

Dear Drs. Renner and McIntosh,

We would like to thank you and the reviewers again for your quick turnaround, your constructive feedback on our manuscript, and for pointing out a number of typos and formatting mistakes.

Below, we address your remarks point-by-point. To facilitate the review process, we have included a version with track changes and a version with all changes accepted. Our most significant change to the manuscript was following your suggestion to remove results interpretation and loneliness measurement recommendations. We agree that both would be premature at this stage of the research process, given that half of the dataset is yet to be analyzed.

On top of the issues you pointed out, we conducted another review of our analysis scripts ourselves, and corrected several programming mistakes that had varying levels of impact on the results (e.g., mis-specified variables or incorrect parameter values in function calls). Specifically, one mistake involved using 'more than  $\frac{2}{3}$  of the correlations' instead of the preregistered 'at least  $\frac{2}{3}$  of the correlations', which affected 9% of the nomological nets proportions. We documented these changes on the GitHub repository of the project: <https://github.com/iropovik/lonelinessMeasurementEU/commit/cb68f5ee515c0539cd17af7a3e7c0bdd80d62eb1>

We hope that this revised version meets the high standards set forth by PCI-RR for In Principle Acceptance. Should you have any further questions or concerns, please do not hesitate to contact us.

Sincerely, and on behalf of all co-authors,  
Bastien Paris

### Clarifications

1. **Drew Altschul:** The authors offer to “explicate which of the reported indices/statistics are not used for inference”. I would indeed appreciate it if the authors did this, I think it would help many readers with less statistical expertise and technical acumen.

**Authors' Response:** We clarified our analytic reasoning in the following sentence, line 429:

Given the large size of the sample included in the study, we expected the  $\chi^2$  test of model fit to consistently return significant p-values. Consequently, we did not use  $p$ -values nor RMSEA confidence intervals to make inferences when evaluating the fit of the factor structures (but still reported them for the sake of transparency and completeness). Instead, we considered the model fit to be sufficient with CFI values  $\geq .90$  and RMSEA values  $\leq .08$  (see also De Roover et al., 2022; Hu & Bentler, 1999).

2. **Drew Altschul:** Under Results, factor analyses and internal consistency: “following exploratory and confirmatory factor analyses in our exploratory fold” – I know the authors did here and I know why this reads as it does, but I can’t help but fear that this phrasing will confuse people.

**Authors’ Response:** We enhanced clarity by changing the phrasing “Following exploratory and confirmatory factor analyses in our exploratory fold” to “Following factor analyses in our exploratory fold”.

3. **Drew Altschul:** Under Construct Validity: “(3) health.” Please be more specific here. From the new figure, I’m guessing this refers to the ubiquitous self-rated health question, but it would be great to have that cleared up right here.

**Authors’ Response:** We clarified as follows: “(3) health, which was reported through a one-item self-rated health question”.

4. **Mary Louise Pomeroy:** Line 297. I would like to see slightly more explanation on how the demographic quotas were determined.

**Authors’ Response:** We edited the paragraph, which now reads as follows:

“We used quotas based on the population of each Member State to reflect the target population in terms of age, gender, educational attainment, and Nomenclature of Territorial Units for Statistics (NUTS) region of residence. These simple, non-interlocking quotas were mapped to population shares calculated from Eurostat’s official population statistics by male/female gender, six age groups (16-25, 26-35, 36-45, 46-55, 56-65, and 65+), three education groups (International Standard Classification of Education ISCED level 0-2; levels 3 and 4 and levels 5-8); and 2-16 geographical regions depending on the country.”

5. **Mary Louise Pomeroy:** Pages 10 and 17. I do not see an initial definition for the acronyms NUTS, CINT, and JRT.

**Authors’ Response:** We added definitions for the NUTS (Nomenclature of Territorial Units for Statistics) and JRC (Joint Research Centre) acronyms at their first occurrence in the manuscript. We edited “CINT” to “Cint” and further specified that it is an online platform consisting of a network of panels.

6. **Mary Louise Pomeroy:** Line 680: I found that the way this data was presented was difficult to follow. Also, add a final percentage sign at end of the sentence.

