

Dear Editors Logan and Chambers,

We resubmit a revision of the following manuscript for consideration at PCI Registered Reports:

Evaluating the pedagogical effectiveness of study preregistration in the undergraduate dissertation: A Registered Report

As you will see from our “Response to Reviewers” below, we have addressed each reviewer’s comments and implemented their suggestions which has improved our manuscript considerably. We would like to take this time to thank the reviewers for their positive and detailed feedback; it is the most constructive and collegial that we have received.

We hope you share our enthusiasm for the revised manuscript and look forward to hearing from you in due course.

Sincerely,

Madeleine Pownall, Charlotte R. Pennington, Emma Norris and Kait Clark

Response to Reviewers

Reviewer 1: Reviewed by Kelsey McCune

1. This preregistration aims to test the effect on statistics confidence and awareness of questionable research practices when undergraduate dissertations are preregistered. I think this is a necessary empirical investigation that will facilitate understanding and methods of the open science movement. The introduction section is thorough, with sufficient background to support the proposed study.

Response: Thank you for your positive appraisal of our paper. We have addressed all of your comments in the PDF file and marked these changes in red font in the manuscript, specifically we have:

- identified and resolved typos
- Np^2 has been defined (partial eta squared)
- Grade categories have been made clearer (p. 15):

This will be scored on a categorical scale **that is in line with the UK conventions of academic grades awarding: 1st class classification (> 70%), 2:1 classification (60 - 69%), 2:2 classification (50 - 59%), 3rd class classification 40 - 49%, fail (< 40%).**

- The table columns have been left in the original format, as this is a requirement of PCI Registered Reports
- Made hypotheses more explicit
- Added a new section titled ‘implications for open science’ in the discussion

2. I suggest adding an additional paragraph commenting on the impact to the open science movement of the findings of the proposed study.

Response: We now add a 'Implications for open science' into our discussion section (p. 29, paragraph 4 - 30) that responds to this comment:

3. The hypotheses, predictions, and interpretation of alternative results could be more detailed and I make specific comments on these in the attached track-changes pdf document. The planned analyses seem sufficient to address the 3 research questions, however I do not have experience in the best analysis practices for survey-based studies.

Response: We have now made our predictions and hypotheses much more detailed in the revised manuscript (p. 10).

We have three confirmatory hypotheses based on a significant two-way interaction between Group and Time. For all of the hypotheses, we predict a significant Time*Group interaction, in that participants in the preregistration group will show improvements above and beyond those that occur due to time differences (Time 1 vs Time 2).

H1: Due to the thoughtful engagement with statistical processes that the preregistration process requires (Lindsay et al., 2016), we predict that students who preregister their dissertation will have higher scores on the four constructs within the Survey of Attitudes Toward Statistics (SATS-28), from Time 1 to Time 2.

H1a. Students who preregister their dissertation will have higher (i.e., more positive) *affect* towards statistics compared to students who do not preregister their dissertation from Time 1 to Time 2.

H1b. Students who preregister their dissertation will have higher self-reported *competence* with statistics compared to students who do not preregister their dissertation from Time 1 to Time 2.

H1c. Students who preregister their dissertation will have higher perceived *value* of statistics compared to students who do not preregister their dissertation from Time 1 to Time 2

H1d. Students who preregister their dissertation will have higher and less *difficulty* with statistics at T2 compared to students who do not preregister their dissertation from Time 1 to Time 2.

H2: Secondly, given that the preregistration process prompts wider consideration of the QRPs that preregistration aims to avoid, we predict that students who preregister their undergraduate dissertations will have a reduced self-reported acceptance of 11 selected QRPs compared with students who do not preregister their dissertation, when comparing Time 1 responses with Time 2.

H3: Relatedly, given that the preregistration process forms part of a wider conversation about open and transparent science, we expect that students who preregister their undergraduate dissertations will have higher perceived confidence in their understanding of 12 selected Open Science terminology terms, compared with students who do not preregister their dissertation, when comparing Time 1 responses with Time 2.

Reviewer 2: Reviewed by Neil Lewis, Jr.

