This is my first registered report, please bear this in mind.

The proposed work seeks to understand measurement and structural characteristics of some of the most widely used loneliness measures. This is the sort of basic, descriptive research we need to be able to put so much prior research into context, to have solid ground to stand on in order to move forward in this area.

My review will be structured into 2 sections – my thoughts on PCI RR stage I criteria A-E, and line by line comments.

**1A. The scientific validity of the research questions.**

According to existing theory, or rather assumptions, there is one general factor of loneliness, except when maybe there are 2. This project cuts through the vast swath of work on loneliness to address a set of key related questions on the basic structure of loneliness measurements. These inquiries are valid and justifiable.

**1B. The logic, rationale, and plausibility of the proposed hypotheses, as applicable.**

The authors present several hypotheses, and acknowledge that they do not have specific hypotheses in certain scenarios. But, they leave the door open for the development of further hypotheses through their exploratory-confirmatory pipeline. This is an appropriate way to proceed with these sorts of questions in this sort of dataset. The hypotheses they do present are meaningful and testable; they follow from theory and research questions, though these links could be clarified more precisely in the text.

**1C. The soundness and feasibility of the methodology and analysis pipeline…**

I reviewed the methods description in the paper, and went through all analytic code linked to on OSF. The authors introduce reasonable sample size criteria in their opening table, and their data set generally meets these criteria. Where the criteria may not be met, the authors have noted this, and should continue to note it as a limitation. The analyses and inferential approaches are appropriate for these types of models (EFA’s and CFA’s). Inclusion and exclusion criteria are generally good, but there are some missing bits – e.g. the authors handling of missing values/NAs. I was able to understand what the authors intend to do through looking at the analytic code, but there is almost no mention of this in the text (this sort of thing should be improved). The authors distancing of themselves from the sample partitioning procedure is appropriate, and the code matches the functionality required, though I hope this might be described in a bit more detail in the actual text in a future version.

**1D. Whether the clarity and degree of methodological detail is sufficient…**

Though the authors are looking at quite a few different groups (e.g. countries) with their analyses, the bulk of their analyses are not overly complex. The inferential criteria for model fit are clear, and the highlighted text in the results sections layout clear pathways for how the results will proceed depending on the outcome of each analysis. The Bonferroni corrections are appropriate. The code is generally clear and well commented. Though I did not find simulated or pilot data to test the analytic scripts, I was able to follow them and I did not see a reason why they would not function. However, I would encourage the authors to put their code on github or some other open versioning system so if changes are made, they can be transparently followed. All in all, I believe this stage 1 submission presents a reproducible workflow and sufficiently protects from bias.

**1E. Whether the authors have considered…**

As this is an observational, not experimental, study, many common/suggested tactics are not applicable here. Again, I would like to see a bit more about how the authors plan to handle unusual data, like NA’s, potential outliers (or if we shouldn’t expect outliers…). However, their use of multiple loneliness inventories is a strong implantation of parallelism, as well as being highly relevant to their research questions in the field more generally.

---

On the opening table:

“given that no comprehensive data exist on its factor structure” – what exactly does “it’s” refer to hear?

“Following these analyses in the exploratory fold, we will decide – per country” – decide based on what? This isn’t particularly clear here; it is somewhat better in the main text but really this ought to be fleshed out more to reduce researcher degrees of freedom.

Further line by line comments follow…

LNs 64-73 are a good first paragraph, but the authors jump too quickly in the next paragraph. What \*is\* loneliness and why is it an issue now? The authors cover some of this in the 2nd section of the introduction, but I’d really like to see some broader scoping text that contextualizes this issue as more than just a “strategic priority”.

LN 89 – please clarify: “increased spending in mental health care” by whom?

LN 90 – not sure this is the best word choice – is there such a thing as “wanted” loneliness?

LN 113 – something to consider more here are these sampling differences. Or if not here, then in the discussion, since sampling differences come in many varieties, and can be extremely important, but this report only really covers national differences.

LNs 147-156 – this is all fair criticism, but in the interest of being balanced before the authors present the results, I think there needs to be a bit more on what single item measures can do, or have done, or what we have learned from them. More brief lit review of this along the lines of the following paragraph on what “insights” composite indexes have yielded (LNs 157-162).

LNs 195, 196 – here the authors write “alpha” and “tau”, but elsewhere they use the proper Greek symbols. I suggest the authors choose one or the other for consistency, and I also suggest they choose to use Greek characters.

LNs 250-251 – it’s given elsewhere but I think it should be here too – state that 0.6 is the cutoff value, or tell the reader where to find it.

LNs 256-257 – “to the level of shape and CI of the correlations” – I don’t understand this phrase. Could the authors please rewrite this and clarify?

LNs 260-261 – so how will the authors proceed if they have no predictions? I have an idea since I’ve read the whole paper over, but I think the authors need to give a little more information here about what they’re going to do.

LN 272 – what are “ex-post weights”?

LNs 314-317 – I’m not sure these Omega values should be included here, it goes against the flow. For instance, Omega\_u-cat: the u\_cat part hasn’t been described yet and the reader may not know what this is. It also isn’t clear just what segment(s) of the sample these Omega values are for. Please move them, or clarify in case I’m missing something.

Footnote 2 – yes, this is all true, but it’s also totally fine given the particular circumstances. I support this decision by the authors.

LNs 349-350 – this is actually not what the authors propose to do with the T-ILS in their analytic code, and I agree with the analytic code. So this should be rewritten – it is only for the DJGLS-6 that exploratory factor analysis will be used.

LNs 368-377 – I leave this decision up to the authors, but I don’t personally think this approach is necessary. The authors could just look at the EFA’s of 1, 2, and even 3 factor solutions and then evaluate model fit, rather than using 2 somewhat subjective criteria. But ultimately I leave this at the authors’ discretion.

LN 383 – this is somewhat correct and somewhat incorrect. Firstly, the authors leave space for RMSEA confidence intervals in the results section, which is a good thing to do, but those 90% CIs are akin to pCLOSE, so ultimately they are still using significance testing here. That’s fine, though the authors should definitely mention the fact that they’re going to be using RMSEA confidence intervals here, before the results.

LN 393… – Good use of WLSMV and ordered responses in SEM.

LN 454 – why does the unraveling happen only at the scalar level? Or am I misunderstanding this? It would be good to know more about what the implications of switching to the mixture multigroup analysis are.

A final note about the construct validity sections, drawing on the code. In the code, the authors appear to be fitting many CFA constructs for the wider nomological net, which is good. But I’m left wondering why the correlational analysis is so simple. Westfall and Yarkoni (2016) demonstrated that correlations between latent constructs are easily inflated, but when one accounts for measurement invariance as one does in SEM, we can get much closer to the true correlations. Westfall and Yarkoni also show how this sort of analysis can be done with single indicators as well, when a multi-indicator latent variable cannot be formed. I suggest that authors consider this approach, or at very least they refrain from using Pearson correlations for this set of analyses. Pearson correlations really only ought to be used on continuous variables, and adding up the items of these constructs doesn’t really make them continuous, there is still clear ordinality in these sorts of data. I strongly suggest the authors use Spearman or Kendall correlations instead.

I look forward to seeing the next stage of this report.