Lisa Peeters
Doctoral Researcher
Methodology of Educational Sciences
KU Leuven
Faculty of Psychology and Educational Science
Andreas Vesaliusstraat 2 - box 3762
3000 Leuven
Belgium

Tel: +32 472 68 16 31
E-mail: lisa.peeters@kuleuven.be
22/11/2024

Manuscript ID: #894

Dear Professor Evans, Dr. Silverstein and anonymous reviewer,

Thank you for your valuable and constructive feedback regarding our Stage 1 manuscript titled "Mapping methodological variation in experience sampling research from design to data analysis: A systematic review", submitted for review to PCI Registered Reports. Your positive opinion of this work was very encouraging and your insightful comments allowed us to further improve the manuscript.

Please find attached the revised version of our Stage 1 manuscript with changes highlighted in red. Since the initial submission of this manuscript, we have included an additional author, Dr. M. Annelise Blanchard, who will be involved in the data extraction and the data synthesis for this review. In addition to the revised manuscript, we also address each of the recommender's and reviewers' comments below.

Comments from the recommender:

Comment #1: *Based upon reviews of other open practices I have contributed to, my main concerns surround the role of poor reporting standards and the implications this can have for the impact of the project, however I can see this is acknowledged in the manuscript and that you have conducted a robust pilot to negate these.*

Response: Indeed, it is very likely that poor reporting standards will limit the data that we will be able to collect regarding methodological practices. We are grateful for your acknowledgement of our efforts to deal with this limitation, both by assuring through the pilot that our data extraction is equipped to handle missing information, and by openly discussing this problem in our manuscript. This 'problem' is exactly the reason why we think the double aim of our work (describing practices and assessing transparency) is appropriate: systematically reviewing methodological practices without a critical discussion of reporting practices would paint only half a picture.

Comments from reviewer 1 (Dr. Silverstein):

R1 comment #1: *Abstract: the abstract currently feels quite long, and could be made more concise. It might be helpful to at this point add placeholders for certain information that you do not yet have (e.g. the number of included records, a summary and an explanation of the main result, and the implications of the review) as I'm not sure how much editing is permitted at Stage 2 (feel free to ignore if the abstract is free to be edited as much as desired at Stage 2).*

Response: Thank you for your feedback. At your suggestion, we have re-structured and shortened the abstract, which now follows an aim-methods-results-implications structure. The sections 'results' and 'implications' are left empty at this stage, as they will be added at Stage 2.

R1 comment #2: *Search dates: I don't believe that the focus on only 2023 is currently well-justified. If you want the most recent work, would it make sense instead to start 1 year before the date of the first search going up to the date of the first search?*

Response: Thank you for your suggestion. We agree that 1 year before the start of the initial search would be a more reasonable period than the arbitrary choice of the year 2023. We have changed this in the 'Search Strategy'-section:

"To ensure the feasibility of the review, the search will be limited to works published at most one calendar year before the start of the search [Exact dates updated at Stage 2]."

R1 comment #3: *Is there a contingency plan to broaden the search if there are not enough articles in this time period?*

Response: Thank you for your critical consideration of our methods. As we expect well over 150 eligible records, we did not consider having a contingency plan in place in case our expectations are wrong. We have decided to expand our search period if one year yields less than 150 eligible records, and have added the following to the 'Screening'-section:

"In case the number of eligible records is unexpectedly lower than this limit, the search will be expanded with an additional six months – to a total of 1.5 years before the start of the initial search."

R1 comment #4: *Is there a minimum sample size of articles that would be deemed sufficient? Is there a maximum that can be analysed due to capacity or the authorship team?*

Response: We find it difficult to make any claims about a minimum sufficient sample size of articles, but identifying enough eligible papers is unlikely to be an issue, based on our pilot and a rough estimate of the number of relevant papers per year based on Wrzus and Neubauer (2023). Ideally, to get a full overview of the recent literature, all eligible articles would be included in the review. However, as you allude to, there are limits to what we can achieve within a reasonable timeframe due to the capacity of the authorship team. The first author of this manuscript will screen and extract data from *all* records in the sample, while several other members of the team will assist by screening and/or extracting data from a subsample to allow for the assessment of reliability of these decisions. As we don't know exactly how many eligible records our search will yield, it was necessary to set a limit on the sample size of articles to ensure that the review remains feasible. We think that a sample of 150 articles will

be sufficient to be able to provide a good overview of current methodological practices, while keeping the review feasible.