“Evaluating the pedagogical effectiveness of study preregistration in the undergraduate dissertation: A Registered Report” is an interesting manuscript that proposes an innovative design to study the effects of implementing preregistration into research methods pedagogy on student learning and attitudes. Overall, I think the manuscript is well-written and informative, asks an important question for both research and practice, and has a sound research design for testing the proposed hypotheses. The authors addressed most of the concerns that came to mind as I read the manuscript, so I only have a few minor questions and suggestions for the authors to consider prior to carrying out their study. [...] Overall, this is a great manuscript, and I commend the authors for conducting this important research.

1. My first question is about why there is such a heavy focus on statistics as an outcome rather than other elements of research design. One of the things that I appreciate about preregistration is that it forces researchers (and students) to think about the relationship between many elements in the research process (strength of manipulation, effect sizes, sample characteristics that might moderate processes, etc.). The statistical elements are of course important, but the preregistration process reveals more than that—it forces us to wrestle with how all elements of

research design are related. I was surprised to not see other research design related competencies (other than statistics and QRPs) being measured.

Response: Thank you for your positive review and detailed comments on our Stage 1 manuscript. Our focus on statistics, rather than other areas of the research process, is largely informed by the vast literature which demonstrates how statistics anxiety or concerns around the statistical content of dissertations is a concern among students and a barrier to engagement (e.g., Onwuegbuzie & Wilson, 2003). Further, we elected to select outcomes that would improve the student *experience* during the dissertation process. As we see it, there are two core reasons why preregistration would be useful for students: 1) it leads to outcomes that are good for science, and 2) it improves students' engagement with (and even enjoyment in) the scientific process. Currently, our study focuses on the latter of these two goals, although we welcome follow-up research which interrogates further how reregistration in undergraduate dissertations may be useful for the wider scientific research issues, as you mention. Moreover, the scales that we use in our study are taken from other published literature and we are thus confident in their validity and reliability. There is currently, to our knowledge, no validated measure of students more general views of research methods, which also justifies our choice of materials. We now have a new section in our Discussion where we note this explicitly and already begin to suggest follow-up areas that may be worthy of study in this context too, as you note (p. 30 of the revised manuscript):

Preregistration may also have benefits beyond those that are captured in the measures of the present study. For example, engagement in the preregistration process may likely improve students' trust in the research they are conducting, inspire ambitions to pursue a career in research, and improve research literacy above and beyond attitudes towards statistics. These potential variables are all worthy of follow-up studies that further interrogate how preregistration, and indeed Open Science tools more broadly, may confer advantages to undergraduate students.

2. My second question is about the "forced entry" strategy for minimizing missing data...is that allowed? This could be a country-level difference so feel free to ignore this question if it is, but in the US we are not allowed to use forced responses—it is considered an ethical violation; we can use "request response" to nudge them toward answering, but participants have to have the option to not answer questions that they do not want to.

Response: While the Ethics Committee has approved the use of 'forced entry', we agree that there are ethical concerns about forced responses, particularly for paid studies. Further, there are also some concerns about forced entries leading to lower quality data, because participants randomly select items in order to progress in the survey (e.g., evidence from Décieux et al., 2015). Therefore, we now use 'requested entry' via Qualtrics, where

participants must confirm that they are happy that they have answered all the questions they wish to, if some are left unanswered. We detail this now on page 21:

Finally, we will avoid missing data adversely impacting our statistical power by using a ‘**requested entry**’ option on Qualtrics, so participants are unable to progress in the survey **without first confirming that they are happy that they have answered all the questions they wish to (if some are left unanswered).**

Reviewer 3: Reviewed by Lisa Spitzer

I want to congratulate the authors to this interesting and thoroughly planned Registered Report, which I enjoyed reading and reviewing very much. Summary: The authors want to conduct a study among undergraduate psychology students in the UK, to assess if preregistration of the final-year dissertation influences attitudes towards statistics, QRPs, and open science. For this, a targeted sample size of 200 students will be recruited, that plan or plan not to preregister their dissertation. The design follows a 2 (pregistration: yes vs. no) between x 2 (timepoint: before and after dissertation) within subjects design.

