

Peer review process documentation

of the article

The Impact of a Climate Crisis Class on Collective Action Participation

by

Raihan Alam, Maxwell Lyons, Tiffany Nguyen, Tal Waltzer, and Adam Aron

published in

Environmental Psychology Open

DOI: <https://doi.org/10.69805/epo.v29.a32>

Handling editor: Florian Lange

Review Round 1

Comments by the editor

1. Thank you for making such diligent use of our manuscript template. I attach a version of your manuscript with minor revisions of the German translations. Please consider them when preparing your revision, but there is no need to respond to those comments in the response letter.
2. Please insert your tables as tables (not as figures) into the manuscript file. You can find a formatting example in our manuscript template.
3. For figures, on the other hand, rather include them in a non-editable figure format. Otherwise, things like the decimal separators can switch when we process your files on our German computers (see Figure 4, for instance). Figure notes should appear below the figure (cf. Figure 4) and figures should include error bars (cf. Figures 2 & 3, but also see Reviewer C's suggestions for alternative figure formats).
4. Please add confidence intervals when reporting effect sizes.
5. I think your manuscript would benefit from including a table that displays zero-order correlations between all study variables, including not only the difference scores but also the original variables from which those differences have been calculated.

Comments by Reviewer B

The manuscript, "The Impact of a Climate Crisis Class on Collective Action Participation," explores whether taking a university course on climate change can shift students' knowledge, beliefs, and real-world action. It also examines how these different factors relate to one another. The topic is timely and fits well with *Environmental Psychology Open*. It's inspiring to see work that takes on the urgent question of what actually motivates real-world climate activism. This is a critical area of research, and the study makes an important contribution by putting it at the center. Still, the paper needs substantial revisions before it is ready for publication. The biggest issue is with the concepts. Terms like knowledge, beliefs, and awareness aren't clearly defined, and they don't always seem to be used in a consistent way. The theoretical side of the paper also is underdeveloped: it doesn't really match the complexity of what ends up being tested. Some concepts, like climate anxiety, show up in the results without being introduced or explained earlier. There are also a few things to sort out on the methods side. The way activism is measured is thoughtful and elaborated, but it still seems to rely fully on self-report. Statistical corrections are applied inconsistently across the analyses, and this should be addressed. The tone could be more consistent throughout, as some parts feel too informal while others are overly technical. The supplementary materials would also benefit from clearer titles, proper figure labels, and fully spelled-out abbreviations.

Strengths of the script

- Good fit for journal's scope.
- Pre-registered hypotheses
- Intervention study

- Studies high impact behavior

Suggestions for improvement

- Redraft all parts of the manuscript to improve clarity and focus
- Clearly state the paper's novelty and added value
- Stick to key concepts
- Use consistent terminology for each key concept throughout the paper.
- Avoid assertions that stray from the main focus of the manuscript.
- Address group size imbalance issue.
- Thoroughly discuss results and limitations.
- Follow APA 7th edition guidelines to structure the manuscript and all related information.

I have outlined below some examples of specific areas where these issues show up and where the paper could be strengthened.

Intro

The manuscript seems to include more psychological constructs than are meaningfully addressed in the literature review. At this stage, there is not enough conceptual grounding to fully support the choice of variables or to clearly explain how they fit into the study's broader rationale. There is also some methodological confusion, for example, it is not always clear how the constructs are defined or why specific measures were chosen. It looks like both qualitative and quantitative approaches were used to assess key constructs, but that mixed-methods aspect is not clearly flagged in the abstract or explained early on. Even after reading the literature review carefully several times, it is still not clear what is already known and what remains open when it comes to how knowledge relates to real-world collective action. There is a fair amount of relevant research out there, especially in fields like educology and environmental education, that could help strengthen this part of the paper. Journals such as *Environmental Education Research* or *The Journal of Environmental Education* might offer useful sources. The general direction makes sense, but the narrative is not quite consistent or fully aligned with the study's scope just yet.

- L90–96. It would make more sense here to focus on papers that directly examine the link between knowledge and activism, rather than broader conceptual discussions that do not clearly connect to the study. There is relevant research out there that could make the argument stronger. Since knowledge is a key part of the study, a clear definition would also be useful at this point.
- L97–99. Why specifically knowledge? I think it would be great to already hint here why knowledge is a target concept.
- Literature review contains bits that belong to the method section
- L124 (and also elsewhere in the manuscript). There seems to be some conceptual mixing between “knowledge,” “beliefs,” and “awareness.” At times they are used almost interchangeably, and in a few places, knowledge is described in ways that lean more toward beliefs. While these concepts are clearly related, beliefs can build on knowledge, and awareness can grow from either, they serve

different functions in how we understand behavior. Knowledge usually refers to verifiable information (<https://doi.org/10.1080/00461520903029006>), beliefs to personal convictions that may or may not align with facts (<https://doi.org/10.3102/00346543062003307>), and awareness to a more general, sometimes affective or attentional orientation to an issue (<https://doi.org/10.1037/1528-3542.6.3.392>). Part of this confusion likely comes from the fact that even widely used sources like the APA Dictionary do not always offer precise or consistent definitions. But especially in an empirical study, it is important to stick to definitions that help keep these constructs distinct and measurable. Personally, I find clearer distinctions more useful, but since the literature does allow some overlap, that just makes it even more important for the authors to explain what they mean by "knowledge" here, and what conceptual framework they are using.

- L157. For now, it is not clear to me that the current facet, confidence in future outcomes, really fits under the category of knowledge. It seems more like a belief or expectation than something based on verifiable information.
- In L176 and elsewhere there is a very commendable effort to be transparent about what the literature does and does not support. Still, the framing could be made more balanced by bringing in related findings, like correlational studies or intervention work that touches on knowledge-related constructs. That way, the lack of experimental research does not come across as a dead end, but more as a clear opening for future work.
- L198-202: this part is completely undeveloped and requires analysis.
- L204-208: belongs to the method section.
- L210-214. I would not group the psychological variables and knowledge together here, they are really quite distinct. It would be clearer and more consistent to keep the focus specifically on knowledge, especially if that is the main thread the paper is trying to follow.
- These sources seem relevant to develop the literature review further:

<https://doi.org/10.1016/j.jenvp.2019.101330>

<https://doi.org/10.1016/j.jenvp.2021.101563>

<https://doi.org/10.1016/j.jenvp.2021.101581>

<https://doi.org/10.3389/fcomm.2019.00004>

<https://doi.org/10.1088/2515-7620/ac9aeb>

<https://doi.org/10.1080/13504622.2025.2468513>

<https://doi.org/10.3389/fpos.2021.713340>

<https://doi.org/10.1177/2378023120925949>

https://doi.org/10.1007/978-3-030-04230-1_8

Method

The overall sample size looks fine, and the power analysis is helpful. But the strong imbalance between groups (103 vs. 29) could reduce the effective power for between-group comparisons and makes some ANOVA assumptions a bit shaky. It would be good to acknowledge this as a limitation. The within-group analyses seem solid, but any conclusions involving the control group should be taken with a bit more caution given the small N and higher attrition. It might also help to compare those in the control group who completed T2 with those who did not, at least key baseline variables. If there are no meaningful differences, that would help justify the final sample. And if there are, it is still useful information that can be addressed in the analysis or discussion.

- L252. perhaps better to say ‘did not report their gender’ instead of missing.
- L251. yet in abstract it is said that 133. And here 132.
- L260. please clarify, is this - #809500 - a case permission number?
- Figure 1. please elaborate what is meant by organisations.
- Figure 1. some words are uppercased and some not. Is there a rationale behind this?
- Figure 1. either mention all the knowledge facets, or do not mention them at all, because they seem to repeat the information in Table 1.
- L297-307 (while describing instrument). Some clarification would be helpful around how different constructs are positioned conceptually, especially in relation to beliefs. For example, self-efficacy and collective efficacy are widely understood as belief-based constructs, as they reflect perceived capabilities (Bandura, 1997; **Bandura, A.** (1997). *Self-efficacy: The exercise of control*. New York: W. H. Freeman. ISBN: 978-0-7167-2850-4.). However, biospheric values are more stable guiding principles and are generally not considered beliefs, though they can influence them. Similarly, climate anxiety is better described as an affective state, often rooted in belief structures, but not a belief in itself. These distinctions do not need to be deeply unpacked in the manuscript, but a clearer framing would support the overall conceptual clarity, especially if all four are being discussed in relation to knowledge or behavior. But this should already be addressed in literature review.
- Figure 2. Please indicate which line is control vs intervention. It is indicated in Figure 3 (next to the first interaction), but I do not think that this is the right place to indicate it.

Procedure: The procedure section includes a lot of valuable detail, but in its current form it is quite difficult to follow. It combines procedural steps, measurement information, detailed intervention content, and coding descriptions all in one long stretch of text. As a result, it is hard to track what was actually done at each time point, what was measured (and how), and what belongs to the course-based intervention versus to the data collection procedure. The section would really benefit from being reorganized or broken into clearer parts, for example:

- (1) What participants did at T1 and T2;
- (2) What was measured, distinguishing between qualitative and quantitative components;
- (3) What the intervention involved (course structure, activities, goals); and
- (4) How coding and scoring were conducted.

Results

Since none of the predictors in the regression turned out to be significant, it is not entirely clear what was gained by using uncorrected p-values in the earlier t-tests. It looks like the idea was to keep more potential predictors, but in the end, it did not lead to a working model. It also creates some inconsistency in how statistical standards are applied across the results. A tighter approach, either sticking to corrected values or selecting predictors based on stronger theory, might have made the test of Hypothesis 2 clearer and easier to interpret.

There is already a lot going on in the results section, especially with the mix of qualitative and quantitative findings. To help the main story land more clearly, it would make sense to focus on the key results that were part of the original plan. Some of the extra bits, like the Public Power group doing more actions, are interesting, but they might fit better in the supplementary materials. That way, the reader is not trying to hold too many threads at once, and the most important takeaways stay front and center.

Discussion

Given how much still needs to be worked out around the core concepts, theoretical framing, and methods, it might not make sense to revise the discussion in depth just yet. But, even as it stands, the discussion does not feel well anchored in the broader literature. The interpretation leans more toward describing the results or offering general ideas, but it does not clearly show how the findings connect to what is already known. Once the main issues are addressed, the discussion will need to be revisited so it better reflects those changes and speaks more directly to existing research.

- L573-574 very technical information.
- L571-572 it is not immediately clear what is ‘system 2’.
- L609. The paper states, this was not a power issue because the scatterplots did not show any relationship between knowledge and action. But that alone does not really rule out power problems. With a modest sample and quite a few predictors in the regression, it is still possible the study did not have enough sensitivity to pick up smaller or more subtle effects. Even if things look flat in the plots, that does not always mean there is nothing there, it might just mean the study was not equipped to detect it. Although the authors ran a power analysis and reported adequate power to detect medium effects ($f^2 = .18$), the results suggest that any existing effects are likely smaller than that threshold. Since none of the predictors came close to significance, it seems quite possible that the study was underpowered to detect smaller, but still meaningful, associations. So while the power analysis is a helpful addition, it does not fully support the claim that power was not an issue here.
- Although the authors refer to their outcome as “real-world collective action,” it is important to note that the data still comes from participant self-report. The weekly reflections were structured and scored, which adds depth, but they are still based on what participants said they did rather than on any direct behavioral trace or consistent external verification. If any form of verification was used (e.g., confirmation from group leaders or documentation of actions), including that detail would

really strengthen the claim. As it stands, it might be more accurate to frame this outcome as a detailed self-report of real-world engagement rather than a direct behavioral measure.

