

RECOMMENDER

Comment 1: Thank you for your revisions. I am not sure if you fully updated the version of the manuscript I see in the osf link. You still need to address the point of specifying your IV more clearly: State explicitly in the results section and the design table that the IV will be absolute prediction error. State explicitly in both places that you will use model 2.

Authors' response: We apologize for the confusion, We have now stated explicitly in the Methods section (in subsections 2.5. Primary Measures, 2.7. Data analytical plan, and 2.8. Pilot data) and in the design table that the IV will be absolute prediction error. We have also stated in design table that we will use the Step 2 model from the pilot data. The rationale for using this model is described in the pilot data section and is also acknowledged in the data analytical plan section.

Comment 2: For power, you still need to justify a roughly smallest effect of interest. I am not familiar with the pre-existing power programs for lme, and it may be easiest to simulate yourselves: generate data from a H1 model with a fixed raw slope (e.g.) a 1000 times, run the lme, and determine the proportion of significant effects. This way you can estimate error variance from your pilot, fix the raw slope at a just interesting value, and vary N. Maybe hiring a graduate statistics student for a few hours would enable this to get done.

Authors' response: Thank you for your proposition and guidance. As recommended, we conducted power simulation using the R package labelled smir based on the data of the pilot study to determine the sample size required for the main study. Accordingly, we completely rewrote the 2.9. section (Sample size estimation) as follow:

"To estimate the sample size of the main study, we used the R package smir on the pilot data. In line with recent guidelines that suggest running power analysis based on the lowest meaningful estimate of the effect size (Dienes, 2021), we ran 1000 simulations with a raw slope of absolute_prediction_error fixed at .10. Results indicated that for an alpha of 0.05, the power was .83 (95% confidence interval [.80 .85]) with 27 participants across 336 observations. Accordingly, if α is chosen at .05, with a minimum effect size of .10, and a power of .80 is desired, then a sample of 27 participants along 12 measurement points (i.e., a running session) is required for testing the step 2 LMM presented in the previous section." (see also page 17, paragraph 3 of the revised manuscript).

As acknowledged in the Data, code, and materials availability subsection, the R codes and outputs of the simulation are available at <https://osf.io/2sb86/>.