I have added comments to the PDF of the Registered Report (see attached). Most of these concern minor points, such as:

- 1) Spelling.
- 2) Structure of the text.
- 3) Consistency of notation and formatting.

Depth and clarity of descriptions. Most often I added some questions I had during reading, for which I would find it beneficial if the authors would answer them in the text to enhance clarity and reproducibility of the study.

Response: Thank you for your useful and detailed comments which have helped us to greatly improve this manuscript. We have now responded to all of your comments and questions provided on the PDF, and have marked all changes in red in-text. Most notably, we have now restructured the introduction, as you suggested, to read: Open Science background -> preregistration and the value of preregistration -> Open science in undergraduate training. We have also amended all “open science” to “Open Science”. We now address your additional comments:

1. Enumeration of hypotheses. I would find it beneficial if the authors would differentiate between different sub-hypotheses and enumerate them, to reference them later during the analysis section.

Response: We have now made the hypothesis and sub-hypotheses clearer on pages 10-11, as per your suggestion. See below:

We have three confirmatory hypotheses based on a significant two-way interaction between Group and Time. For all of the hypotheses, we predict a significant Time*Group *interaction*, in that participants in the preregistration group will show improvements above and beyond those that occur due to time differences (Time 1 vs Time 2).

H1: Due to the thoughtful engagement with statistical processes that the preregistration process requires (Lindsay et al., 2016), we predict that students who preregister their dissertation will have higher scores on the four constructs within the Survey of Attitudes Toward Statistics (SATS-28), from Time 1 to Time 2.

H1a. Students who preregister their dissertation will have higher (i.e., more positive) *affect* towards statistics compared to students who do not preregister their dissertation from Time 1 to Time 2.

H1b. Students who preregister their dissertation will have higher self-reported *competence* with statistics compared to students who do not preregister their dissertation from Time 1 to Time 2.

H1c. Students who preregister their dissertation will have higher perceived *value* of statistics compared to students who do not preregister their dissertation from Time 1 to Time 2.

H1d. Students who preregister their dissertation will have higher and less *difficulty* with statistics at T2 compared to students who do not preregister their dissertation from Time 1 to Time 2.

H2: Secondly, given that the preregistration process prompts wider consideration of the QRPs that preregistration aims to avoid, we predict that students who preregister their undergraduate dissertations will have a reduced self-reported acceptance of 11 selected QRPs compared with students who do not preregister their dissertation, when comparing Time 1 responses with Time 2.

H3: Relatedly, given that the preregistration process forms part of a wider conversation about open and transparent science, we expect that students who preregister their undergraduate dissertations will have higher perceived confidence in their understanding of 12 selected Open Science terminology terms, compared with students who do not preregister their dissertation, when comparing Time 1 responses with Time 2.

2. Bot protection. Since the authors plan to advertise their study via social media, I recommend to implement some kind of bot protection.

Response: Thank you for raising this important point. We have looked into best practice guidance on this issue and will use 'Prevent multiple submissions' via Qualtrics security settings to reduce the likelihood of fraudulent/bot responses. We will also employ the 'prevent indexing' security option, which means that the study's URL will not be picked up by search engines that may be targeted by bots (see <https://www.qualtrics.com/support/survey-platform/survey-module/survey-checker/fraud-detection/>). We note this in the revised manuscript (p. 17):

We will also employ Qualtrics' 'prevent multiple submissions' and 'prevent indexing' (i.e., block search engines from including the study URL in search results) security options to minimise chances of fraud/bot responses.

3. Definition of preregistration in the study. I recommend that the authors update their definition of preregistration, since the current definition describes that preregistration is possible "before you collected your data". However, to address secondary data analysis preregistration, I would add that preregistration is also possible before data is analyzed.

Response: We have amended the wording in the study materials to refer to pre-registration as: "before you look at the data" and "before you have accessed your data" to respond to this comment. We include this in the manuscript (p. 15)

"Preregistering a research project involves creating a record of your study plans before you look at the data. The plan is date-stamped and uneditable. The main purpose of preregistration is to make clear which hypotheses and analyses were decided on before you

have *accessed* your data and which were more exploratory and driven by the data.”