- The discussion does acknowledge the main null result, that none of the knowledge or belief variables predicted real-world action, which is an important finding. But most of the explanation leans on measurement issues or data limitations. It would be helpful to reflect on what this result means for the broader question the study set out to explore. If the aim was to see whether analytic knowledge drives action, and that did not hold up, that is something worth engaging with more directly, not just as a limitation, but as a possible insight.
- Some of the results are described as “striking” or “surprising,” but they come across as fairly modest, especially given the overall pattern of findings. For example, a small increase in collective efficacy in a class context is not exactly unexpected, and the lack of a relationship between knowledge and action, while disappointing, is not shocking either. Toning down the language a bit would help the discussion stay more in line with what the results actually show.
- The discussion refers to the weekly reporting approach as “pioneering,” - that might be a bit strong. Structured weekly self-report is not uncommon in this type of research. It is a useful method and applied well here, but might be better described as a helpful contribution rather than something entirely new.
- There is a lot of statistical detail carried into the discussion section, exact p-values, references to scatterplots, and repeated mention of significance levels. APA style encourages keeping the technical results in the Results section and focusing the Discussion on interpretation and meaning.
- The manuscript does acknowledge a few key limitations, such as sample size, generalizability, and some measurement concerns, but the reflection remains fairly thin. Since this is a single intervention study, conducted in one course and relying entirely on self-report, it would help to frame the findings more clearly as preliminary and context-specific. There is also no mention of the need for replication of results, which seems like an important gap, especially given the exploratory nature of some analyses and the null findings on the main hypothesis (not all results were null).
Regardless, right now, some of the conclusions feel a bit too confident given the limits of the design.

Other notes

The language is generally understandable, but there are many places where phrasing feels cumbersome or overly complex. In particular, there is a mix of formal (e.g., L9-12) and casual expressions that do not sit well together. Phrases like “research study” (L100) appear where simpler wording would be clearer, and some sentences are weighed down by redundancy or unclear structure. A focused language edit would help improve clarity and flow, especially for readers outside of the immediate research context.

- L13, perhaps “knowledge aspects/facets” would be a better fit than “items”?
- L16, abstract. Is it considered the same construct? knowledge / belief elements
- L18-19, abstract. Too general for the abstract, vague formulation.
- L25, abstract. Might be better for the sake of consistency to indicate “climate crisis” instead of climate change.

To summarize, the topic is an important one, and there is clear potential of the script, but it still needs quite a bit of shaping before it can really deliver on that promise. The conceptual foundations need tightening, the methods need a clearer throughline, and the results need to be distilled so the reader is not left navigating through too many pieces. Once those core issues are addressed, the contribution will be easier to see and easier to stand behind.

I really appreciate the opportunity to review this paper. I know it is a lot to take in, but I truly hope the suggestions are helpful for the next round, that is the spirit they were offered in.

Comments by Reviewer C

This work tests a real-world educational intervention: a university class on the climate crisis. Before and after the class, students were measured on various knowledges and attitudes, and hypotheses were preregistered about an expected increase in knowledge, sense of collective efficacy, and real-world action. The results are important and caution the field to take self-report as the gold standard for behavior change: as others have pointed out, it is poorly correlated with real behavior. This is an important piece of work that contributes to recent 'barriers first' calls for behavior change in environmental psychology. I applaud the authors for doing this complex real-world study, which of course also comes with complexities that are hard to mitigate (such as unequal group sizes).

Major suggestions

- Statistics: my main concern is the lack of an active control, and unequal sample sizes between conditions. I understand it's too late to change this now, but for the future I would strongly suggest an active control group (e.g. students taking a course around the same time that also includes some active group work, but on a different topic). This may allow stronger conclusions about the intervention, rather than with the passive 'no class' baseline now. Do we know anything about the demographic match between the control and intervention students (were the control participants also enrolled in the same program and taking similar classes, but just not enrolled in this particular class)?
 - For the future, it could be interesting to consider a variant of people only taking the lectures, or only engaging in group work. Now, the passive lectures and active social-change participation cannot be readily distinguished.
 - While I do not expect the authors to redo the data collection with an active control and/or more control participants (although I hope they do so for the next course iteration), there should be a broader discussion about statistical inference and potential statistical problems (e.g. power).
- data visualization: the figures could be much improved by adding individual data points (with connecting slopes), group error bars, and/or full distributions (for instance like a raincloud plot)? Perhaps using JASP as a free alternative to SPSS helps here, as they by default output more comprehensive graphs that show the individual variability of the data (see for example <https://jasp->

stats.org/2021/10/05/raincloud-plots-innovative-data-visualizations-in-jasp; <https://wellcomeopenresearch.org/articles/4-63> Figure R11).

- replicability: I was able to download the data in SPSS format, but there is no codebook or publicly available instructions in a format that I could open (I do not have an SPSS license). I can open the .sav file in JASP, but without instructions I could not replicate the basic plots: there are many variables, not all of which are relevant for the figures (I think), and I did not see which variable corresponds to the T1/T2 measurement (or how to otherwise pair the T1-T2 datapoints within the same person). I strongly suggest adding a codebook for step-by-step replication of the main results, which could be aided if the raw data was complemented with a spreadsheet containing processed data (and ideally, steps that describe how the raw data is converted into the processed data).

Minor suggestions

- in the introduction, 'before the class begun' is a bit abrupt. Before going to the measured target variables, perhaps explain briefly what class students are taking (elective/mandatory, BSc/MSc, class size, duration of the class), and why specifically an educational intervention is proposed (similar to the paragraph that is now at line 318; perhaps cite the Otto 2020 tipping points paper which lists education as one key leverage point)? It would help to briefly preview the type of intervention at the same level as all the outcome measures are justified in the introduction. For instance, it would help to briefly mention things like: were there also lectures next to the focus groups? What kinds of organizations were people assigned to? Now, some of this information is missing, or it comes very late in the paper.
- share the Qualtrics template, so others can reuse these more easily?
- research assistants who ranked knowledge and action scores: did they rate the responses blinded to the timepoint at which these were collected, and blind to the condition (class or control)? if not, may these rankers have been biased by knowing which responses were given at T1/T2?
- <https://www.nature.com/articles/s44159-024-00305-0> and perhaps <https://gep.psychopen.eu/index.php/gep/article/view/13089> would be helpful to cite (especially the Albarracin paper which to me is the new gold standard reference for behavioral interventions).
- the preregistration mentions a linear model, but the main text uses an ANOVA. could you explain why? the linear model may be more powerful in grouping participants within action cohorts (to account for e.g. different group dynamics as a function of each group's coordinator etc).

Author response to the reviewers Round 1

Dear Editor,

I am very grateful for the great of level of detail in both reviews and I think that the comments nicely complement each other. In line with the reviewers, I appreciate the impact-oriented nature of your study, the

focus on a real-world collective action intervention, and the compelling use of open-science practices, and I think your work is a great fit for Environmental Psychology Open. The reviewers also noted a number of aspects that need to be improved for clarity and consistency, and I think that these issues can be addressed in a thorough revision of your manuscript.

Thanks for your enthusiastic reception and kind remarks. We are also grateful to the two reviewers for their thorough attention and great suggestions. Below we respond to each and every comment and indicate how we have changed the manuscript. In the tracked document, we highlight changes in yellow.

**Sincerely,
Your Authors**

Editor

1. *Please insert your tables as tables (not as figures) into the manuscript file. You can find a formatting example in our manuscript template.*

Thank you for making this clear. Each table in the manuscript is now in table format instead of figure format.

2. *For figures, on the other hand, rather include them in a non-editable figure format. Otherwise, things like the decimal separators can switch when we process your files on our German computers (see Figure 4, for instance). Figure notes should appear below the figure (cf. Figure 4) and figures should include error bars (cf. Figures 2&3, but also see Reviewer C's suggestions for alternative figure formats).*

Thank you again for your clarity on this point. We have replaced the previous figures with non-editable images, moved figure notes to appear below the figures, and have included standard error bars in all of our graphs.

3. *Please add confidence intervals when reporting effect sizes.*

In Tables 2 and 3 we include the 90% confidence intervals for our effect sizes.

4. *I think your manuscript would benefit from including a table that displays zero-order correlations between all study variables, including not only the difference scores but also the original variables from which those differences have been calculated.*

Thank you for this suggestion. We have now created three zero-order correlation tables: one for Time 1 variables, one for Time 2 variables, and one for the difference scores. Each table also includes the summed collective action score, allowing readers to see how each variable at each time point, and their difference scores, relate to real-world action. These tables are in the Supplemental Materials.

Reviewer B

RB.1. The manuscript, "The Impact of a Climate Crisis Class on Collective Action Participation," explores whether taking a university course on climate change can shift students' knowledge, beliefs, and real-world action. It also examines how these different factors relate to one another. The topic is timely and fits well with Environmental Psychology Open. It's inspiring to see work that takes on the urgent question of what actually motivates real-world climate activism. This is a critical area of research, and the study makes an important contribution by putting it at the center. Still, the paper needs substantial revisions before it is ready for publication. The biggest issue is with the concepts. Terms like knowledge, beliefs, and awareness aren't clearly defined, and they don't always seem to be used in a consistent way. The theoretical side of the paper also is underdeveloped: it doesn't really match the complexity of what ends up being tested. Some concepts, like climate anxiety, show up in the results without being introduced or explained earlier. There are also a few things to sort out on the methods side. The way activism is measured is thoughtful and elaborated, but it still seems to rely fully on self-report. Statistical corrections are applied inconsistently across the analyses, and this should be addressed. The tone could be more consistent throughout, as some parts feel too informal while others are overly technical. The supplementary materials would also benefit from clearer titles, proper figure labels, and fully spelled-out abbreviations.

We thank the reviewer for the positive comments and concrete suggestions. We take these main points in turn below.

RB.2. Terms like knowledge, beliefs, and awareness aren't clearly defined, and they don't always seem to be used in a consistent way.

We have taken the reviewer's advice and have tried to deal with the complicated differences between knowledge, beliefs and awareness by adopting the formulation of Hornsey et al. 2022 of "analytical reasoning" through the manuscript, and in particular in the introduction on pages 4-8. That rubric appears to cover the kind of concept we were testing and teaching, be it knowledge, belief or awareness, without having to be specific. It is to be contrasted with non-analytical reasoning such as social norms and group identity.

RB.3. The theoretical side of the paper also is underdeveloped: it doesn't really match the complexity of what ends up being tested.

We have now reframed the introduction substantially. From the second paragraph onwards, we get straight to the topic of why analytical reasoning is an important topic to study, what has been done so far, and why our approach is different. For example, after reviewing the various studies on the relationship between analytical reasoning and climate action we write on pages 4-5:

“Notwithstanding this discouraging picture of the relationship between analytical reasoning/knowledge and climate and behavioral action we nevertheless set out in the current study to explore this issue further. We reasoned that many of these prior studies have not looked specifically at the relationship between analytical reasoning and collective climate action, have not provided adequate opportunities for real-world collective action (we used a class format where people were assigned to groups), and have not perhaps focused on the types of analytical reasoning that might be important. The prospect of identifying key types of analytics information that does drive people to collective action seems like a critical endeavor given that other ways of driving social mobilization are clearly very insufficient on climate, ecological and wider social crises.”

In that sense we have a theory here that our provision of information to the students about different kinds of analytical reasoning, informed by grassroots experience and in conjunction with the provision of real-world action, might reveal relationships that hitherto have been elusive.

RB.4. Some concepts, like climate anxiety, show up in the results without being introduced or explained earlier. There are also a few things to sort out on the methods side.

Climate anxiety was not one of our primary theoretical variables of interest; rather, we included it as a covariate in our multiple regression model to control for baseline beliefs. This is stated in the Methods section, Given that our central research question focused on the relationship between *analytical reasoning* and *collective action*, we didn’t highlight climate anxiety in the Abstract or Introduction. We hope this is acceptable.

RB.5. The way activism is measured is thoughtful and elaborated, but it still seems to rely fully on self-report.

We thank the reviewer for pushing us to clarify our measure. The students filled out self-reported surveys each week but we corroborated those against their group members, so this is objectively ratified action, which is quite different from 99% of studies in the field. We omitted to mention details in the original which we now do as follows on page 11:

“To provide a concrete estimate of alignment between individual self-reports and group-level descriptions, we conducted a structured consistency check on a random subsample of participants’ weekly reports. Specifically, one coder reviewed entries from four different subgroups (N = 20 participants) drawn from two time points in the course: Week 5 and Week 7. The sampled subgroups included Fossil Free Degree Subgroup 2 (6 participants) and Youth Will Subgroup 4 (4 participants) from Week 5, and Bike SD Subgroup 1 (5 participants) and Public Power Subgroup 2 (5 participants) from Week 7. For each participant in the subgroup, the coder compared their report of individual and group activity to the reports provided by their fellow group members for the same week. We found full consistency across all cases, such that every participant’s reported actions were

corroborated by at least one other group member's description, and no substantial mismatches were identified."

