

Dear Corina,

Thank you for highlighting these points and providing us with the opportunity to address these further. Thank you also for your patience while we revised our manuscript, we appreciate the flexibility as we worked through these revisions. We have now revised the manuscript further and respond to the comments below.

Kind regards,  
The authors

1) Neil Lewis, Jr. noted that “It could be beneficial to connect the current results with those broader calls about what is necessary for moving the open science movement forward”. In your revision, you added only one general sentence that briefly cited other articles. Your results could be more meaningfully connected with the broader literature if you provided more detail and delved more in to the interesting practical and theoretical aspects of these connections.

**Response:** Thank you for this recommendation. We have reviewed our Discussion and our connections with the literature and note that we have previously highlighted some noteworthy practical and theoretical connections:

- Lines 12-13: We note the evidence for the value of the COM-B behaviour change approach (Norris & O’Connor, 2019).
- Lines 18-20: We note the relationship between obstacles to adopting preregistration expressed by students and those of established researchers (Toth et al., 2021).
- Lines 46-48: We respond directly to the call of Pownall et al. (2022) and present the first study using quasi-experimental methods to understand the impact of preregistration on student outcomes.
- Lines 73-78: We note potential discrepancies in preregistration experiences for students and how these may parallel differences in preregistration for researchers (Bakker et al., 2020).
- Lines 108-111: We note the barriers to preregistration for researchers in general (Spitzer & Mueller, 2023) and how these may be tied to the experience for students.
- Lines 113-115: We build upon the ongoing metascientific efforts to maximise the potential of open science practices (Gervais et al., 2021; Suls et al., 2022).

However, we agree that we could strengthen our argument that open science should be embedded into higher education with further connections to the wider literature and have added some additional text to this effect:

“Our findings also contribute to the case that open science should be embedded into higher education for improved student scientific literacy and confidence (see Pownall et al., 2023 for a review). In response to the UK House of Commons Science and Technology Committee’s call for evidence of the contributors to research integrity,

the Framework for Open and Reproducible Research Training (FORRT) argues the importance of the pedagogical consequences of how we teach, mentor, and supervise students (Azevedo et al., 2021). A wide range of resources have recently been developed to support student learning of open science, including a student guide (Pennington, 2023), the FORRT project materials (Azevedo et al., 2023), and the Collaborative Replications and Education Project (CREP; Wagge et al., 2019). These efforts aim to strengthen student knowledge and engagement in research in order to become more savvy consumers of science (Korbmacher et al., 2023).”

2) Lisa Spitzer expressed interest in “exploratory analyses of students who wanted to preregister but then did not. Perhaps it might be interesting to look at their results of capability, opportunity, and motivation?” In your response, you declined to conduct the analyses due to “keeping the paper within scope” and a “very long word count”. At PCI RR, there are no word limits, and, while you might have a target journal in mind that imposes word limits, your article at PCI RR is independent from this. I err in favor of adding value to the research to get as much as you can out of all of your hard work. While it is entirely your choice about whether you conduct a post-hoc analysis, you might consider whether this would add value to the data you were able to collect and, if you feel it relevant, go ahead with it. To be clear, you don't need to re-address this point in your revision, I just wanted to bring it up in the context of there not being any word limits at PCI RR so you are free to conduct the analysis if you want to.

**Response:** Thank you for this recommendation. The number of students who indicated that they planned to preregister but then did not preregister is low ( $n = 8$ ). However, we have run this analysis comparing students who planned to preregister and did not ( $n = 8$ ) versus did ( $n = 29$ ) and have reported this in-text:

“As a final exploratory analysis, we explored whether there were differences in capability, motivation, and opportunity for preregistration between the students who indicated at Time 1 that they initially *planned* to preregister and then at Time 2 did not ( $n = 8$ ) versus did preregister ( $n = 29$ ). Independent samples t-tests showed that there was no difference in reported opportunity ( $t(7.52) = 1.79, p = .057$ ) or motivation ( $t(35) = .58, p = .28$ ), but there was a small but significant difference between capability, such that students who planned to

preregister and then did preregister rated their capability to be higher ( $M = 4.48, SD = .738$ ) than students who planned but did not preregister ( $M = 4.0, SD = .6$ ),  $t(35) = 1.7, p = .049$ .”

3) The anonymous reviewer had doubts about your Stage 2 sample size ( $n=89$ : 52=experimental group, 37=control group) being much smaller than what was expected ( $n=200$ : 100=experimental group, 100=control group) at Stage 1. Here is the reviewer's comment:

“My most serious concern with this Stage 2 report is the drastic difference in the planned and achieved sample size. While the Stage 1 proposed to collect of final sample of 200, with 100 participants in each group, the final sample comprised less than half of this planned amount, and only 37 subjects in one group. I appreciate that this study was subject to recruitment and retention issues, and that the study was conducted under time pressure, but this strikes me as a major drawback in the Registered Report context. What is the achieved power, based on the analyzed sample, for the effect size previously proposed at Stage 1? This concerns me both in terms of the reliability of the observed effects, as well as our ability to confidently interpret the null findings.”

This calls into question whether your Stage 2 meets the review criterion “2C. Whether the authors adhered precisely to the registered study procedures” and, consequently, “2E.