4. Attention check. I recommend to implement a different attention check, since participants that simply agree to anything in the study, will pass the current attention check.

Response: To prevent the likelihood of straight-lining impacting the attention check (i.e., participants answering ‘agree’ to all items including the check), we now have ‘strongly disagree’ as the attention item. We have also consulted with the open science community for advice on appropriate attention checks online and have added another attention check, in which participants must write a word (“purple”) into a textbox (see end of page 15-16 for detail in the manuscript). This attention check is endorsed by the Prolific guidelines as a suitable and fair way of spotting for low quality data: <https://researcher-help.prolific.co/hc/en-gb/articles/360009223553-Using-attention-checks-as-a-measure-of-data-quality>. We detail this in the revised manuscript (p. 17):

As an attention check (i.e., to ensure that participants are actively paying attention to the survey materials and to prevent spam/bot respondents), we will add an item “*Please select strongly disagree to this question*” in the COM-B measure, to assure data quality. This will be repeated in Time 1 and Time 2. As a second attention check, we will use a protocol from the Prolific guidelines and will ask participants “*Please enter the word ‘purple’ in the textbox below.*” accompanied by a textbox. Any participant who fails both of these attention check (i.e., who does not select strongly agree and correctly enter the word ‘purple’) will be excluded from the final analyses.

5. Addition of some questions in the study. This is only based on personal interest.

Response: We agree that a question asking about whether students felt pressure to find significant results would be very interesting; we have added this item into the measures (see p. 19).

One item will also be “*I felt pressure from my supervisor to find significant results in my dissertation*” (reverse scored).

6. Combination of frequentist and Bayes statistics. Unfortunately, I am no expert to Bayes statistics, but I understand that one method (frequentist vs. Bayes) would be

sufficient, however, here both are combined. Maybe the other reviewers can contribute additional input to this question.

Response: At present, frequentist statistics still tend to prevail within the psychological literature, and thus we wish to include p-values in order to report our findings to be easily understood by a wide audience. However, given the problems with the interpretation of p-values (e.g., Wagenmakers, 2007), we also wish to report Bayes factors as indicators of the strength of evidence for our against our hypotheses. Bayes factors will also be necessary to evaluate the strength of evidence against our hypotheses in the event of null results and are recommended to be reported alongside p-values for this reason (Dienes, 2016).

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic bulletin & review*, 14(5), 779-804.

7. One hypothesis concerns the "understanding/awareness" of open science. Since understanding is not directly tested, but rather perceived understanding is assessed, I would recommend to use the term "perceived understanding".

Response: We have now changed this terminology throughout to use the term 'perceived understanding'.

8. Sometimes, an ANOVA is implemented for a comparison of two groups (preregistration: yes vs. no). From my perspective, a t-test would be the more appropriate analysis. (more detailed descriptions of these comments are provided in the attached PDF)

Response: We have amended this for the COM-B items, where there is only one score taken at Time 1 - thank you for the suggestion.

Additionally, I would like to discuss some major points which I think are important to consider before conducting the study:

9. I would like to discuss more the data collection at Time 1. First, I suggest that quota sampling could be used to control for different group sizes (preregistration: yes vs. no). Alternatively, if no quota sampling is used, I think the authors should discuss more what would be the consequences of the case that there may be very unequal group sizes.

Response: We will use Cross Logic Quota sampling via Qualtrics - thank you for this suggestion. Please see page 13 for detail of this quota sampling in the revised manuscript:

We will use 'Cross Logic Quota' sampling within Qualtrics (see Qualtrics, <https://www.qualtrics.com/support/survey-platform/survey-module/survey-tools/quotas/>) to roughly monitor group allocation at Time 1, although this will be done using the preregistration *plan* questions (see below), which could differ from the final preregistration group allocation

at Time 2 (i.e., some participants could plan to preregister but do not actually preregister at Time 2).