RB.6. Statistical corrections are applied inconsistently across the analyses, and this should be addressed.

We thank the reviewer for their attention to detail. We did in fact apply Holm–Bonferroni corrections to the paired-sample t-tests conducted within the intervention condition, and all results reported as significant survived correction. We have clarified this in the manuscript by replacing “(we used $p < .05$ uncorrected)” with “(using Holm–Bonferroni p-values corrected for 8 comparisons)” on page 16.

RB.7. The tone could be more consistent throughout, as some parts feel too informal while others are overly technical.

We thank the reviewer for pointing that out, we hope we have fixed this now with changes to the writing throughout the manuscript.

RB.8. The supplementary materials would also benefit from clearer titles, proper figure labels, and fully spelled-out abbreviations.

We thank the reviewer for this helpful recommendation. In response, we have made several improvements to the supplementary materials. Specifically, we removed the exploratory scatterplots and now report only clearly labeled correlation tables and coded thematic tables, each with consistent and descriptive titles. We have also ensured that all abbreviations in the tables are fully spelled out.

RB.9. Suggestions for improvement

- *Redraft all parts of the manuscript to improve clarity and focus*
- *Clearly state the paper's novelty and added value*
- *Stick to key concepts*
- *Use consistent terminology for each key concept throughout the paper.*
- *Avoid assertions that stray from the main focus of the manuscript.*
- *Address group size imbalance issue.*
- *Thoroughly discuss results and limitations.*
- *Follow APA 7th edition guidelines to structure the manuscript and all related information.*

We thank the reviewer for these overarching suggestions. Below, we address each of these points in detail where they arise.

RB.10. Intro

The manuscript seems to include more psychological constructs than are meaningfully addressed in the literature review. At this stage, there is not enough conceptual grounding to fully support the choice of variables

or to clearly explain how they fit into the study's broader rationale. There is also some methodological confusion, for example, it is not always clear how the constructs are defined or why specific measures were chosen. It looks like both qualitative and quantitative approaches were used to assess key constructs, but that mixed-methods aspect is not clearly flagged in the abstract or explained early on. Even after reading the literature review carefully several times, it is still not clear what is already known and what remains open when it comes to how knowledge relates to real-world collective action. There is a fair amount of relevant research out there, especially in fields like educology and environmental education, that could help strengthen this part of the paper. Journals such as Environmental Education Research or The Journal of Environmental Education might offer useful sources. The general direction makes sense, but the narrative is not quite consistent or fully aligned with the study's scope just yet.

- *L90–96. It would make more sense here to focus on papers that directly examine the link between knowledge and activism, rather than broader conceptual discussions that do not clearly connect to the study. There is relevant research out there that could make the argument stronger. Since knowledge is a key part of the study, a clear definition would also be useful at this point.*

We thank the reviewer for this helpful suggestion. In response, we revised the Introduction to focus more directly on studies that examine the relationship between knowledge and climate-related action, while also acknowledging the limitations of that literature. On page 4, we now write:

“Although it might seem *prima facie* obvious that providing analytical reasoning information to people is necessary for driving them to act collectively on climate and ecological problems, there are few demonstrations in the literature of this relationship (here we use the phrase “analytical reasoning” to encompass varieties of knowledge, awareness and belief). For example, an over-arching review of the relationship between environmental behaviors in general and knowledge suggested a negligible relationship (Albarracín, Fayaz-Farkhad et al., 2024), albeit this was mostly individual rather than collective behavior. One meta-analysis in that review covered dozens of studies on the relationship between climate change knowledge and climate change adaptation behaviors such as supporting environmentally friendly policies and relocating, finding an overall r value of only 0.14 (Van Valkengoed & Steg, 2019). Further, an article based on the identification of weak relationships between analytical information and personal efficacy beliefs on climate concluded that non-analytical factors such as social norms and group identity are probably much more important (Hornsey, Chapman et al., 2022), fitting earlier critiques that the problem with climate action is not to be fixed by filling information deficits but rather through cultural cognition, that is, forms of information that connect with people's values and ideologies (Kahan & Braman, 2006; Moser & Dilling, 2007). In the more specific domain of the university campus, our earlier study found that while a set of video interventions about climate change issues and social movements was related to an increase in collective efficacy beliefs, it scarcely triggered real-world participation (Castiglione et al., 2022). Another class-based study made presentations on topics such as the relationship between people and nature, waste management, and energy production, and found a subsequent tiny increase in adolescent's subsequent self-reported willingness to engage in climate-related activities and protest (Balundé & Poškus, 2025).”

As mentioned in our responses to RB.2 and RB.3, we addressed the definitional concerns raised, we reframed the construct throughout the paper as *analytical reasoning on pages 4-8*, drawing on the conceptual approach articulated by Hornsey et al. (2022). This framing allows us to unify previously inconsistent terminology (e.g., knowledge, beliefs, awareness) under a more cohesive conceptual umbrella. We hope this revision more clearly articulates the theoretical foundation of the study and better aligns with prior empirical work directly examining the knowledge–action relationship.

RB.11. L97-99. Why specifically knowledge? I think it would be great to already hint here why knowledge is a target concept.

We thank the reviewer for this question. We focused on analytical reasoning because it is often assumed to drive climate action, yet evidence for this link is limited — both in terms of the types of analytical reasoning studied and the extent to which real-world action is measured. This gap motivated our test of whether gains in analytical reasoning from a climate class would relate to real-world activism, a question with clear implications for climate education and movement strategies. We now make this rationale explicit in the introduction on pages 4-8, as described in RB.2, RB.3, and RB.10.

RB.12. Literature review contains bits that belong to the method section

We thank the reviewer for this comment and have moved the references to our study design from the Introduction into the Methods section.

RB.13. L124 (and also elsewhere in the manuscript). There seems to be some conceptual mixing between “knowledge,” “beliefs,” and “awareness.” At times they are used almost interchangeably, and in a few places, knowledge is described in ways that lean more toward beliefs. While these concepts are clearly related, beliefs can build on knowledge, and awareness can grow from either, they serve different functions in how we understand behavior. Knowledge usually refers to verifiable information (<https://doi.org/10.1080/00461520903029006>), beliefs to personal convictions that may or may not align with facts (<https://doi.org/10.3102/00346543062003307>), and awareness to a more general, sometimes affective or attentional orientation to an issue (<https://doi.org/10.1037/1528-3542.6.3.392>). Part of this confusion likely comes from the fact that even widely used sources like the APA Dictionary do not always offer precise or consistent definitions. But especially in an empirical study, it is important to stick to definitions that help keep these constructs distinct and measurable. Personally, I find clearer distinctions more useful, but since the literature does allow some overlap, that just makes it even more important for the authors to explain what they mean by “knowledge” here, and what conceptual framework they are using.

See our responses to RB.2 and RB.3.

RB.14. For now, it is not clear to me that the current facet, confidence in future outcomes, really fits under the category of knowledge. It seems more like a belief or expectation than something based on verifiable information.

We thank the reviewer for this helpful observation. We agree that “confidence in future outcomes” is better conceptualized as a belief or expectation, not as a form of analytical reasoning. In revising the manuscript, we clarified this distinction and removed “confidence in future outcomes” from our core list of analytical reasoning constructs. As part of this revision, we also realigned Table 1 to include only the eight constructs that we now consistently frame as types of analytical reasoning, and we labeled them accordingly. Each construct is now clearly matched to its corresponding survey question. We feel that removing “confidence in future outcomes” makes our paper much better, and doesn’t affect the OSF pre-registration (since it wasn’t one of the core questions asked). However, this minor change has also resulted in a small change in the results of the multiple regression, which we discuss below.

RB.15. In L176 and elsewhere there is a very commendable effort to be transparent about what the literature does and does not support. Still, the framing could be made more balanced by bringing in related findings, like correlational studies or intervention work that touches on knowledge-related constructs. That way, the lack of experimental research does not come across as a dead end, but more as a clear opening for future work.

We appreciate the reviewer’s thoughtful suggestion to provide a more balanced framing around the literature on urgency. In response, we have removed the earlier sentence stating “For this element, we didn’t identify any specific literature related to action,” and have expanded our discussion to incorporate related correlational and intervention studies. We spent a lot more time in this revision period looking over the large literature. Specifically, on page 6, we now write:

“There is some evidence that climate-science based urgency - the idea that we face a fairly imminent crisis - can increase people’s stated willingness to take action, although this appears to depend on how urgency is framed. For example, prior research found that perceived urgency related to intentions to support policy, intentions to vote and intentions to engage in personal actions (Leiserowitz *et al.*, 2025; van der Linden, 2015). Other papers suggest though that expressing the urgency as carbon budget timelines and tipping points rarely leads to action unless it is framed in actionable terms (Moser, 2016) or ways that connect with people’s values, identities and sense of behavioral control (Nisbet & Scheufele, 2009). Other research suggests that urgency might backfire without combining it with messages of efficacy and hope (O’neill & Nicholson-Cole, 2009). The topic of urgency also relates to that of psychological distance (i.e., how close or far away one feels climate change is to one’s prospects). One meta-analysis suggests that individuals tend to engage in more pro-environmental behaviors when they perceive the climate crisis as a proximal and concrete issue (Maiella *et al.*, 2020). However, this relationship has been contested by Van Valkengoed and colleagues’ (2023) meta-analysis, which found that only 9 out of 26 reviewed studies showed a positive association between psychological distance and climate action. Overall, the literature does support the importance of communicating urgency, and this, combined with the instructor’s ability over many weeks to lay out the physical-science-basis and political-economic basis of urgency suggested to us that analytical reasoning about urgency would relate to action.”

RB.16. L198-202: this part is completely undeveloped and requires analysis.

We appreciate the reviewer's criticism of this section. We have fleshed out our reasoning better and on page 7 under the *Collective Climate Action and Individual Climate Action* section, we now write the following:

"Some research has shown that people's perception of policy effectiveness or campaign effectiveness is a function of their collective action beliefs (Lubell, 2003; Seiferth, Tengö et al., 2024). Other research, at the individual level, has shown that people sometimes prefer simple curtailment behavior such as turning off the lights over more impactful individual actions such as shifting banking and driving an electric vehicle (Lundberg, Tang et al., 2019). We thus planned to measure participants' ability to reason about the types of actions they can take. We predicted that their ability to name concrete examples of collective actions (e.g., joining campaigns, going to protests) and individual actions (e.g., carpooling) would be related to their propensity to take action."

RB.17. L204-208: belongs to the method section.

We appreciate the reviewer's recommendation. In response, we have deleted the sentence "The participants were assigned to collective action groups and needed to act as a small group to earn class credit – although, to be clear, the amount of action required to get that credit was low and we anticipated many would do much more" from the literature review section. This information is already covered in the Study Overview and Methods section.

RB.18. L210-214. I would not group the psychological variables and knowledge together here, they are really quite distinct. It would be clearer and more consistent to keep the focus specifically on knowledge, especially if that is the main thread the paper is trying to follow.

Thank you. As explained above, we removed the "confidence in future outcomes" item from our analytic reasoning measures and maintained a consistent focus on analytical reasoning throughout the manuscript. We believe this addresses the reviewer's point and helps maintain conceptual clarity around the constructs under investigation.

RB.19. These sources seem relevant to develop the literature review further:

<https://doi.org/10.1016/j.jenvp.2019.101330>

<https://doi.org/10.1016/j.jenvp.2021.101563>

<https://doi.org/10.1016/j.jenvp.2021.101581>

<https://doi.org/10.3389/fcomm.2019.00004>

<https://doi.org/10.1088/2515-7620/ac9aeb>

<https://doi.org/10.1080/13504622.2025.2468513>

<https://doi.org/10.3389/fpos.2021.713340>

<https://doi.org/10.1177/2378023120925949>

https://doi.org/10.1007/978-3-030-04230-1_8

We thank the reviewer for the above references. We looked at all of these and now refer to some of them throughout the Introduction section of the manuscript, as well as beefing up the background literature to provide a better theoretical grounding.

