  For instance, on page 33 the authors refer to "Neff’s conceptualization of self-compassion as a singular, bipolar construct (Neff, 2022, Neff, 2023)."  This is not an accurate characterization of Neff's model. In Neff's publications, self-compassion is described as a multi-dimensional rather than singular construct.  Saying that CS and UCS operate in tandem along a bipolar continuum is not the same as saying that self-compassion is unidimensional.  Rather, Neff argues that "self-compassion represents a dynamic system in which the various elements of self-compassion are in a state of synergistic interaction" (Neff, 2016a, 2016b, Neff & Toth-Kiraly, 2022), and that this system is best conceptualized as moving along a continuum from UCS to CS (Neff, 2023).  This needs to be clarified throughout the manuscript. Rather than setting up a simplistic straw man argument, the authors would be better served by discussing the difficulties of measuring dynamic systems, the pluses and minuses of using a total score rather than two (including loss in variance when using CS and UCS separately which greatly reduces predictive power), and the challenges and complexities of trying to use statistical modeling to capture messy real life processes.

**AR#2.1:** We thank the Reviewer for pointing out this aspect that was necessary to clarify. In the revision, we addressed this point by changing several sections of the manuscript.

In the Introduction, we state:

This perspective presents self-compassion as a multidimensional and dynamic system in which CS and UCS interact synergistically to regulate emotional well-being (Neff, 2022, 2023).

In the section “State Self-Compassion Dynamics: Partial Evidence for the Bipolar Continuum Hypothesis” of the Introduction, we write:

This perspective presents self-compassion as a multidimensional and dynamic system in which CS and UCS interact synergistically to regulate emotional well-being (Neff, 2022, 2023). Supporting this view, psychometric research demonstrates that the Self-Compassion Scale captures both a global self-compassion factor and six specific subfactors (Neff et al., 2017, 2021).

In the General Discussion:

This study investigated the Bipolar Continuum Hypothesis by examining the dynamic relationship between CS and UCS in real time, drawing on Ferrari et al.’s (2022) conceptualization of self-compassion as a dynamic, multidimensional, and adaptive system. Recognizing that CS and UCS represent distinct yet interrelated components of a bipolar continuum, we tested four hypotheses addressing their temporal dynamics, contextual modulation, and individual variability. By framing self-compassion as a multidimensional system, we investigated whether CS and UCS respond asymmetrically or synergistically to contextual influences, while maintaining their interdependence.

and

A key challenge is reconciling the theoretical multidimensionality of self-compassion with measurement approaches.

and

This study highlights the need for balancing simplicity and nuance when investigating dynamic constructs like self-compassion. While total scores on the Self-Compassion Scale offer practical utility, they may overlook the multidimensional and context-sensitive nature of self-compassion. Approaches integrating the simplicity of total scores with the depth of multidimensional analyses could provide a more robust and theoretically aligned understanding of state self-compassion.

Concerning “difficulties of measuring dynamic systems, the pluses and minuses of using a total score rather than two (including loss in variance when using CS and UCS separately which greatly reduces predictive power), and the challenges and complexities of trying to use statistical modeling to capture messy real life processes” we have added the following sections to the revised General Discussion.

Beyond measurement, this study underscores the challenges of modeling self-compassion as a dynamic, context-sensitive construct. Real-life processes are shaped by a complex interplay of traits, situational factors, and moment-to-moment regulatory mechanisms. While EMA is a powerful tool for capturing these dynamics in naturalistic settings, it introduces methodological complexities. Temporal modeling requires addressing issues such as autocorrelation, lagged effects, and within-person variability—factors that complicate interpretation, increase computational demands, and necessitate larger sample sizes to ensure robust statistical power.

Later on:

This study highlights the need for balancing simplicity and nuance when investigating dynamic constructs like self-compassion. While total scores on the Self-Compassion Scale offer practical utility, they may overlook the multidimensional and context-sensitive nature of self-compassion. Approaches integrating the simplicity of total scores with the depth of multidimensional analyses could provide a more robust and theoretically aligned understanding of state self-compassion.

Thank you.