We have also added some additional information regarding the choice of our sample size to the manuscript:

"Random sampling of eligible records has been used in systematic reviews (e.g., Wrzus & Neubauer, 2023) to allow for sufficient coverage of the literature while safeguarding the review's feasibility (considering the small team). The sample size was decided based on a pilot (described in the Supplementary Materials), and a rough estimate of the number of relevant studies that are published in a single year based on previous systematic reviews (e.g., Wrzus & Neubauer, 2023). In case the number of eligible records is unexpectedly lower than this limit, the search will be expanded with an additional six months – to a total of 1.5 years before the start of the initial search."

R1 comment #5: *You say that citation tracking will not be carried out as there would be few additional records due to the time period – would this not just mean that it would be very quick to do, and therefore still worth doing?*

Response: 1. Thank you for your suggestion. However, the sample size for this review is fixed (to 150 articles). Most likely, our final sample of 150 articles will already be a random sample of all eligible articles. This means that citation tracking, even if it is relatively quick and easy, would only make the pool of all eligible articles larger, not the final sample included in the review. Additionally, we think there is no reason to believe that including studies identified through citation tracking would make the sample more representative of the research field. Therefore, we have chosen not to carry out citation tracking at all.

R1 comment #6: *Data synthesis: how many reviewers will synthesise the results? Describe how the reliability of the decisions will be assessed if only one reviewer will be involved. In order to avoid bias more than one reviewer should contribute to this process. (Topor et al., 2020)*

Response: Thank you for making us aware of this lack of transparency, and for providing a very informative reference. We added some additional information about our synthesis procedure:

"As the synthesis of numerical and categorical data is largely predetermined, this synthesis will be carried out by one researcher and the reliability of the decisions will not be assessed. An RMarkdown document containing the preliminary analysis plan for this data can be found on https://osf.io/abvxp/. Open-ended items will be coded by two researchers independently, and reliability will be assessed. Decisions regarding presentation of the numerical results and the narrative synthesis (Popay et al., 2006) of the open-ended items will be discussed with multiple members of the research team."

Comments from reviewer 2:

R2 comment #1: *It would be beneficial if the concept of adaptivity is better explained. Neither figure 1 nor the first time the concept is mentioned in the text establishes this concept thoroughly. However, adaptivity is one of the criteria to assess transparency and methodological variation. I could (at least in the SOM) envision a table, providing a non-*

exhaustive overview of adaptivity - and frankly I am not 100% I know what you mean by adaptivity.

Response: Thank you for bringing this lack of clarity about the conceptualization of adaptivity to our attention. To remedy this, we have added a broad definition of adaptivity:

"'Adaptivity' is quite a broad term, that as of yet is not used often in the ESM literature. Based on related literature in the field of educational sciences (Wauters et al., 2010), we formulated a definition of adaptivity in ESM designs. This definition is intentionally very broad, so that the current work can maximally inform a more precise conceptualization of adaptivity. We define adaptivity as the adjustment of one or more characteristics of the ESM study design to the individual participants' characteristics and preferences, preceding measurements, and/or the context."

We hope to create the overview table that you suggested, and a tentative framework that can inform future discussion of adaptivity, based on the data that we collect in this review.