RB.19. Method

The overall sample size looks fine, and the power analysis is helpful. But the strong imbalance between groups (103 vs. 29) could reduce the effective power for between-group comparisons and makes some ANOVA assumptions a bit shaky. It would be good to acknowledge this as a limitation. The within-group analyses seem solid, but any conclusions involving the control group should be taken with a bit more caution given the small N and higher attrition. It might also help to compare those in the control group who completed T2 with those who did not, at least key baseline variables. If there are no meaningful differences, that would help justify the final sample. And if there are, it is still useful information that can be addressed in the analysis or discussion.

We thank the reviewer for this important point. As suggested, we now explicitly acknowledge the group size imbalance as a limitation in the Results section (page 12) and write:

“Although the control group was substantially smaller than the intervention group ($N = 29$ vs. 103), post hoc power analyses confirmed that the significant interaction effects were adequately powered ($> 80\%$), given their medium effect sizes ($\eta^2 = .05-.08$). Nevertheless, we note that conclusions involving the control group should be interpreted with some caution due to the group size imbalance.”

RB.20. perhaps better to say ‘did not report their gender’ instead of missing.

We thank the reviewer for this suggestion and have replaced “missing” with “did not report their gender” on page 8.

RB.21. yet in abstract it is said that 133. And here 132.

We thank the reviewer for catching this inconsistency. We have corrected the abstract to reflect the correct sample size of 132 participants.

RB.22. please clarify, is this - #809500 - a case permission number?

#809500 refers to the IRB protocol number associated with the ethical approval for this study. We have not added further clarification in the manuscript itself, as we believe this will be understood by readers familiar with standard research protocols, but we are happy to clarify further if needed.

RB.23. Figure 1. please elaborate what is meant by organisations.

In this context, “organizations” refers to the five climate-focused social-change groups that intervention participants were randomly assigned to. We have not modified the manuscript text, as we believe this becomes clear in the Procedure section, but are happy to add clarification if it is needed.

RB.24. Figure 1. some words are uppercased and some not. Is there a rationale behind this?

We have fixed this to consistently uppercase words in the Figure.

RB.25. Figure 1. either mention all the knowledge facets, or do not mention them at all, because they seem to repeat the information in Table 1.

We decided to list only a few of the facets in Figure 1 to give readers a high-level sense of the intervention content without overloading the visual. Including all eight facets would make the figure overly busy and detract from its primary purpose as a simplified overview of the study design. Since Table 1 already provides full detail, we feel this balance allows readers to access the complete information without crowding the visual summary. We hope this approach strikes an effective balance between clarity and brevity, but we are happy to revise this if need be.

*RB.26. L297-307 (while describing instrument). Some clarification would be helpful around how different constructs are positioned conceptually, especially in relation to beliefs. For example, self-efficacy and collective efficacy are widely understood as belief-based constructs, as they reflect perceived capabilities (Bandura, 1997; **Bandura, A.** (1997). *Self-efficacy: The exercise of control*. New York: W. H. Freeman. ISBN: 978-0-7167-2850-4.). However, biospheric values are more stable guiding principles and are generally not considered beliefs, though they can influence them. Similarly, climate anxiety is better described as an affective state, often rooted in belief structures, but not a belief in itself. These distinctions do not need to be deeply unpacked in the manuscript, but a clearer framing would support the overall conceptual clarity, especially if all four are being discussed in relation to knowledge or behavior. But this should already be addressed in literature review.*

We agree with the reviewer that climate anxiety, biospheric values, and efficacy beliefs are distinct constructs and not forms of analytical reasoning. These variables were not central to our hypotheses but were included as covariates in the regression model to examine whether analytical reasoning predicted real-world action independently of these dispositional factors. We clarify this further on pages 9-10, and now write:

“The survey contained eight open-ended questions designed to assess both analytical reasoning about the climate crisis and also four Likert-scale items, capturing belief-based

constructs, core values, and affective states. These variables were climate anxiety (e.g., ‘I feel anxious about climate change’...), biospheric values..., self-efficacy..., and collective efficacy.... These variables were measured to be included as covariates to examine whether analytical reasoning predicted action independently of these factors.”

RB.27 Figure 2. Please indicate which line is control vs intervention. It is indicated in Figure 3 (next to the first interaction), but I do not think that this is the right place to indicate it.

We thank the reviewer for their suggestion. In our new figure, we have placed the legend at the top to indicate the groups.

RB.28. Procedure: The procedure section includes a lot of valuable detail, but in its current form it is quite difficult to follow. It combines procedural steps, measurement information, detailed intervention content, and coding descriptions all in one long stretch of text. As a result, it is hard to track what was actually done at each time point, what was measured (and how), and what belongs to the course-based intervention versus to the data collection procedure. The section would really benefit from being reorganized or broken into clearer parts, for example:

- (1) What participants did at T1 and T2;*
- (2) What was measured, distinguishing between qualitative and quantitative components;*
- (3) What the intervention involved (course structure, activities, goals); and*
- (4) How coding and scoring were conducted.*

Thank you for this helpful suggestion about the structure of the Procedure section. We have reorganized the section to align with the elements you proposed. Specifically, we have divided the section into the following subsections:

- 2.3.1 T1 and T2 Survey Measures**
- 2.3.2 T1 and T2 Coding of Analytical Reasoning**
- 2.3.3 The Class**
- 2.3.4 Climate Action Measure**
- 2.3.5 Climate Action Measure Coding**

This structure improves clarity while retaining all original content and ensures that the timeline, measurement strategies, intervention design, and coding procedures are clearly distinguished.

RB.29. Results

Since none of the predictors in the regression turned out to be significant, it is not entirely clear what was gained by using uncorrected p-values in the earlier t-tests. It looks like the idea was to keep more potential predictors, but in the end, it did not lead to a working model. It also creates some inconsistency in how statistical standards are applied across the results. A tighter approach, either sticking to corrected values or selecting predictors based on stronger theory, might have made the test of Hypothesis 2 clearer and easier to interpret.

See our response to RB.6. We now use Holm-Bonferroni corrected p-values for the t-tests as well.

RB.30. There is already a lot going on in the results section, especially with the mix of qualitative and quantitative findings. To help the main story land more clearly, it would make sense to focus on the key results that were part of the original plan. Some of the extra bits, like the Public Power group doing more actions, are interesting, but they might fit better in the supplementary materials. That way, the reader is not trying to hold too many threads at once, and the most important takeaways stay front and center.

We appreciate the reviewer's concern about maintaining clarity and focus in the Results section. However, we believe the finding that the Public Power group performed significantly more actions than other groups adds valuable context. While it was not part of our pre-registered hypotheses, this post hoc result provides useful fodder for thinking through the practical and structural drivers of action—such as ease of access and regularity of opportunities—which can inform both future research and real-world organizing strategies. For this reason, we have opted to retain this brief analysis in the main text, while ensuring that it does not detract from the central findings of the study.

Discussion

RB.31. Given how much still needs to be worked out around the core concepts, theoretical framing, and methods, it might not make sense to revise the discussion in depth just yet. But, even as it stands, the discussion does not feel well anchored in the broader literature. The interpretation leans more toward describing the results or offering general ideas, but it does not clearly show how the findings connect to what is already known. Once the main issues are addressed, the discussion will need to be revisited so it better reflects those changes and speaks more directly to existing research.

- L573-574 very technical information.

We respectfully disagree with the suggestion to remove the noted sentence. While the information is somewhat technical, we believe it is important to acknowledge as a concrete limitation of the control group data.

RB.32. L571-572 it is not immediately clear what is 'system 2'.

We thank the reviewer for pointing out this concern. On page 19, we cut the phrase “System II type information” and stick with “analytical reasoning.”

RB.33. The paper states, this was not a power issue because the scatterplots did not show any relationship between knowledge and action. But that alone does not really rule out power problems. With a modest sample and quite a few predictors in the regression, it is still possible the study did not have enough sensitivity to pick up smaller or more subtle effects. Even if things look flat in the plots, that does not always mean there is nothing there, it might just mean the study was not equipped to detect it. Although the authors ran a power analysis and reported adequate power to detect medium effects ($f^2 = .18$), the results suggest that any existing effects are likely

smaller than that threshold. Since none of the predictors came close to significance, it seems quite possible that the study was underpowered to detect smaller, but still meaningful, associations. So while the power analysis is a helpful addition, it does not fully support the claim that power was not an issue here.

We agree with the reviewer's concern and have revised the manuscript accordingly. When discussing the null results on page 20, we no include the following:

“First, given our modest sample and several predictors, it is entirely possible that smaller—but still meaningful—effects went undetected. Our power analysis indicated we had sufficient sensitivity to detect medium-sized effects ($f^2 = .18$) but not smaller ones. Thus, we encourage readers to interpret the null findings with caution.”

RB.34. Although the authors refer to their outcome as “real-world collective action,” it is important to note that the data still comes from participant self-report. The weekly reflections were structured and scored, which adds depth, but they are still based on what participants said they did rather than on any direct behavioral trace or consistent external verification. If any form of verification was used (e.g., confirmation from group leaders or documentation of actions), including that detail would really strengthen the claim. As it stands, it might be more accurate to frame this outcome as a detailed self-report of real-world engagement rather than a direct behavioral measure.

See our response to RB.5.

RB.35. The discussion does acknowledge the main null result, that none of the knowledge or belief variables predicted real-world action, which is an important finding. But most of the explanation leans on measurement issues or data limitations. It would be helpful to reflect on what this result means for the broader question the study set out to explore. If the aim was to see whether analytic knowledge drives action, and that did not hold up, that is something worth engaging with more directly, not just as a limitation, but as a possible insight.

We thank the reviewer for this thoughtful prompt. We agree that the null results speak to more than just methodological constraints, and in the discussion, we do engage with this point directly. As we write in the discussion (page 19):

“It could reflect measurement and statistical issues, or it could reflect a psychologically interesting insight into the limits of providing analytic information to drive action.”

Moreover, on page 21, we discuss the need to look at different, non-analytic variables from the organizational behavior literature for example.

RB.36. Some of the results are described as “striking” or “surprising,” but they come across as fairly modest, especially given the overall pattern of findings. For example, a small increase in collective efficacy in a class context is not exactly unexpected, and the lack of a relationship between knowledge and action, while disappointing, is not shocking either. Toning down the language a bit would help the discussion stay more in line with what the results actually show.

We have removed the word “striking” when discussing the collective efficacy result. That said, we continue to describe this finding as notable, given the well-documented difficulty of shifting efficacy beliefs in the climate domain, as discussed by Hornsey et al. (2021). We have also retained the use of “surprising” to describe the null relationship between analytical reasoning and action, as this was a preregistered hypothesis based on both relevant literature and theoretical expectations around how information might influence collective behavior. We believe this framing accurately reflects our initial predictions and the unexpected nature of the findings within that context.

RB.37. The discussion refers to the weekly reporting approach as “pioneering,” - that might be a bit strong. Structured weekly self-report is not uncommon in this type of research. It is a useful method and applied well here, but might be better described as a helpful contribution rather than something entirely new.

In the Abstract and Discussion sections we now replace “pioneers” with “develops”.

RB.38. There is a lot of statistical detail carried into the discussion section, exact p-values, references to scatterplots, and repeated mention of significance levels. APA style encourages keeping the technical results in the Results section and focusing the Discussion on interpretation and meaning.

We thank the reviewer for their recommendation and have made edits throughout the Discussion to avoid describing exact p-values and technical results.

RB.39. The manuscript does acknowledge a few key limitations, such as sample size, generalizability, and some measurement concerns, but the reflection remains fairly thin. Since this is a single intervention study, conducted in one course and relying entirely on self-report, it would help to frame the findings more clearly as preliminary and context-specific. There is also no mention of the need for replication of results, which seems like an important gap, especially given the exploratory nature of some analyses and the null findings on the main hypothesis (not all results were null). Regardless, right now, some of the conclusions feel a bit too confident given the limits of the design.