10. Also, in general, I am unsure if enough participants can be collected at Time 1, since only a short time period is indicated for this data collection (September - October 2021). I would appreciate it if the authors could elaborate on how they will proceed if less than the targeted 240 students can be collected by the end of October, or if only a small proportion of participants indicate planning to preregister.
(more detailed descriptions of these comments are provided in the attached PDF)

Response: If we fail to reach our required N of 200 participants and/or have unequal groups, as per your previous comment, we will first conduct a sensitivity power analysis on the available data and use this to contextualise our discussions and interpretation of final results (see p. 12). Whilst this is one of the challenges of the Registered Report publication model ([Chambers and Tzavella \(2021\)](#)), we have mitigated against this risk as much as feasibly possible. We are thus confident that we can successfully recruit 200 participants (after accounting for failed attention checks and attrition) due to: incentives for participation, our contacts with departments who operate a preregistration model in their dissertations, and our connections with other Open Science communities who will be able to assist in recruiting participants in the preregistration group (e.g., communities such as UKRN, ReproducibiliTea etc).

11. Overall, I think that the study addresses a valid research question with coherent and plausible hypotheses. I perceived the described methods as sound, feasible, and for the most parts, clear. Whenever open questions remain, I have marked them in the attached PDF. Additionally, I thought it was excellent that the authors had already addressed possible limitations themselves. I feel like the study will be an important contribution and I hope that my comments will help the authors improve their study and manuscript.

Response: Thank you for your detailed and constructive feedback.

Reviewer 4: anonymous reviewer,

1. This proposed study aims to test the influence of study pre-registration on undergraduate attitudes towards statistics and questionable research practices. Given the enthusiasm for open science practices among early career researchers, and the certainty with which many of those practices are endorsed by some members of the field, it is important to empirically test the influence of those practices on meaningful research outcomes. Open science practices themselves should indeed be subject to the scrutiny of the Registered Report format, so I commend the team for taking on this sort of research. In order to meet the demands of this format, however, I believe the current report could use more detail in several key areas related to establishing the scientific premise, justifying the choice of measures and associated hypotheses, and ensuring that any results can be interpreted with confidence.

Response: Thank you for these very helpful comments. We go through each one in turn below:

2. Overall, I think the study motivation and scientific premise should be strengthened. While undergraduate teaching and learning are inherently important, it's a bit unclear exactly what issue this RR is tackling, or what guides the choice of measures and hypotheses. For instance, the intro lays out a number of dissertation struggles that undergraduates may face related to anxiety, disengagement, writing ability, and supervisory relationships (p. 6). Is pre-registration meant to mitigate these concerns? If so, how? It's unclear to me whether the focus is on student well-being, research quality, or pedagogical value. If it is all of these things, I think the intro should more clearly establish why each is important to study in this context, and make a more explicit case for why/how pre-registration should be expected to improve these problem areas. Is the thinking that the mere act of preregistration will impart these benefits? Or will these students undergo some additional learning, which might have the same benefits without the actual preregistration? The intro also implies that open science practices are universally endorsed, but there is in fact a vocal pushback against many of these practices (e.g., Szollosi et al., 2020, TICS). I mention that, not to undermine the current study, but because I believe this perspective should be recognized and may offer further fuel to justify the proposed work.

Response: We now describe our thinking in more explicit detail on page 10 in our hypotheses. Essentially, we suggest that preregistration may improve outcomes such as positive statistical affect and competence because of the level of engagement and thinking that preregistration prompts researchers (in this case, students) to consider at the start of the research process. For each of our hypotheses, we now also add in a rationale for the predicted direction of our results (p. 10):

We have three confirmatory hypotheses based on a significant two-way interaction between Group and Time. *For all of the hypotheses, we predict a significant Time*Group interaction, in that participants in the preregistration group will show improvements above and beyond those that occur due to time differences (Time 1 vs Time 2).*

H1: *Due to the thoughtful engagement with statistical processes that the preregistration process requires (Lindsay et al., 2016), we predict that students who preregister their dissertation will have higher scores on the four constructs within the Survey of Attitudes Toward Statistics (SATS-28), from Time 1 to Time 2.*

H1a. Students who preregister their dissertation will have higher (i.e., more positive) *affect* towards statistics compared to students who do not preregister their dissertation from Time 1 to Time 2.

