Reply to 3rd round PCIRR decision letter reviews #496:

Norton et al. (2007) replication and extensions

We would like to thank the editor and the reviewers for their useful suggestions and below we provide a detailed response as well as a tally of all the changes that were made in the manuscript. The editor's and reviewers' comments are in bold with our reply underneath in normal script.

A track-changes comparison of the previous submission and the revised submission can be found on: https://draftable.com/compare/OjRjMfeVffBJ

A track-changes manuscript is provided with the file: "PCIRR-S1-RNR3-Norton-etal-2007-rep-ext-main-manuscript-track-changes.docx" (https://osf.io/76h8r)

Reply to Editor: Dr./Prof. Yuki Yamada

> **We have asked the reviewer to check the manuscript again. As you can see, some minor issues have been raised. I also feel that all these should be resolved before granting an IPA. Please consider them and I would appreciate it if you could revise them again.**

Thank you for the reviews obtained, your feedback, and the invitation to revise and resubmit. We addressed all the final minor issues.

Reply to Reviewer #1: Dr./Prof. Zoltan Kekecs

> **I am grateful for the authors' detailed response to my suggestions. I have a few further observations and suggestions that may help the authors to improve the study and the manuscript:**
>
> **- Regarding the order effect: I appreciate the authors' concern that addressing the order effect in formal statistical inference would create unwanted complexity to the situation, which could threaten the confirmatory nature of this investigation. But instead of doing a confirmatory analysis on the order effect, I simply suggest to do an exploratory sensitivity analysis and/or some other investigations that could hint at the effect of study presentation order. I especially don't like the authors current proposal that the order effect analysis would only be done if the effect was not confirmed. This practically "stacks the deck" in favor of the authors. If they find the effect, no further investigation is done (which could question the authors' interpretation), but if the effect is not found, an analysis of the order effect could still salvage the situation and gives an extra chance for finding the effect. I suggest that the authors simply state that an exploratory analysis will be undertaken to investigate the possible influence of order of presentation. This exploratory analysis could include visual analysis of graphs plotted according to presentation order, and displaying descriptive statistics by order of presentation. These graphs and figures could be included in a supplement if these are too big to include in the main article. As the authors say, these analyses are straightforward and do not require too much effort, and this way they also don't threaten confirmatory power.**

We appreciate the feedback, and understand that you are asking us to conduct these analyses regardless, and we gladly will do so and revised accordingly.

It is important for us to note that we have no vested interest in whether the target's findings replicate or not, and there are no results that would be "in favor of the authors" or analyses that would "stack the deck". This is a Stage 1 Registered Report where there is an acceptance regardless of the outcomes, with full transparency and sharing of all procedures, materials, data, and code, and so each step can be evaluated and with a procedure that guards against outcome bias.

In our previous revision, we explained in detail why these additional analyses are an issue and asked for clear editorial guidelines as to how to proceed.

In case there was some ambiguity, we reiterate that our interpretation of the replication success is only based on the full higher-powered sample. We clearly labeled the order effect "exploratory", and we addressed multiple analyses by adjusting the alpha. Therefore, there was never any stacking of any decks, no increase in chances, or salvaging. The criteria was clear, and the additional order effects could not impact that or bias that in any way. These analyses were added to address in advance the possibility that things might not replicate, and offer possibilities as to why that might happen.

We appreciate the view that these analyses are potentially of interest regardless of the outcomes, even if there are issues of complexity and interpretability, and so we adjusted accordingly. We also added an explanation of what "moderator analyses" would look like.

We revised this paragraph:

> We therefore pre-register that if we fail to find support for our hypotheses that we rerun exploratory analyses for the failed study by focusing on the participants that completed that study first, and examine order as a moderator.

To this:

> We will run exploratory analyses focusing on the participants that completed that study first, and examine order as a moderator, meaning that we will run the analyses first with the study displayed first and then with the study not displayed first, and report the differences between the two, and examine whether the confidence intervals of the effect overlap.

> - **Sample size rationale: I am happy that the authors have revised their power analysis and now provide reproducible R code to support the sample size rationale of their proposal. I would like to point out that the sample size rationale in the current version of the manuscript is still inconsistent. The authors say that "multiplying the largest required sample size among all target studies (208) by 2.5 to 723". However, 208 x 2.5 = 520.**

Thank you for catching that. As indicated in the previous paragraph, the required sample is 289, and therefore this should have been 289 * 2.5 = 722.5 ~ 723.

Please see changes in the paragraphs:

> We conducted a series of a priori power analyses based on these effect sizes and we found that we require **289** participants to detect the effects reported in the target article with 95% statistical power at alpha = .05 (see supplementary materials Table S1 and analysis code for more details).

Given the likelihood that the original effects are overestimated, we used the suggested Simonsohn (2015) small telescopes approach with the generalized rule of thumb of multiplying the largest required sample size among all target studies (~~208~~**289**) by 2.5 to 723, rounding up to 800 participants.

- **It seems that the authors have re-classified H3 as an exploratory analysis. However, this is not properly reflected in the current version of the manuscript. Please, explicitly state in the main text that this research question is not a confirmatory hypothesis, rather, this will be an exploratory analysis. I would also not characterize this as a hypothesis ("H") anymore, since no inferential statistics should be run on exploratory analyses.**

We amended all references to H3 and the mediation model analyses as exploratory.