We appreciate the reviewer’s comment and agree that our framing could more clearly acknowledge the contextual and preliminary nature of the findings. In response, we now write the following on page XX of the discussion:

“Given the study’s context—one class at one university—these findings should be interpreted as preliminary. We strongly encourage replication and extension in more diverse contexts and with larger, more balanced samples to test the robustness and generalizability of these relationships. Future studies would also benefit from including an active control condition, such as a comparable class without climate content, to more rigorously isolate the effects of such interventions.”

RB.40. Other notes

The language is generally understandable, but there are many places where phrasing feels cumbersome or overly complex. In particular, there is a mix of formal (e.g., L9-12) and casual expressions that do not sit well together. Phrases like “research study” (L100) appear where simpler wording would be clearer, and some sentences are weighed down by redundancy or unclear structure. A focused language edit would help improve clarity and flow, especially for readers outside of the immediate research context.

L13, perhaps “knowledge aspects/facets” would be a better fit than “items”?

We thank the reviewer for their suggestion and replace “items” with “types of analytical reasoning.”

RB.41. L16, abstract. Is it considered the same construct? knowledge / belief elements

Per our response to RB.2 and RB.3, we replace “knowledge / belief elements” with “types of analytical reasoning.”

RB.42. L18-19, abstract. Too general for the abstract, vague formulation.

We thank the reviewer for their recommendation and have deleted the sentence “These results raise questions about how analytic knowledge of the climate and ecological predicament is measured, and how it is taught.”

RB.43. L25, abstract. Might be better for the sake of consistency to indicate “climate crisis” instead of climate change.

We thank the reviewer for their suggestion and replace “climate change” with “climate crisis.”

RB.44. To summarize, the topic is an important one, and there is clear potential of the script, but it still needs quite a bit of shaping before it can really deliver on that promise. The conceptual foundations need tightening, the methods need a clearer throughline, and the results need to be distilled so the reader is not left navigating through too many pieces. Once those core issues are addressed, the contribution will be easier to see and easier to stand behind.

I really appreciate the opportunity to review this paper. I know it is a lot to take in, but I truly hope the suggestions are helpful for the next round, that is the spirit they were offered in.

We are grateful for these kind words and believe the manuscript has significantly improved because of the reviewer’s suggestions.

Reviewer C

RC.1. This work tests a real-world educational intervention: a university class on the climate crisis. Before and after the class, students were measured on various knowledges and attitudes, and hypotheses were preregistered about an expected increase in knowledge, sense of collective efficacy, and real-world action. The results are important and caution the field to take self-report as the gold standard for behavior change: as others have pointed out, it is poorly correlated with real behavior. This is an important piece of work that contributes to recent 'barriers first' calls for behavior change in environmental psychology. I applaud the authors for doing this complex real-world study, which of course also comes with complexities that are hard to mitigate (such as unequal group sizes).

We are grateful for these kind words and tackle each of the reviewer's comments below.

RC.2. Major suggestions

- *Statistics: my main concern is the lack of an active control, and unequal sample sizes between conditions. I understand it's too late to change this now, but for the future I would strongly suggest an active control group (e.g. students taking a course around the same time that also includes some active group work, but on a different topic). This may allow stronger conclusions about the intervention, rather than with the passive 'no class' baseline now. Do we know anything about the demographic match between the control and intervention students (were the control participants also enrolled in the same program and taking similar classes, but just not enrolled in this particular class)?*
 - *For the future, it could be interesting to consider a variant of people only taking the lectures, or only engaging in group work. Now, the passive lectures and active social-change participation cannot be readily distinguished.*
 - *While I do not expect the authors to redo the data collection with an active control and/or more control participants (although I hope they do so for the next course iteration), there should be a broader discussion about statistical inference and potential statistical problems (e.g. power).*

We appreciate the reviewer's thoughtful recommendation. As noted in our response to RB.39, we now explicitly discuss the importance of using an active control group in future research to better isolate the specific effects of climate crisis education versus general engagement or course participation. We also highlight the need for larger and more demographically balanced samples. This broader framing, which appears on page xx of the Discussion, helps clarify the limits of the current design and our commitment to refining it in future iterations.

RC.3. data visualization: the figures could be much improved by adding individual data points (with connecting slopes), group error bars, and/or full distributions (for instance like a raincloud plot)? Perhaps using JASP as a free alternative to SPSS helps here, as they by default output more comprehensive graphs that show the individual variability of the data (see for example <https://jasp-stats.org/2021/10/05/raincloud-plots-innovative-data-visualizations-in-jasp>; <https://wellcomeopenresearch.org/articles/4-63> Figure R11).

We thank the reviewer for this suggestion regarding data visualization. In response, we have revised our figures to include group means and standard error bars for each condition and timepoint (see Figures 2 and 3), which improves clarity without overcomplicating the visual layout. We also explored the use of raincloud plots and generated several versions internally. However, after review, we felt that including both interaction plots and full rainclouds would make the manuscript overly visually dense, especially given the number of variables and timepoints plotted. That said, in our codebook in the supplementary materials, we have included directions for how to visualize our data in JASP using raincloud plots.

RC.4. replicability: I was able to download the data in SPSS format, but there is no codebook or publicly available instructions in a format that I could open (I do not have an SPSS license). I can open the .sav file in JASP, but without instructions I could not replicate the basic plots: there are many variables, not all of which are relevant for the figures (I think), and I did not see which variable corresponds to the T1/T2 measurement (or how to otherwise pair the T1-T2 datapoints within the same person). I strongly suggest adding a codebook for step-by-step replication of the main results, which could be aided if the raw data was complemented with a spreadsheet containing processed data (and ideally, steps that describe how the raw data is converted into the processed data).

Thank you for this extremely helpful suggestion. In response, we have created a dedicated replication codebook and instruction guide tailored for JASP users, which is now included in the supplementary materials. The codebook includes a full variable map showing how each construct corresponds to T1 and T2 variables, along with step-by-step instructions for running all key analyses—repeated-measures ANOVAs, paired-sample t-tests, regressions, and raincloud plots—entirely within JASP.

RC.5. Minor suggestions

- *in the introduction, 'before the class begun' is a bit abrupt. Before going to the measured target variables, perhaps explain briefly what class students are taking (elective/mandatory, BSc/MSc, class size, duration of the class), and why specifically an educational intervention is proposed (similar to the paragraph that is now at line 318; perhaps cite the Otto 2020 tipping points paper which lists education as one key leverage point)? It would help to briefly preview the type of intervention at the same level as all the outcome measures are justified in the introduction. For instance, it would help to briefly mention things like: were there also lectures next to the focus groups? What kinds of organizations were people assigned to? Now, some of this information is missing, or it comes very late in the paper.*

We thank the reviewer for this suggestion. In response, we have clarified early in the manuscript that the intervention was an undergraduate elective course that met twice weekly for lecture over ten weeks. We also believe we adequately justify the use of an educational intervention in the Introduction. Additionally, we provide more details about the organizations and the student's final projects in the supplemental materials.

RC.6. share the Qualtrics template, so others can reuse these more easily?

We thank the reviewer for this suggestion and have uploaded our OSF database with the T1 and T2 Qualtrics survey templates.

RC.7. research assistants who ranked knowledge and action scores: did they rate the responses blinded to the timepoint at which these were collected, and blind to the condition (class or control)? if not, may these rankers have been biased by knowing which responses were given at T1/T2?

We appreciate the reviewer's question, and now on pages X and X where we explain the coding procedures, we clarify that the research assistants were "blinded to condition and timepoint."

RC.8. <https://www.nature.com/articles/s44159-024-00305-0> and perhaps <https://gep.psychopen.eu/index.php/gep/article/view/13089> would be helpful to cite (especially the Albarracin paper which to me is the new gold standard reference for behavioral interventions).

We appreciate the reviewer's recommendation and integrate Albarracin et al.'s paper on page 4 of the Introduction.

RC.9. the preregistration mentions a linear model, but the main text uses an ANOVA. could you explain why? the linear model may be more powerful in grouping participants within action cohorts (to account for e.g. different group dynamics as a function of each group's coordinator etc).

We appreciate the reviewer's comment. While the preregistration referred to a "linear model," we note that repeated-measures ANOVAs are in fact a type of linear model—specifically, a general linear model for within-subject designs. We chose this approach for reporting because it provides a clear and interpretable framework for testing our pre-registered condition x time hypotheses. We agree that a multilevel linear model (e.g., with participants nested within action cohorts or coordinators) could offer additional analytic flexibility. However, we believe our modest sample size, especially with respect to group clusters, would not be sufficient to support such a complex model.

Reviews Round 2

Comments by the Editor

1. After having read the reviews, I had another look at your manuscript to see how the terminological issues identified by Reviewer B could be addressed. I left comments on this in the attached version of your manuscript. In that version, I also highlighted some smaller issues (e.g., formal/translation issues) that I would ask you to fix during the revision of your manuscript. You do not need to respond to those in-text comments in your response letter.
2. I agree with Reviewer B that the change to “analytical reasoning” as an umbrella concept renders the rationale confusing/inconsistent at many points. Talking about participants’ understanding (or knowledge) of different facets/aspects of the climate crisis might be more consistent and effective. If what you have assessed does not correspond perfectly to how understanding/knowledge is commonly assessed, describe how it relates to understanding/knowledge or belief-like constructs. More important than the perfect term would be a clear definition of what you mean by the term you select.
3. In line with Reviewer B, I think there might be inconsistencies between the text (or your analytical decisions) and the preregistration. Please carefully check again and make sure to report all analyses in the way they were preregistered. If certain analytical choices were not made at the preregistration stage, please state that explicitly and show how the results differ if different plausible analytical choices are made. If you want to add additional analyses that go beyond what is preregistered (e.g., the analysis of difference scores), please clearly label them as exploratory analyses.

Comments by Reviewer A (corresponds to Reviewer C in Round 1)

I thank the authors for updating their manuscript. I have only one major suggestion: why not add Bayes Factors (should be very easy if you’ve already replicated the analysis in JASP, simply switch from frequentist ANOVA to Bayesian ANOVA)? Beyond the power analysis (which is always quite sensitive to assumptions about effect sizes), Bayes Factors allow not just saying ‘we did not find significant evidence for this hypothesis’ but also ‘we find evidence for the null hypothesis of no effect’. Given that the main conclusions rely on the absence of effects, adding this small number to each test would allow a stronger conclusion on whether the results are ‘inconclusive’ or actually support the null hypothesis (as Reviewer B also points out). In an example, quickly running KCCI vs. KCCI, the BF10 of the interaction effect with condition seems inconclusive (0.017).

Bayes Factors also don’t require multiple comparison correction. This would allow you to much strengthen the discussion on ‘the main result of a null relationship between analytical reasoning and collective action.’ (l. 587) <https://link.springer.com/article/10.3758/s13423-020-01798-5>

I very much apologize for not thinking about this when I first read the paper - I have been recently adding Bayes Factors to one of my own projects and realized that it would be very helpful here too.

I hope this will further allow you to be more confident in what we can conclude from the data. Thank you for this work.

Minor

- line 386, what is ‘tabling’ in this context?
- could figure 2 have an indication of which panels are significant and which are not?

- are there plans to follow up with these students to test the long-term effects? Perhaps some seeds were planted by the class that would only show in e.g. students' career or educational choices a year later.
- As the authors mention (especially in section 3.3), social norms and group dynamics are probably a much larger determinant of real-world collective action. I wonder if this could be addressed using a nested/hierarchical linear model, where each group gets a random intercept (assuming that if there are strong group norms, then different students within a group should be more alike than across groups). In general, I feel like there is more to explore on the differences in collective action levels between organisations (and between groups within an organisation), which may give further pointers to what kind of variables organisers can take into account to build effective groups and coalitions (you discuss this qualitatively in the discussion). This dataset could perhaps be used as an exploration of some ideas to test in further follow-up studies (I realize you probably don't want to over-interpret or lengthen the list of non-preregistered ideas).
- Thank you for providing a codebook on the OSF. It would be great to mention somewhere in the methods where this codebook can be found (it does not seem to be referenced from the main text). It could also be nice to add in the methods that you used SPSS to generate the main figures, and that you found the same results with JASP (and refer to the supplement). On OSF, please describe on the main page what each file includes, and that the JASP instructions/codebook can be found inside the supplement (where it may not be found if someone just goes to OSF and needs to know where to start).