H1b. Students who preregister their dissertation will have higher self-reported *competence* with statistics compared to students who do not preregister their dissertation from Time 1 to Time 2.

H1c. Students who preregister their dissertation will have higher perceived *value* of statistics compared to students who do not preregister their dissertation from Time 1 to Time 2.

H1d. Students who preregister their dissertation will have higher and less *difficulty* with statistics at T2 compared to students who do not preregister their dissertation from Time 1 to Time 2.

H2: Secondly, given that the preregistration process prompts wider consideration of the QRPs that preregistration aims to avoid, we predict that students who preregister their undergraduate dissertations will have a reduced self-reported acceptance of 11 selected QRPs compared with students who do not preregister their dissertation, when comparing Time 1 responses with Time 2.

H3: Relatedly, given that the preregistration process forms part of a wider conversation about open and transparent science, we expect that students who preregister their undergraduate dissertations will have higher perceived confidence in their understanding of 12 selected Open Science terminology terms, compared with students who do not preregister their dissertation, when comparing Time 1 responses with Time 2.

We also note the push-back to preregistration in our introduction (p. 4).

Despite these practices being increasingly endorsed and embraced by the scientific community (however, see Szollosi et al., 2019 for an alternative perspective), scant research assesses the pedagogic value of Open Science practices in improving teaching and learning.

3. After establishing the problem, I also think the report could clarify why these particular proposed measures were chosen, and also offer stronger justification that the measures will test what we think they'll test. It's not totally clear to me that these self-report questionnaires actually tap into the pedagogical effectiveness of pre-registration. For instance, is there any evidence that student attitudes toward statistics relate to their knowledge or competence with statistics?

Response: Given the novelty of this study and the notion that this is the first study, to our knowledge, that empirically assesses pedagogical effectiveness of preregistration, our measures were chosen predominantly due to recent theoretical commentaries that suggest how preregistration may confer such advantages to students (e.g., Creaven et al., 2021; Pownall, 2020). Whilst much literature has made the moral *argument* that the thoughtful nature of the process of preregistration may benefit student's statistical attitudes, acceptance of QRPs, and understanding of Open Science, this has not yet been addressed by empirical research, which is the main rationale for these measures.

4. Instead of asking for confidence in open science terms, why not actually measure ability to define the terms accurately? I think the report could be strengthened overall if it included more objective measures of knowledge (stats and open science), and engagement in QRPs (i.e., did you do any of these things?), rather than just attitudes.

Response: The reason why we did not want to measure this is because we did not want to make judgments on 'correctness' of students' open science definitions. Indeed, the metascience literature demonstrates that many open science tools and terms have contested definitions (for example, to use a crude example: does 'open data' refer to useable, accessible data that are publicly available to all? Does data that are available only to academics count as 'open data'? and so on). Therefore, we see our way of measuring awareness of terms as a more robust, reproducible method of capturing this outcome. Similarly, use of this measure also reflects previous literature that asks researchers about their perception of open science terms, rather than their ability to 'correctly' define them (e.g., Krishna & Peter, 2018; Stürmer et al., 2017).

Stürmer, S., Oeberst, A., Trötschel, R., & Decker, O. (2017). Early-career researchers' perceptions of the prevalence of questionable research practices, potential causes, and open science. *Social Psychology*.

5. I would also like to see more details about recruitment and study criteria. What do we know right now about the uptake of pre-registration in this cohort? As in, how feasible will it be to recruit equal numbers in both groups, and for those groups to be

matched? The report acknowledges this might be an issue, but I think it would be a rather serious one if all the pre-registration participants end up coming from a certain set of targeted schools, and all the non-pre-registered are from different schools.

Response: We appreciate that recruitment of the pre-registration group is one of the core challenges in this study's proposed methodology. However, as we outline in the manuscript, we have identified a range of schools that already implement pre-registration processes. Beyond these, we plan to recruit participants via social media (e.g., on Academic Twitter). We will also use Cross Logic Quota sampling via Qualtrics to ensure balanced groups, as we now note in p. 13:

We will use 'Cross Logic Quota' sampling within Qualtrics (see Qualtrics, <https://www.qualtrics.com/support/survey-platform/survey-module/survey-tools/quotas/>) to roughly monitor group allocation at Time 1, although this will be done using the preregistration *plan* questions (see below), which could differ from the final preregistration group allocation at Time 2 (i.e., some participants could plan to preregister but do not actually preregister at Time 2).