Whether the authors' conclusions are justified given the evidence”

([https://rr.peercommunityin.org/help/guide\\_for\\_recommenders#h\\_6759646236401613643390905](https://rr.peercommunityin.org/help/guide_for_recommenders#h_6759646236401613643390905)). The much smaller sample size was not discussed with PCI RR as this deviation was starting to unfold during the course of the research. I would like a full justification about 1) whether this small sample size is a deviation from the Stage 1 plan and, if so, 2) explain exactly how it differs from the Stage 1 plan and why you think it is still scientifically valid using the details set out in your power analyses, as well as any other pieces of evidence that can show this. 3) If your small sample size is not a deviation from the Stage 1 plan, please explain exactly why and use details from your Stage 1 power analysis (as well as any other evidence you can bring to bear on the issue) to show why this is the case. A summary of these details should also be included in the article to help readers understand this point because this question will come up for future readers as it already has during the review process. I appreciate that you attempted to address this comment in your response, however, there isn't enough detail in your response for me to be able to empirically evaluate whether your article meets the above two Stage 2 review criteria.

**Response:** Thank you for this comment. We acknowledged that sample size would be a risk in the Stage 1 manuscript and, importantly, run and report a sensitivity power analysis in the Stage 2 manuscript which shows that, although our sample is low, we are able to detect effect size of  $np^2 = .10$  for interactions, which is crucial for addressing our research questions (see below). Therefore, while our sample size is lower than planned (owing primarily too stringent and robust attention checks to ensure high data quality), we remain well powered to address our research questions.

*“Therefore, our final sample comprised 89 participants ( $M_{age} = 21.84$ ,  $SD = 3.457$ , 77.5% female,  $n = 60$  White British) with 52 students confirming they had preregistered their dissertation (preregistration group) and 37 who did not preregister (control group). Based on the lowest cell size ( $n = 37$ ), sensitivity power analyses indicate that we could reliably detect an effect size of  $np^2 = .10$  for the Group\*Time interaction and pairwise comparisons of  $d = /> .66$  with 80% statistical power, which was higher than planned’.*

Furthermore, at the Stage 1 manuscript, we discussed in depth the risks of sample size and this was explicitly factored into the planning of the study and discussed with reviewers (e.g., see text from the manuscript below).

*“The first risk was participant attrition from Time 1 to Time 2, leading to incomplete data across measures. We aimed to mitigate this by accounting for average attrition rates in our planned sample as per other longitudinal studies conducted on Prolific (7%-24%; Palan & Schitter, 2018) and utilising a varied recruitment approach. At Time 2, participants not recruited via Prolific were entered into a prize draw in order to incentivise participation. Similarly, recruitment of the preregistration group required a level of buy-in from institutions that embed a preregistration model into their undergraduate dissertation process. Members of the research team had contacts with these institutions listed in Table 1, which should mitigate barriers to student access in the preregistration group. We ran a sensitivity power analyses on the complete data and used this to contextualise our discussions and interpretation of final results. Our final sample size is smaller than planned, largely owing to our stringent attention checks and matching of data from Time 1 to Time 2; we discuss this in the Limitations.”*

To your point about helping the reader to understand this point, we also discuss sample size explicitly in detail in the Discussion section under ‘Limitations’ and here we reiterate the sensitivity analysis also:

*“We must acknowledge certain limitations of the present study. First, our sample size was smaller than we initially planned, owing largely to attrition from Time 1 to Time 2 of the survey, as well as the implementation of rigorous data quality checks. This meant that instead*

*of being able to detect effect sizes of approximately  $d = .40$  for the pairwise comparisons of interest, we were able to detect effect sizes of  $d \geq .66$  with 80% power. This means that we were only able to detect stronger effects rather than moderate effects, of which none were found. Therefore, it is possible that null results reported here were owing to an inability for us to detect significant effects with our smaller than planned sample size, rather than the absence of a true effect. Therefore, future research should aim to conceptually replicate our findings with larger sample sizes that are better equipped to detect smaller effect sizes. The issue of sample size is a challenge inherent within all quasi-experimental and longitudinal research, and we implemented multiple approaches to mitigate this, such as close contact with study participants through their supervisors, and follow-up emails to participate (see Recruitment). Therefore, we call now to other pedagogical scholars to take these reported findings as one early investigation into the impact of preregistration and urge the discipline to continue to provide high-quality, rigorous, nationally-representative data to shine empirical light onto Open Science tools and their value. That is, current findings should be regarded as a useful first step in the exploration of preregistration and its pedagogic value and we call on other researchers to shine further empirical light onto Open Science tools within education”.*

We feel strongly that this change in sample size does not constitute a deviation from the accepted Stage 1 protocol, because of how explicit we were in incorporating risk of attrition at Stage 1 and how we detailed extensively our plans to mitigate this risk and evaluate the strength of our final sample. Importantly, our sensitivity power analysis tells us that the high-quality data we did collect was sufficient to detect effect sizes of interest for interactions (which are a key part of our research question) and thus the findings remain scientifically and theoretically important. Thus, given a) how sample size risk was discussed in depth at Stage 1 and Stage 2, and the limits of our sample discussed explicitly in our limitations section, and b) the novelty of this research question and the ambitious, longitudinal study design that inherently comes with risk of data attrition, we are confident in the scientific soundness of our findings and this change in sample size does not constitute a deviation from the Stage 1 protocol. No changes to the introduction, methods, or analysis plan have been made and potential drop-out of participants was noted explicitly in the Stage 1 manuscript which received IPA.

3. Additionally, I checked for your data and code and was able to find the data sheet (<https://osf.io/download/zdu8f/>), but I was only able to find the code for the Stage 1 power analysis (<https://osf.io/download/jpmbt/>) and not the code for the Stage 2 Results section. I

checked your submission checklist and you state that all analysis code has been provided and is at the following URL: [https://osf.io/5qshg/?view\\_only=](https://osf.io/5qshg/?view_only=). However, I was not able to find the code at this repository. Please provide the remaining code and a direct link to the file at OSF (rather than a general link to the broader project).

***Response:*** Thank you for checking this! We have now provided a link to the remaining code for the core confirmatory analyses in the OSF project page that we link to in the manuscript.