Comments by Reviewer B (corresponds to Reviewer B in Round 1)

The manuscript titled "The Impact of a Climate Crisis Class on Collective Action Participation" is a revised version of the paper I previously reviewed. I have carefully read the updated manuscript, other related materials and the authors' detailed response. I appreciate the effort put into addressing earlier concerns and providing updates. That said, I still see theoretical framing, particularly around the introduction of the term "analytical reasoning", as a core unresolved issue. If anything, bringing in this broad label to cover a diverse set of constructs seems to add more conceptual confusion rather than clarify the study's focus. Several other issues (i.e. conceptual clarity, definition of core concepts), from the first round also remain to be addressed, also a few new points emerged. I outline these in more detail below. I think the paper could get there, but these points need to be sorted out first before it is ready to be published.

Literature review/Theoretical reasoning

There is still a persisting issue with how one of the core constructs is conceptualized. In the first version of the manuscript, it was labeled "knowledge," but some of the components included go beyond what in psychology is typically defined as knowledge. In the current version, these components are grouped under "analytical reasoning," but this does not resolve the issue. According to the source the authors cite - Hornsey et al. - analytical reasoning refers to "effortful, deliberative, verbal processing of information, largely detached from emotion." The source explicitly distinguishes this form of reasoning from emotion-based or intuitive processes, which are not part of the analytical reasoning concept (and this is also true for the beliefs). Also, analytical reasoning describes a cognitive process, not a set of outcome variables. In this

manuscript, the term is used as an umbrella for a diverse mix of conceptual elements, some of which may potentially involve analytical thinking, others not. This creates a mismatch between the label and the content. Rather than relabeling, a more straightforward solution might be to name these multiple constructs, including **knowledge and beliefs**, and explain why it makes sense to consider them together. That would provide a more transparent and theoretically sound framing. Below I provide my understanding of each of eight components and how they could be defined and labeled. Please note, that I do not suggest that authors should strictly follow what I propose, but rather use this as a reference information/starting point to clarify and define concepts.

Given the body of literature on general psychology but also the fact that data on those eight components were collected based on rather qualitative methodology, it seems that “Climate change impacts”, “Climate-science based urgency”, “Climate Change Solutions”, “Barriers to Addressing Climate Change” could fall under the “knowledge” category. Given this is a qualitative study (open-ended prompts, no accuracy checks), and mental representations are coded, it is reasonable and justified to label this as knowledge. But it is important to be clear about what is meant by “knowledge” (i.e., subjectively held understandings, not factual correctness). From the quantitative paradigm perspective, those concepts could be labeled as beliefs since there is no accuracy check provided. It’s up to authors which conceptualisation will be chosen as long as it is robustly backed by literature. Next, “Climate Justice Framework” seems to be capturing distributive climate justice sensitivity guided by an equity-based moral frame. It could be stretched to a “belief” category (normative belief perhaps), but that would require robust theorizing. Further, “Role of Social Movements” seem to reflect a belief (i.e. conceptual stance or opinion on what movements should or do contribute). Finally, “Individual Climate Action” and “Collective Climate Action” both concepts reflect beliefs about potential behaviors, one at the individual level, the other at the collective level.

One component was removed after my comment from the first round saying that it does not seem to fit under the knowledge category. But now when authors provide more information in Table 1 it seems that many of those components might not exactly fit under the “knowledge” (let alone “analytical reasoning”) umbrella. My suggestion here is not to trim components if they do not match the larger pattern, but rather rethink based on the literature, which theoretical category each of them might fit in and indicate it.

If this conceptual unclarity is not addressed it will continue affecting other parts of the manuscript (**more examples in #Language & conceptual clarity**)

The manuscript includes climate anxiety, biospheric values, self-efficacy, and collective efficacy as covariates in the regression models. While these are not the primary variables of interest, they are theoretically relevant, as each may have implications for collective action. I propose briefly acknowledging this in the literature review, so readers can understand why these constructs were included in the analysis and how they connect to action/behavior.

Definitions

I was struggling to find clear definitions for key concepts such as **knowledge, beliefs, actions, collective action, etc.**, (but also biospheric value, efficacy, CC anxiety) which made it harder to fully grasp how these were understood throughout the paper.

Consistent use of the terminology

Throughout the manuscript, the terms "activism", "collective action", "climate movement", "social movements of activists and advocates", etc., are used almost interchangeably, but they are not quite the same. "Activism" usually refers to more intentional, organized efforts to push for change, while "collective action" is a broader term, it can include things like signing petitions, joining events, or other group-based efforts that are not necessarily activism. The overlap is understandable, and there is some debate in the literature, but it is important to pick one framing and stick with it consistently. It would really help with the clarity and flow of the text. Since the paper's title refers specifically to 'Collective Action Participation,' it might help to consistently use this term throughout the manuscript. I bring this up because even seemingly similar behaviors, like recycling plastic versus organic waste, are quite often predicted by different psychological constructs. The same logic applies to various forms of activism: engaging in one type does not always stem from the same motivations or beliefs as another. Clarifying which specific construct of activism is being focused on would help clarify which kinds of actions the intervention is expected to influence, and it would also make future replications (if any) more straightforward.

The same comment is valid for "climate change" and "climate crisis".

Conceptual clarity & some language issues

After revisions, issues with conceptual clarity and language remain. The manuscript still requires a careful, step-by-step read to catch inconsistencies in both argumentation and use of key terms. The examples below illustrate the broader pattern, but there is still a need for revision throughout the entire paper. The goal is not flawless English or unrelenting clarity, but just enough precision to make sure the ideas are understandable and key concepts are applied consistently:

Example 1: L10-13. Cumbersome sentence structure: "All were required to answer questions about their analytical reasoning, and biospheric values, beliefs about efficacy, and climate anxiety, at two time points (separated by nine weeks), while we measured real-world collective action in the intervention condition through the 9 weeks." The "while..." clause makes the timeline feel like it is contrasting, but here it is not really a contrast, it is just a parallel part of the design.

Example 2: L419-422. "For collective action beliefs, there was a significant interaction effect between condition and time $F(1, 126) = 10.933, p = .001, \eta^2 = .08$. Participants in the intervention showed a stronger ability to reason about collective action at T2 compared to T1 $t(100) = 6.134, p < .001, d = .61$, while those beliefs did not significantly change at T2 in the control condition $t(26) = .55, p = .587$." The phrasing here is conceptually confusing. Reasoning is a cognitive process; beliefs are psychological states. Presenting them as interchangeable outcomes, first referencing reasoning, then beliefs, blurs the constructs. It is unclear what was actually measured: beliefs, reasoning, or both. This is not a statistical issue, but one of clarity. Without consistent definitions, the meaning of the findings becomes difficult to interpret.

Example 3: L506-508. "For the multiple regression analysis, we also included as covariates the belief measures from T1 of climate anxiety, biospheric values, self-efficacy, and collective efficacy (on the view that we needed to try to control differences in baseline beliefs and values in our participants)." Values, anxiety and efficacy are still being grouped under the umbrella of "beliefs," despite concerns I raised in an earlier review. These are distinct constructs that have established definitions in the literature. For example, climate anxiety is an emotional state, and biospheric values reflect a general goal in life to take care about the

environment, neither fits the standard definition of a belief. I suggest using the original terms consistently and avoiding umbrella terms that blur conceptual clarity.

Example 4: L601. I still stand by my earlier point: it is confusing to call it “surprising” that certain elements did not change. That phrasing implies there is a long history of consistent findings that this result somehow contradicts. But given that this is a newly developed and first-time-tested program, there is no real precedent to be surprised about.

Example 5: L622-625. “Hornsey et al.’s argument, which was mainly about self-efficacy beliefs, was that providing analytical reasoning is insufficient to persuade people to act socially, even hypothetically; instead non-analytic information such as social identity, social norms and threat-levels are more important (...)”.

The current phrasing misrepresents Hornsey et al.’s original argument, where analytical reasoning is understood as a cognitive process, not a belief, and Hornsey’s work treats it as distinct from belief constructs. Some of the components included in the paper’s own reasoning seem to blur the line between analytical and non-analytical content, which makes the comparison to Hornsey’s claim problematic.

#Real world action

After carefully reviewing the full manuscript, I still find it difficult to see how the behavior measured in this study qualifies as real-world action. From what is presented, I would say it is a very well-designed and thoughtful self-report measure, one of the really good examples out there, with validation from multiple sources. But real-world action typically involves observable behaviors, ideally reported or verified by someone other than the participant. For example: measuring changes in recycling volume in a lab setup, weighing household waste before and after an intervention, or comparing energy usage through smart meters. In this study, however, the primary source of data is participant self-report. Given that, I would recommend reconsidering the use of the term “real-world action”.

Preregistration

The preregistered hypotheses focused on improvement in the quality of answers, suggesting a shift in depth, nuance and insight. But in the manuscript, this is reframed as an increase in analytical reasoning and beliefs, which leans toward quantity (e.g., more items mentioned or points scored). This appears to be a conceptual shift in what is being measured. Richer thinking and longer lists are not equivalent. While the rubric awards points for more detailed responses, it remains unclear whether the outcome captures better reasoning or more of it. Both approaches are perfectly valid, but since the preregistration emphasized quality, a brief acknowledgment or footnote explaining the shift would be helpful. Whether the mismatch is acceptable is ultimately an editorial decision, but I found it worth flagging.

In preregistration document

In the manuscript

H1

For Week 9 compared to Week 1, participants in the intervention condition will show improvement in the quality of their answers related to the climate crisis and its solutions, compared to participants in the control condition

Analytical reasoning and beliefs will increase from T1 to T2 for participants in the intervention condition, and more than compared to the control condition

H2

In the intervention condition, the quality of answers related to the climate crisis and its solutions will be positively related to the level of the participant's collective action participation

For participants in the intervention, the change in analytical reasoning quality from T1 to T2 will be associated with greater engagement in real-world collective action

#Precision of results interpretation

"Although some literature suggests that biospheric values are quite stable (Martin & Czellar, 2017), this result suggests that the ecological features of the class content did change people's ability to rate their personal identification with nature" (L635-637). It is established that values might fluctuate because of the short term influences, but this does not necessarily mean that they change in the long run. Classical studies from Shalom H. Schwartz's team shows that things people consistently value, need really strong stimulation to induce lasting change. Another flaw in the argument is saying that "...class content did change people's ability to rate their personal identification with nature", but biospheric values are about general goals in life not identification processes (identification is of course part of the process but is not measured in this study).

Technical issues

L92 and elsewhere. "(e.g. (Agostini & van Zomeren, 2021; Bamberg, Rees et al., 2018; Dablander, Sachisthal et al., 2024; Hornsey, Blackwood et al., 2006; Roser-Renouf, Maibach et al., 2014; Van Zomeren, Postmes et al., 2008))". 'E.g.' should not be outside the brackets.

L143. The last two rows of Table 1 use the same question to address individual and collective action.

L154. Spelling error "chang" ('e' is missing).

L226-229. Belongs to the method.

L541-543. Belongs to the discussion.

Discussion. Repeating technical results in the Discussion does not add much and does not quite line up with APA guidelines, which are more about interpretation than repetition. Of course, I get that this might be something the editor decides in the end.

#Final/summary comment

- There are eight components (nine in the earlier version) targeted in the intervention. One possible way to address them in the literature review is to identify a theoretical framework that can encompass them (unfortunately, "analytical reasoning" does not align well with the content). Alternatively, they can be presented as stand-alone concepts, but it is important to explain why it makes sense to consider them together.
- Each component should be clearly defined based on existing literature and allocated to a recognisable category (e.g., knowledge, belief, justice sensitivity, etc.).
- Climate anxiety, biospheric values, self-efficacy and collective efficacy are also worth mentioning in the literature review, along with definitions.
- It would be very helpful for the reader if key concepts were used consistently across the manuscript.