**Authors’ Response:** We edited the paragraph to enhance clarity, and it now reads as follows:

“Using the same criteria as for the full nomological network, loneliness measures show predictive validity for the three domains in the following number of EU

countries: DJGLS-6, 24 countries (88.89%) for social activities and attitudes, 27 (100%) for emotional states, and 25 (92.59%) for health; T-ILS, 19 countries (70.37%) for social activities and attitudes, 27 (100%) for emotional states, and 27 (100%) for health; Single-item measure, 14 countries (51.85%) for social activities and attitudes, 27 (100%) for emotional states, and 26 (96.30%) for health.”

7. **Mary Louise Pomeroy:** Line 121. Please consider citing the following article which discusses research measurement issues that stem from conflating social isolation and loneliness. Pomeroy ML, Mehrabi F, Jenkins E, O'Sullivan R, Lubben J, Cudjoe TKM. Reflections on measures of social isolation among older adults. *Nat Aging*. 2023 Dec;3(12):1463-1464. doi: 10.1038/s43587-023-00472-4. PMID: 37640906.

**Authors' Response:** We cited this article in lines 121 and 134.

8. **Elizabeth Renner and Robert McIntosh:** Please ensure that the design table is fully updated to reflect the current status of the analyses and what is yet to be done. (In addition, in one place the manuscript makes reference to certain actions to take place 'after the analyses in the exploratory fold.')

**Authors' Response:** We updated the “hypothesis” column of the design table by adding the hypotheses for the confirmatory fold, as derived from the analyses on the exploratory fold. Given their important number, we provide country-by-country hypotheses in Appendix A. We also updated the “interpretation given different outcomes” column to describe how we will determine if the results on the confirmatory fold are consistent (inconsistent) with the results on the exploratory fold.

We have removed the sentence “Exact cross-validation predictions will be sharpened – to the level of direction, magnitude, and CI of the correlations – after the analyses in the exploratory fold.” on page 16.

### **Typos and Formatting**

9. **Elizabeth Renner and Robert McIntosh:** Please ensure that tables are called out in numerical order (currently Table 2 is called out before Table 1). Please note also that the data in Table 2 do not look quite right (are columns shifted?).

**Authors' Response:** We appropriately renamed the Tables and fixed the content of Table 2 (now renamed Table 1).

10. **Elizabeth Renner and Robert McIntosh:** Generally, it would be helpful to keep listed concepts (e.g., factor structure, internal consistency, measurement invariance, and construct validity) in the same order throughout the abstract, design table, and manuscript. For the nomological network analysis, the three main concepts (social

activities and attitudes, emotional state, and health) appear in various orders throughout the manuscript. Consistent ordering will assist readers' understanding of what was done. Further, in the construct validity section, a list of items begins with an "a)" under the social activities and attitudes constructs, but there is no "b)" and so on.

**Authors' Response:** We edited the manuscript (as well as Figure 1) to display a consistent ordering of the nomological network concepts as follows: a) social activities and attitudes, b) emotional states, and c) self-reported health. We further fixed the construct validity section by adding the missing b) list.

**11. Mary Louise Pomeroy:**

- a. Line 121 and elsewhere. Please correct author's last name "Probaska" to "Prohaska."
- b. Line 203. "Gap" should be plural.
- c. Line 309. I believe "panelist" should be plural.

**Authors' Response:** We corrected these different typos. Thank you for pointing them out.

## **Results interpretation and recommendations**

**12. Mary Louise Pomeroy:**

- a. Line 721. The authors go back-and-forth on whether the T-ILS is intended to measure general loneliness or social loneliness. I found it to be a bit contradictory and it confuses the authors' interpretation of findings discussed in the final "Summary of the Exploratory Fold" section.
- b. Line 735. Remove "the" from the sentence "...that the how..."
- c. Line 738. Very interesting discussion that I truly enjoyed reading. After digesting the authors' interpretation, there are two points I would like to push back on, asking the authors explore them further.
  - i. In hindsight, I think it makes sense that the emotional loneliness subscale correlates more closely with depression, whereas the social loneliness subscale correlates more closely with social support. To some extent, I'd expect most measures of loneliness to correlate closely with indicators of depression. Social support might better capture access to resources, or possibly levels of social contact that are indicative of social isolation (or a lack thereof). Loneliness and social isolation are weakly to moderately correlated at best, whereas research consistently shows higher correlations between loneliness with depressive symptoms. Loneliness might be thought of as an aspect of mental health, not unlike depression. Thus, while counterintuitive, I am not sure I agree that these correlational relationships point to weaknesses in three of the four loneliness measures. I could be swayed, but need the authors to provide further rationale to convince me.

