

Dear Corina,

Thank you for the prompt response. We are very pleased that our revision has been well received, and we have addressed the remaining comments below.

Lindsay J. Alley  
Jordan Axt  
Jessica K. Flake

## **Response to recommender: Corina Logan**

You wrote: "On the advice of a colleague, we have switched from using Bartlett's factor scores for the sensitivity analysis to regression factor scores"

Please justify why this change was needed and why the new analysis is better suited to your study.

**Both of the factor score estimation methods implemented in lavaan, regression and Bartlett's, result in biased estimates of downstream effects (Devlieger, Mayer, and Rosseel 2016). In the previous version we had planned to use the default: Bartlett's. However, we have since learned that regression factors perform better in simulation studies compared to Bartlett's factor scores, which result in the largest amount of bias (Devlieger, Mayer, and Rosseel 2016). We have added the following to the paper to clarify this decision:**

**"Regression factor scores were used (Thurstone 1935) because they exhibit less bias in the estimation of downstream effects compared to Bartlett's factor scores (Devlieger, Mayer, and Rosseel 2016), the other factor score estimation method implemented in lavaan for continuous data." p. 21**

## **Response to reviewer: Shinichi Nakagawa**

The authors did a wonderful and diligent job of addressing my comments.

**We are glad to hear that the revisions were clear and satisfactory, thank you for the helpful review.**

There is one more thing to address. I wrote:

4: I think the authors are mainly talking about what is known as "selection bias" (in medicine) and "corridor bias" (in casual inference). Do the authors want to mention these words so this work will be more relevant to a broader audience?

They replied:

We would be happy to add these terms, as we would like our paper to be relevant to a broad audience and understand that differences in disciplinary language may impede this. Unfortunately, we are not

clear which specific part of the paper this is referring to or how to address it. We leave this to the recommender to provide guidance as to if and how we can address this comment.

I think they could add a sentence about 'selection bias' and 'corridor bias' at the end of the first section or just before the section "Measurement and Replication Research". These ideas are similar to a convenience sample not being equivalent to a true population.

**Thank you for the clarification. While selection bias is very relevant to research using convenience samples, it is a separate issue from measurement equivalence. In fact, when using multiple convenience samples in a single study, both selection bias and a lack of measurement equivalence could impact results simultaneously and interact to decrease the validity of conclusions. It is necessary to examine measurement equivalence in order to disentangle measurement differences from the impact of sampling on the estimated effect. If there are hidden measurement differences, and sum scores are used, it is not possible to know whether differences between convenience samples are due to this hidden non-equivalence, or due to selection bias.**

**Additionally, we added the following discussion of selection bias to the section on convenience samples (added text italicized):**

**"However, Strickland & Stoops (2019) point out that crowdsourced samples may differ from "the populations to which the results ideally would generalize" (p. 9), *a type of selection bias*. To deal with this limitation, they recommend that researchers collect samples through diverse methods and consider aggregate results. If this approach is to be effective, it's important that aggregated samples demonstrate ME, or that researchers employ a statistical model that accounts for non-equivalence across samples. If selection bias and hidden measurement differences are both impacting the results of a study, it is important to correct for ME in order to disentangle these two sources of bias." p. 9**

All the best for the research

Singed

Shinichi Nakagawa

Devlieger, Ines, Axel Mayer, and Yves Rosseel. 2016. "Hypothesis Testing Using Factor Score Regression: A Comparison of Four Methods." *Educational and Psychological Measurement* 76 (5): 741–70.

Thurstone, L. L. 1935. *The Vectors of Mind: Multiple-Factor Analysis for the Isolation of Primary Traits*. Vol. 274. Chicago: University of Chicago Press.