-
- Consider relabeling real-world action to robust self-report.
 - It is also concerning that the framing of the hypothesis in the preregistration does not match the framing in the paper.

Thank you for the opportunity to revise your work. I believe these are my final suggestions, largely reiterating what was raised in the first review, but this time more precisely outlining what I see as problematic. With these comments I meant to contribute to strengthening the theoretical background of the manuscript.

Author response to the reviewers Round 2

Dear Dr. Lange,

We are extremely grateful for your hard and thorough work helping us with this manuscript. Below we reply to each of your comments and those of your reviewers.

We also received the PDF document with your suggested textual changes, and have adopted most of your suggestions.

Sincerely,

Raihan Alam and Dr. Adam Aron

Editor comments:

1. After having read the reviews, I had another look at your manuscript to see how the terminological issues identified by Reviewer B could be addressed. I left comments on this in the attached version of your manuscript. In that version, I also highlighted some smaller issues (e.g., formal/translation issues) that I would ask you to fix during the revision of your manuscript. You do not need to respond to those in-text comments in your response letter.

We now take Reviewer B's suggestion to a) label them as "knowledge and beliefs about the climate crisis", b) clarify which items are knowledge and which are beliefs, and c) say what we mean by "knowledge" and "belief".

Thank you for your comments in the manuscript; we have incorporated most of them.

2. I agree with Reviewer B that the change to "analytical reasoning" as an umbrella concept renders the rationale confusing/inconsistent at many points. Talking about participants' understanding (or knowledge) of different facets/aspects of the climate crisis might be more consistent and effective. If what you have assessed does not correspond perfectly to how understanding/knowledge is commonly assessed, describe how it relates to understanding/knowledge or belief-like constructs. More important than the perfect term would be a clear definition of what you mean by the term you select.

See our response to your first comment.

3. In line with Reviewer B, I think there might be inconsistencies between the text (or your analytical decisions) and the preregistration. Please carefully check again and make sure to report all analyses in the way they were preregistered. If certain analytical choices were not made at the preregistration stage, please state that explicitly and show how the results differ if different plausible analytical choices are made. If you want to add additional analyses that go beyond what is preregistered (e.g., the analysis of difference scores), please clearly label them as exploratory analyses.

See answers below to Reviewer B.

Moreover, we have included the other changes you requested:

- All d, t, and F value now have a leading zero
- On page 17 we now mention the lack of pre-registered specificity in our analysis of H2:
 “We did not pre-register this analysis at the level of detail reported here (e.g., restricting predictors to significant T2–T1 difference scores within the intervention condition and standardizing variables prior to regression).”
- Instead of 50% more collective actions we now say 50% “higher quality” action in reference to the comparison of actions between groups
- We now mention that the increase in biospheric values could also reflect a priming effect or demand effect.

 Reviewer A:

I thank the authors for updating their manuscript. I have only one major suggestion: why not add Bayes Factors (should be very easy if you've already replicated the analysis in JASP, simply switch from frequentist ANOVA to Bayesian ANOVA)? Beyond the power analysis (which is always quite sensitive to assumptions about effect sizes), Bayes Factors allow not just saying 'we did not find significant evidence for this hypothesis' but also 'we find evidence for the null hypothesis of no effect'. Given that the main conclusions rely on the absence of effects, adding this small number to each test would allow a stronger conclusion on whether the results are 'inconclusive' or actually support the null hypothesis (as Reviewer B also points out). In an example, quickly running KCCI vs. KCCI, the BF10 of the interaction effect with condition seems inconclusive (0.017).

Bayes Factors also don't require multiple comparison correction. This would allow you to much strengthen the discussion on 'the main result of a null relationship between analytical reasoning and collective action.' (l. 587) <https://link.springer.com/article/10.3758/s13423-020-01798-5>

I very much apologize for not thinking about this when I first read the paper - I have been recently adding Bayes Factors to one of my own projects and realized that it would be very helpful here too.

I hope this will further allow you to be more confident in what we can conclude from the data. Thank you for this work.

We thank the reviewer for their suggestion. We have now added Table 6 (our pre-registered test of H2) to the Supplementary Materials with Bayes Factor results. Those Bayes Factor results are consistent with the frequentist results and, as you suggest, provide an estimate of the degree of evidence in favor of the null hypothesis for all variables but biospheric values. We think it would be too confusing to report both frequentist and Bayes Factor results in the main manuscript, so we keep the Bayes results in the Supplementary Material.

Minor

- line 386, what is 'tabling' in this context?

“Standing at a table on the main library walk” (page 12).

- could figure 2 have an indication of which panels are significant and which are not?

We thank the reviewer for the suggestion. The revised figure now includes stars to indicate significance.

- are there plans to follow up with these students to test the long-term effects? Perhaps some seeds were planted by the class that would only show in e.g. students' career or educational choices a year later.

Not as part of this paper, but the reviewer is correct that this class has had quite an impact on some students. Over the years many have gone to journalism, policy (e.g. in the California state legislature) and several are now studying environmental law. In future iterations of this design we can indeed try to track students over time and relate their outcomes back to their Knowledge and Beliefs during the study.

- As the authors mention (especially in section 3.3), social norms and group dynamics are probably a much larger determinant of real-world collective action. I wonder if this could be addressed using a nested/hierarchical linear model, where each group gets a random intercept (assuming that if there are strong group norms, then different students within a group should be more alike than across groups). In general, I feel like there is more to explore on the differences in collective action levels between organisations (and between groups within an organisation), which may give further pointers to what kind of variables organisers can take into account to build effective groups and coalitions (you discuss this qualitatively in the discussion). This dataset could perhaps be used as an exploration of some ideas to test in further follow-up studies (I realize you probably don't want to over-interpret or lengthen the list of non-preregistered ideas).

We agree with the reviewer that adding more non-preregistered analyses would clutter the paper. We encourage researchers to examine this model with our accessible data.

- Thank you for providing a codebook on the OSF. It would be great to mention somewhere in the methods where this codebook can be found (it does not seem to be referenced from the main text). It could also be nice to add in the methods that you used SPSS to generate the main figures, and that you found the same results with JASP (and refer to the supplement). On OSF, please describe on the main page what each file includes, and that the JASP instructions/codebook can be found inside the supplement (where it may not be found if someone just goes to OSF and needs to know where to start).

We appreciate the reviewer's suggestions and on page 8 mention that this codebook can be found in the Supplemental Materials, and that data can be analyzed with the freely available JASP. We also provide instructions to do that. Moreover, we now also mention that we conducted analyses in SPSS and produced figures in R. We also include a more detailed description of the OSF components in the repository.

 Reviewer B:

The manuscript titled “The Impact of a Climate Crisis Class on Collective Action Participation” is a revised version of the paper I previously reviewed. I have carefully read the updated manuscript, other related materials and the authors’ detailed response. I appreciate the effort put into addressing earlier concerns and providing updates. That said, I still see theoretical framing, particularly around the introduction of the term “analytical reasoning”, as a core unresolved issue. If anything, bringing in this broad label to cover a diverse set of constructs seems to add more conceptual confusion rather than clarify the study’s focus. Several other issues (i.e. conceptual clarity, definition of core concepts), from the first round also remain to be addressed, also a few new points emerged. I outline these in more detail below. I think the paper could get there, but these points need to be sorted out first before it is ready to be published.

Overall, we really appreciate this reviewer’s thoughtfulness on these definitional and conceptual issues.

Literature review/Theoretical reasoning

There is still a persisting issue with how one of the core constructs is conceptualized. In the first version of the manuscript, it was labeled “knowledge,” but some of the components included go beyond what in psychology is typically defined as knowledge. In the current version, these components are grouped under “analytical reasoning,” but this does not resolve the issue. According to the source the authors cite - Hornsey et al. - analytical reasoning refers to “effortful, deliberative, verbal processing of information, largely detached from emotion.” The source explicitly distinguishes this form of reasoning from emotion-based or intuitive processes, which are not part of the analytical reasoning concept (and this is also true for the beliefs). Also, analytical reasoning describes a cognitive process, not a set of outcome variables. In this manuscript, the term is used as an umbrella for a diverse mix of conceptual elements, some of which may potentially involve analytical thinking, others not. This creates a mismatch between the label and the content. Rather than relabeling, a more straightforward solution might be to name these multiple constructs, including **knowledge and beliefs**, and explain why it makes sense to consider them together. That would provide a more transparent and theoretically sound framing. Below I provide my understanding of each of eight components and how they could be defined and labeled. Please note, that I do not suggest that authors should strictly follow what I propose, but rather use this as a reference information/starting point to clarify and define concepts.

Given the body of literature on general psychology but also the fact that data on those eight components were collected based on rather qualitative methodology, it seems that “Climate change impacts”, “Climate-

science based urgency”, “Climate Change Solutions”, “Barriers to Addressing Climate Change” could fall under the “knowledge” category. Given this is a qualitative study (open-ended prompts, no accuracy checks), and mental representations are coded, it is reasonable and justified to label this as knowledge. But it is important to be clear about what is meant by “knowledge” (i.e., subjectively held understandings, not factual correctness). From the quantitative paradigm perspective, those concepts could be labeled as beliefs since there is no accuracy check provided. It’s up to authors which conceptualisation will be chosen as long as it is robustly backed by literature. Next, “Climate Justice Framework” seems to be capturing distributive climate justice sensitivity guided by an equity-based moral frame. It could be stretched to a “belief” category (normative belief perhaps), but that would require robust theorizing. Further, “Role of Social Movements” seem to reflect a belief (i.e. conceptual stance or opinion on what movements should or do contribute). Finally, “Individual Climate Action” and “Collective Climate Action” both concepts reflect beliefs about potential behaviors, one at the individual level, the other at the collective level.

We appreciate the reviewer’s suggestions and have adopted them in the revised manuscript. Specifically, we now a) refer to the key variables as “knowledge and beliefs about the climate crisis”, b) we clarify which are knowledge and which are beliefs in Table 1 and c) clarify what we mean by “knowledge” and “belief” on page 4. There, we write:

“By “knowledge,” we refer to accumulated scientific and lived understanding of such constructs such as “climate impacts” (Ford et al., 2016). By “beliefs” we refer to a different set of constructs such as “climate justice” which, amongst other things, is a kind of thinking that guides our responses to the climate crisis (Schlosberg & Collins, 2014). In the case of knowledge and beliefs, we aimed to derive and score them from coding participants’ open-ended textual responses to class prompts.”

One component was removed after my comment from the first round saying that it does not seem to fit under the knowledge category. But now when authors provide more information in Table 1 it seems that many of those components might not exactly fit under the “knowledge” (let alone “analytical reasoning”) umbrella. My suggestion here is not to trim components if they do not match the larger pattern, but rather rethink based on the literature, which theoretical category each of them might fit in and indicate it.

We have elected to keep the key results as is, and not rewind. There were three rationales for “trimming” that element. First, we realized it was confusing to have 8 textual prompts and 9 variables; Second, the element was “confidence in future outcomes” which didn’t seem like a construct that clearly really fits knowledge or belief either. Third, the original construct appears to be closely related to the “climate crisis anxiety” variable we measured using the Likert scale, which made it feel redundant.

If this conceptual unclarity is not addressed it will continue affecting other parts of the manuscript **(more examples in #Language & conceptual clarity)**

The manuscript includes climate anxiety, biospheric values, self-efficacy, and collective efficacy as covariates in the regression models. While these are not the primary variables of interest, they are theoretically relevant,

as each may have implications for collective action. I propose briefly acknowledging this in the literature review, so readers can understand why these constructs were included in the analysis and how they connect to action/behavior.