- ii. I do not find myself fully satisfied with the authors' determination that the DJG social loneliness subscale is preferable to the TIL-S. I think this determination should be clarified by revisiting the overarching purpose of the study. Is it to advance a single measure that can reliably monitor loneliness across the entire E.U. population? If so, the TIL-S demonstrated scalar invariance across all countries, plus superior factor structures and reliability. Or, is the goal to explore differences in loneliness by culture and country, so that we may inform the most accurate measurement of loneliness, even if measurement tools may necessarily differ by region? If so, perhaps the social loneliness subscale is the best fit, as it may vary by clusters of EU countries. I think each goal is equally valid and, if I am not misinterpreting the authors' findings, each may lend itself to a different measurement recommendation. Further clarification and discussion may elucidate this.
- d. Line 749. "with the exception of, for instance, Bulgaria." What finding is this referring to?
- e. General comment: At some point it will be helpful to see a discussion of the limitations, which I imagine will be written in a subsequent stage. I would like the authors to address two issues. First, what are the limitations of using an online survey? What does this mean for how the sample might be skewed? For instance, are we measuring these constructs in a slightly younger sample, or a sample with higher socioeconomic status? Are we excluding the most lonely individuals or those with less social contact, who by way of their social exclusion may be less likely to participate in this study? A second minor point is that I am curious about country-specific findings. For example, what are the implications for Finland, France, and Romania, wherein the authors did not observe insufficient internal consistency of the DJGLS-6?

### 13. Drew Altschul:

- a. The paragraph beginning – "Somewhat concerning are the modest correlations between scales are supposed to test the same constructs...". Realistically, 0.6 - 0.7 is really not that low for situations like this. It is worth taking note of, and I think my primary issue here may lie with the authors language – what are you really telling the reader by saying this is "somewhat concerning"? "Modest" is not an appropriate descriptor of a correlation this large. But that said, I think the numbers here are very important to consider.
- b. The paragraph beginning – "What is clear is that the how loneliness should be measured...". First, please note the typo. Second, however, is that I disagree with much of the interpretation here, and I think it may be the most important aspect of this manuscript so far. The authors' support of the social loneliness subscale makes sense from some perspectives, but not all. Social loneliness seems to have better predictive validity, this is true, important and worthy of mention, and very much worth trying to understand. But it can't be understood without understanding what is going on with the other "side" of loneliness. We must consider *face* validity. The TLS items, to me, are more closely to the feeling of "loneliness". This is bolstered by the fact that the single-item measure, for all its faults, arguably has the best face validity: "have you been

feeling lonely?" And that item converges more with the TLS and emotional side of DJGLS. So emotions, negative emotionality, and mental health (depression & anxiety are particularly) are wrapped up in loneliness, we already know this, but it needs to be actively considered. The particular criteria variables chosen to relate the loneliness measures too matter as well. There is the potential for selection bias and imbalance there. There are a lot of questions about family, friends, and social contact. The authors ought to specifically consider what those questions say about the complex phenomenon of loneliness. I think a more judicious conclusion is that \*both\* aspects of the DJGLS are important, and the TLS and single-item measure are missing out on the social aspect. The emotional part of loneliness may overlap a great deal with typical mental health measures, but that emotional part is still very important! There is still unique variance there. Someone who wants to study loneliness is going to want to look at both sides of the loneliness the authors are finding. So these findings are very important but the framing and communication of findings is critical as well. Of course, this all depends on the confirmatory analyses as well. Some of what I just wrote may be premature; much of this is certainly discussion material, but I think more care is warranted here.

**Authors' response:** At this stage, and in line with the editors' suggestion, we have decided to remove the interim results interpretations and measurement recommendations from the manuscript. We believe it is more prudent to revisit these discussions after completing all data analyses, specifically after analyzing the confirmatory fold. This approach will allow us to provide a more comprehensive and robust interpretation of the findings and may lead to different discussions depending on the outcomes. Therefore, we will postpone the integration of this feedback into the manuscript. However, we are grateful for the reviewers' feedback and assure them that we will consider it carefully when we analyze the confirmatory fold.