R2 comment #2: *Given the amount of ESM studies, the restriction to one year and 150 randomly selected studies is sensible. I do though wonder whether one could apply either a stricter exclusion criterion or consider even two separate papers - published jointy. My reasoning is that ESM is used in various fields, including clinical psychology / psychiatry. I am curious whether the transparency differs (is lower/higher?) in ESM studies conducted on patients. Here I would expect very little data sharing, and even a too concise data analysis section. Some PSA members have written a book chapter on how to improve open science in clinical psychology (see https://link.springer.com/chapter/10.1007/978-3-031-04968-2_19) as this field of psychology might be behind cognitive psy. Note, I do not have data, only own collaboration with researchers from both fields. Transparency with resepct to describing the patient population is high, but method section are often less detailed in clinical psy papers, including ESM studies.*

Response: Thank you for your suggestion of this undoubtedly interesting research question. We agree that the large amount of data that will be collected for this work has the potential to answer many specific research questions, which cannot possibly be contained in one paper. We have made the choice to limit this initial paper to a general description of the research field, which is reflected in our broad, descriptive synthesis plan. In the future, we hope to use the extracted data to answer more specific research questions, such as whether transparency varies with research field, study population, or other study characteristics. Hopefully, some of the plans to answer these specific research questions will be preregistered before the start of the data extraction for this initial review.

R2 comment #3: *That brings me to another point worth to look at in your analysis: where is the article published. In the advent of OSF and alike, SOMs are a no brainer, and necessary, as some journals have word limitations! This should be taken into account, not least if the ESM is published as "short report" with a e.g., 4000 word limit. That does severly hamper full transparency. With 150 articles, it should be doable to check the journal requirements / restrictions if any.*

Response. It is indeed reasonable to assume that transparency will vary based on journal requirements, which we had not considered before. Thank you for the suggestion! To enable easy access to the journal information, we have added an extra item in the data extraction,

recording the journal the article was published in. In the data synthesis of this initial review, we will not statistically assess the potential association between journal requirements and transparency (for the same reason why we will not assess the association between other study characteristics and transparency, please see the response to your previous comment). However, I think these possible associations warrant at least an acknowledgement in the Discussion-section of our Stage 2 manuscript, as it would be unwise to imply that any differences in transparency are necessarily due to poor reporting standards.

R2 comment #4: *Regarding figure 2, software should be software and version.*

Response: We have followed your suggestion by adapting the data extraction item regarding software to include the software version:

"Q810. Which software and version was used to conduct statistical analyses?

"Software, version" or "Software, Not reported" (if the version is not reported) or "Not reported"

Please only report the software (e.g., R) and the software version, not the packages used.

____"

R2 comment #5: *study design may include hardware, some ESM software is only for android phones, other researchers hand out mobile devices to participants (not uncommon in clinical studies). This has IMO an effect on compliance. Please note that I have not checked whether anybody has investigated this.*

Response: We agree that our data extraction regarding ESM hardware (participants' smartphones, researcher-provides smartphones, etc.) did not capture the full picture of the manner of data collection. To remedy this, we have added an item to record which specific software (app/website) was used to collect the ESM data:

"Q25. Which software was used to collect the ESM data?

Name of app (e.g., m-Path) or website (e.g., Qualtrics) or "Not reported".

____"

Recording the name of the specific software allows us to determine other potentially relevant information (such as the operating systems that the software can be used on) ourselves.

R2 comment #6: line 391: "... through MS forms" - are you refering to having used a microsoft survey tool? Following FAIR principle, this acronym may not be understood in 10 years, who knows.

Response: Thank you for noticing this. We have followed your suggestion and changed "MS forms" to "Microsoft Forms" in the entire manuscript.

R2 comment #7: *I am not sure you can score a non-reporting as 0 (absent), as there might be absent due to N/A, and absent due to lack of transparency (i.e. the study design or analysis clearly implies that this and that has been done)*

Response: We completely agree that coding a non-reporting due to lack of transparency and an item that is simply not applicable to the study the same way would not be appropriate. In

our data extraction, this is not the case. Because of the branching structure of the items, items that are not relevant to a study (and are thus not presented to the coder) are coded as N/A, while items that are relevant but not reported are coded as “Not reported”.

Yours sincerely,

Lisa Peeters

and on behalf of co-authors Prof. Wim Van Den Noortgate, Dr. M. Annelise Blanchard, Dr. Gudrun Eisele, Prof. Olivia Kirtley, Dr. Richard Artner and Prof. Ginette Lafit.