

Dear Editor,

Thank you and the reviewers for this second round of reviews. We adjusted the Stage 1 manuscript to reflect one of the reviewer's main suggestion, regarding the differential recruitment of our experimental and control groups. Following the reviewer's concerns, we have constrained this recruitment method only to the very first study in the series of five (study 1 on university students). For this first study, such method may still be an appropriate low-severity test, and we worry that changing the recruitment method now may raise a considerable challenge given our strict time constraint.

We hope that you find our reply satisfying, and thank you and the reviewers for your helpful feedback during this review process.

Anna, Cameron, Gianluca, and Andrea

Reviewer 1: Jana Kesenheimer (signed)

The authors have addressed my comments (and also those of the other reviewers) thoroughly and appropriately. Especially the idea of the coding list I find to be a valuable addition! I appreciate this work for the numerous additions and endorse the publication of the registered report. I am eager to see the results and wish the researchers success in recruiting participants and carrying out their project!

Thank you Jana for the nice wishes, and for your very constructive and helpful comments through the review process.

Reviewer 2: Helen Landmann (signed)

Thank you very much for this excellent revision that addressed all points that I mentioned. I have just one major concern left: In the revised version, it becomes clear that **the recruiting strategy is different for the experimental group and the control group**. "The experimental group will be recruited to participate in six hour meetings on climate education and fill online questionnaires, while the control group only to complete the questionnaires due to the difficulty of recruiting all 120 participants for the in-person meetings." This would make it very difficult to infer causal effects from the study. Every difference between the two conditions could be attributed to the different recruiting strategy and the **different motivation** that participants had from the start. The authors describe that they would use initial motivation as a control variable in their analyses. However, this only partly addresses the problem. The study requires a lot of time and effort. It would be a pity if causal inferences would be severely restricted by the different sampling strategies. I would like to motivate the authors to consider a **waitlist control group that receives the treatment later than the experimental group**. With this design all participants would be recruited with the same information. I'm aware that it is more difficult to recruit participants for the in-person meetings than to just fill in online questionnaires. However, you planned **five different studies with the above mentioned design**. I would suggest to reduce the number of studies but to choose the more robust design.

Thank you Helen for this follow-up suggestion regarding our recruitment strategy. We would like to clarify that the differential recruiting proposed for study 1 (where the experimental group is recruited for intervention + questionnaires, and the control group for questionnaires only), is *only valid for study 1*. In fact, with this first study we aim to perform a low-severity test of whether the experimental group's pro-climate behavior increases after exposure to our intervention to start with. The control group in this first study will be a passive one, to rule out external events that may cause the behavioral change other than our intervention (e.g. extreme weather events occurring during the period of the intervention). If we observed a behavioral change in the experimental group in study 1, for study 2 onwards we will switch to an active control group (i.e. undergoing a mock intervention); this will enable us to make more robust causal inference on the specific role of our intervention in triggering behavioral change. We elaborate this in the "Optimization procedure" section (under "Control group type"). In this section of our manuscript we previously said we would run two studies with a passive control group (study 1 for each life stage); however, following your concerns, we have now changed this to one single study (study 1 on university students), while the other 4 studies will all have an active control group.

For study 1 on university students, due to some strict time constraints related to the academic year schedule in Italy, we decided to proceed with recruitment following the differential recruiting method, which is now completed. This is because the intervention will take about 6 weeks, and the EMA will happen for 2 weeks before and 2 weeks after; we want to avoid the second round of EMA to overlap with the exam period at the end of June, leading to biased replies (i.e. during exams students carry out a different lifestyle than usual, with behaviors mostly centered around studying). However, as you mentioned, we plan on investigating potential differences in initial motivation due to this differential recruitment, by comparing the EMA scores of the baseline period across the two groups. While this is not a perfect solution, it will point at potential biases to take into account, if we observed a difference in behavior across the two groups in this very first study.

Lastly, that differences in initial motivation would be a larger problem if there was only one time point and the key test was between groups. We are most interested in the change in behavior between baseline and final EMA within groups. It is theoretically possible that experimental participants join the study with a higher motivation (because they are prepared to meet 6 times for 3 hours), but this alone should not make them necessarily more prone to increase their pro-climate behaviors from pre- to post-intervention, compared to the effect of the active treatment (the intervention). Nevertheless, as a robustness check, we will measure initial motivation and compare between conditions, and move to an active control group for the later studies. This is a compromise for feasibility and it doesn't seem to threaten the inferences of the whole project, and we hope you agree.

Thank you for your help in making our study design much stronger,

The Authors