We appreciate the reviewer's suggestion and have now included a paragraph regarding this on page 8, where we write:

“We also measured four other variables using Likert scales: climate crisis anxiety, biospheric values and self and collective efficacy beliefs. These variables were included as covariates in our analysis to examine whether knowledge and beliefs related to verified collective action independently of these factors. Several studies suggest that baseline levels of these variables might relate to the amount of action participants do, and that changes in these variables might occur with changes in beliefs or knowledge (see Schwartz et al., 2023 for climate crisis anxiety; see Brick & Lewis, 2016 for biospheric values, and see Bamberg et al., 2015 and Sarasin et al., 2022 for efficacy beliefs).”

In the remainder of the manuscript we now refer to these variables as “covariates”.

Definitions

I was struggling to find clear definitions for key concepts such as **knowledge, beliefs, actions, collective action, etc.**, (but also biospheric value, efficacy, CC anxiety) which made it harder to fully grasp how these were understood throughout the paper.

See our response above. We have now adopted the reviewer's suggestion about Knowledge and Belief, and we identified them and defined them on page 4 and in Table 1.

Consistent use of the terminology

Throughout the manuscript, the terms "activism", "collective action", "climate movement", "social movements of activists and advocates", etc., are used almost interchangeably, but they are not quite the same. "Activism" usually refers to more intentional, organized efforts to push for change, while "collective action" is a broader term, it can include things like signing petitions, joining events, or other group-based efforts that are not necessarily activism. The overlap is understandable, and there is some debate in the literature, but it is important to pick one framing and stick with it consistently. It would really help with the clarity and flow of the text. Since the paper's title refers specifically to 'Collective Action Participation,' it might help to consistently use this term throughout the manuscript. I bring this up because even seemingly similar behaviors, like recycling plastic versus organic waste, are quite often predicted by different psychological constructs. The same logic applies to various forms of activism: engaging in one type does not always stem from the same motivations or beliefs as another. Clarifying which specific construct of activism is being focused on would help clarify which kinds of actions the intervention is expected to influence, and it would also make future replications (if any) more straightforward.

The same comment is valid for “climate change” and “climate crisis”.

We have adopted this suggestion and have replaced references to “climate change” with “climate crisis” and have likewise tried to use “collective action” almost everywhere throughout the manuscript.

Conceptual clarity & some language issues

After revisions, issues with conceptual clarity and language remain. The manuscript still requires a careful, step-by-step read to catch inconsistencies in both argumentation and use of key terms. The examples below illustrate the broader pattern, but there is still a need for revision throughout the entire paper. The goal is not flawless English or unrelenting clarity, but just enough precision to make sure the ideas are understandable and key concepts are applied consistently:

Example 1: L10-13. Cumbersome sentence structure: “All were required to answer questions about their analytical reasoning, and biospheric values, beliefs about efficacy, and climate anxiety, at two time points (separated by nine weeks), while we measured real-world collective action in the intervention condition through the 9 weeks.” The “while...” clause makes the timeline feel like it is contrasting, but here it is not really a contrast, it is just a parallel part of the design.

We appreciate the reviewer’s attention to detail and have removed the “while” with “and” in that sentence.

Example 2: L419-422. “For collective action beliefs, there was a significant interaction effect between condition and time $F(1, 126) = 10.933, p = .001, \eta^2 = .08$. Participants in the intervention showed a stronger ability to reason about collective action at T2 compared to T1 $t(100) = 6.134, p < .001, d = .61$, while those beliefs did not significantly change at T2 in the control condition $t(26) = .55, p = .587$.” The phrasing here is conceptually confusing. Reasoning is a cognitive process; beliefs are psychological states. Presenting them as interchangeable outcomes. first referencing reasoning, then beliefs, blurs the constructs. It is unclear what was actually measured: beliefs, reasoning, or both. This is not a statistical issue, but one of clarity. Without consistent definitions, the meaning of the findings becomes difficult to interpret.

We thank the reviewer for their attention to detail and have replaced “stronger ability to reason about collective action” with “increased belief in collective action” in that sentence.

Example 3: L506-508. “For the multiple regression analysis, we also included as covariates the belief measures from T1 of climate anxiety, biospheric values, self-efficacy, and collective efficacy (on the view that we needed to try to control differences in baseline beliefs and values in our participants).” Values, anxiety and efficacy are still being grouped under the umbrella of “beliefs,” despite concerns I raised in an earlier review. These are distinct constructs that have established definitions in the literature. For example, climate anxiety is an emotional state, and biospheric values reflect a general goal in life to take care about the environment, neither fits the standard definition of a belief. I suggest using the original terms consistently and avoiding umbrella terms that blur conceptual clarity.

We thank the reviewer for their comments and have removed the reference to “beliefs” in that sentence.

Example 4: L601. I still stand by my earlier point: it is confusing to call it “surprising” that certain elements did not change. That phrasing implies there is a long history of consistent findings that this result somehow contradicts. But given that this is a newly developed and first-time-tested program, there is no real precedent to be surprised about.

We take the reviewer’s point and have removed all mentions of “surprise” or “surprising” from the manuscript.

Example 5: L622-625. “Hornsey et al.’s argument, which was mainly about self-efficacy beliefs, was that providing analytical reasoning is insufficient to persuade people to act socially, even hypothetically; instead non-analytic information such as social identity, social norms and threat-levels are more important (...)”. The current phrasing misrepresents Hornsey et al.’s original argument, where analytical reasoning is understood as a cognitive process, not a belief, and Hornsey’s work treats it as distinct from belief constructs. Some of the components included in the paper’s own reasoning seem to blur the line between analytical and non-analytical content, which makes the comparison to Hornsey’s claim problematic.

We appreciate the reviewer’s point and have now changed that part of the text to state:

“Moreover Hornsey et al. (2021) argued that knowledge or reasoning is often insufficient to persuade people to act socially, even hypothetically; instead other kinds of variables such as social identity, social norms and threat-levels are more important (also see Angill-Williams & Davis, 2022 and Whitmarsh, Poortinga, & Capstick, 2021).”

#Real world action

After carefully reviewing the full manuscript, I still find it difficult to see how the behavior measured in this study qualifies as real-world action. From what is presented, I would say it is a very well-designed and thoughtful self-report measure, one of the really good examples out there, with validation from multiple sources. But real-world action typically involves observable behaviors, ideally reported or verified by someone other than the participant. For example: measuring changes in recycling volume in a lab setup, weighing household waste before and after an intervention, or comparing energy usage through smart meters. In this study, however, the primary source of data is participant self-report. Given that, I would recommend reconsidering the use of the term “real-world action”.

We take the reviewer’s point and have replaced the term “real-world collective action” with “verified collective action” throughout the manuscript.

Preregistration

The preregistered hypotheses focused on improvement in the quality of answers, suggesting a shift in depth, nuance and insight. But in the manuscript, this is reframed as an increase in analytical reasoning and beliefs, which leans toward quantity (e.g., more items mentioned or points scored). This appears to be a conceptual

shift in what is being measured. Richer thinking and longer lists are not equivalent. While the rubric awards points for more detailed responses, it remains unclear whether the outcome captures better reasoning or more of it. Both approaches are perfectly valid, but since the preregistration emphasized quality, a brief acknowledgment or footnote explaining the shift would be helpful. Whether the mismatch is acceptable is ultimately an editorial decision, but I found it worth flagging.

We appreciate the reviewer's attention to detail and now include the following footnote on page 8: "In our pre-registered document we used the language "quality of answers related to the climate crisis and its solutions", however, based on extensive discussions with reviewers, we now refer to "knowledge and beliefs about the climate crisis".

Also, looking at our pre-registration document we noticed that there was a lack of clarity around H2, as we did not specify that it was the "change in quality of answers" rather than "quality of answers" themselves. But this is implicit in our entire study design which is to look at how the change in acquisition of knowledge and beliefs in a climate crisis class would relate to action. See very first sentence of the synopsis: "The main aim of the study is to teach participants about the climate crisis and the importance of activism in combating it and to see how such knowledge relates to actual collective action behavior". We have now fixed this issue both in the footnote and in the Abstract of the paper.

So the footnote now also states on page 8: "Also, in the pre-registration, for H2, we did not specify precisely enough that our intent was to test how the *change* in knowledge and beliefs related to action – even though this was clear in the first line of the pre-registration: "The main aim of the study is to teach participants about the climate crisis and the importance of activism in combating it and to see how such knowledge relates to actual collective action behavior".

We now reproduce the original language of H1 and H2 while adding these clarifications in brackets.

#Precision of results interpretation

"Although some literature suggests that biospheric values are quite stable (Martin & Czellar, 2017), this result suggests that the ecological features of the class content did change people's ability to rate their personal identification with nature" (L635-637). It is established that values might fluctuate because of the short term influences, but this does not necessarily mean that they change in the long run. Classical studies from Shalom H. Schwartz's team shows that things people consistently value, need really strong stimulation to induce lasting change. Another flaw in the argument is saying that "...class content did change people's ability to rate their personal identification with nature", but biospheric values are about general goals in life not identification processes (identification is of course part of the process but is not measured in this study).

We take the reviewer's point and have removed the reference to Martin & Czellar, 2017 that we used to imply that biospheric values are difficult to influence. Moreover, we have changed the other sentences to now say: "This finding is notable given that biospheric values were not a direct target of the class, suggesting that the ecological framing of the content may have shifted

participants' relationship with, or understanding of, nature. Alternatively, the effect may reflect priming of environmental concepts or a demand characteristic."

Technical issues

L92 and elsewhere. "(e.g. (Agostini & van Zomeren, 2021; Bamberg, Rees et al., 2018; Dablander, Sachisthal et al., 2024; Hornsey, Blackwood et al., 2006; Roser-Renouf, Maibach et al., 2014; Van Zomeren, Postmes et al., 2008)))". 'E.g.' should not be outside the brackets.

We thank the reviewer for their attention to detail and have made sure "e.g." is not outside the brackets throughout the manuscript.

L143. The last two rows of Table 1 use the same question to address individual and collective action.

This is not a mistake. We used the same question to gauge the extent to which participants identified individual and/or collective actions they could take.

L154. Spelling error "chang" ('e' is missing).

We thank the reviewer for their attention to detail and have fixed the typo.

L226-229. Belongs to the method.

We thank the reviewer for their recommendation and now include the Study Overview under the Methods section.

L541-543. Belongs to the discussion.

Discussion. Repeating technical results in the Discussion does not add much and does not quite line up with APA guidelines, which are more about interpretation than repetition. Of course, I get that this might be something the editor decides in the end.

We thank the reviewer for their comment; however, we believe that our mention of the results in the Discussion section summarizes the findings and helps orient the reader to our discussion points.

#Final/summary comment

- There are eight components (nine in the earlier version) targeted in the intervention. One possible way to address them in the literature review is to identify a theoretical framework that can encompass them (unfortunately, "analytical reasoning" does not align well with the content). Alternatively, they can be presented as stand-alone concepts, but it is important to explain why it makes sense to consider them together.

In our manuscript, we consider each component separately in the literature review.

- Each component should be clearly defined based on existing literature and allocated to a recognisable category (e.g., knowledge, belief, justice sensitivity, etc.).

We have taken the reviewer's suggestion above about referring to them as "knowledge and beliefs about the climate crisis" rather than "knowledge elements" in our first submission and "analytical reasoning" in our second submission. We allocate them to each broader category in Table 1.

- Climate anxiety, biospheric values, self-efficacy and collective efficacy are also worth mentioning in the literature review, along with definitions.

See our response above. We have done this, and refer to them throughout the manuscript now as covariates.

- It would be very helpful for the reader if key concepts were used consistently across the manuscript.

We appreciate the reviewer's comment and have tried to reference the concepts consistently in the revised manuscript.

- Consider relabeling real-world action to robust self-report.

See our response above. We now use the term "verified collective action".

- It is also concerning that the framing of the hypothesis in the preregistration does not match the framing in the paper.

See our response above. We have now added a footnote about the minor discrepancy.

Thank you for the opportunity to revise your work. I believe these are my final suggestions, largely reiterating what was raised in the first review, but this time more precisely outlining what I see as problematic. With these comments I meant to contribute to strengthening the theoretical background of the manuscript.

We thank the reviewer for their helpful comments and suggestions.