The concern of all pre-registration participants deriving from the same schools is indeed a possibility, but we do not necessarily share the concern that this is a serious methodological issue. Many of the schools that we have already identified as adopting preregistration in undergraduate dissertations are from a range of types of institutions (including post-92 universities and Russell Group). Because this study is the first, to our knowledge, that assesses the pedagogical effectiveness of one open science practice, pre-registration, there are indeed likely to be other open science practices, teaching, and attitudes that coexist with pre-registration practices. It may be argued that departments that are flexible enough to offer pre-registration as a route for students also have more open attitudes towards implementation of other open science practices and teachings. However, because we have Time 1 measures as the baseline, we will be able to assess any differences between the groups *prior* to students undertaking their preregistration journey, which mitigates this concern. We hope this satisfies that concern, and we will be sure to include a more comprehensive discussion of this in the Stage 2 manuscript's discussion section.

6. Is the **only** inclusion criteria that participants be a final-year undergraduate student, studying Psychology at a U.K institution. For instance, what happens at Time 1 if they respond that they've already pre-registered their study? Is the only possible reason that data would be excluded if a participant fails the attention check? I'm just asking to ensure that all the methods are specified in enough detail to be truly reproducible and transparent.

Response: At Time 1, participants are UK-based final-year undergraduate students who are planning an empirical quantitative undergraduate dissertation who have not already pre-registered their study. We have ensured that data collection for Time 1 starts at the end of September/beginning of December to correspond with the timeline of the UK undergraduate dissertation process (e.g., supervisors are allocated their students in September, so it is highly unlikely that they will have preregistered their study by then - if they state that they have, they will not be eligible to complete the study). We have added details of the latter of these criteria more explicitly to page 11-12:

To be eligible for inclusion, participants are required to confirm that they are a final-year undergraduate student, studying Psychology at a UK institution **and planning an empirical quantitative undergraduate dissertation. Participants must have not already pre-registered their proposed undergraduate study at Time 1, and must confirm this in the beginning of the study.**

This is to ensure that the study can contribute directly to existing pedagogic policy discussions regarding embedding Open Sciences within the undergraduate dissertation (e.g. the British Psychological Society's course accreditation standards, 2019). **To be eligible to participate at Time 2, participants must have completed Time 1 (and have a corresponding participant ID number to match up responses).**

7. Regarding the sample size, I understand that the proposal is based on time and resource considerations, but we would still need to be reasonably assured that the outcomes would be valuable. What was the input to the power analyses, and do we have reason to believe that the cited effect size (80% statistical power to detect an effect size of $np2 = .04$) would constitute a meaningful effect in this context?

Response: As the Reviewer states, our rationale for this sample size is based upon both time and monetary resource limitations - we have enough funding for 200 participants ($n = 100$ per condition) completing the survey at two timepoints. This is a reasonable justification as per Lakens (2021). To assess whether we had sufficient power (80%), we inputted this target sample size within the superpower shiny package (Lakens & Caldwell, 2021) and iteratively changed the standardised mean difference (SMD; e.g. Cohen's $d = .30$, $d = .40$, $d = .50$ with SD set at 1.00) until it yielded sufficient power for our focal pairwise comparison - that is, the difference between preregistration vs. control at Time 2. We have now also included the input on our OSF page, as well as the output (subfolder "power analyses": <https://osf.io/5qshg/>). After collecting the data, we will perform sensitivity power analyses on the final sample size to assess the effect size that we were able to detect given 80% power and alpha at .05. We will interpret our findings in line with this effect size to ensure that we do not go beyond the data. Importantly, we will also conduct Bayesian analyses using the default JZS prior ($r = 0.707$; Rouder et al., 2009) to evaluate strength of evidence.

The JZS prior is a noninformative default and objective prior designed to minimise assumptions about the expected effect size, hence being appropriate for this study. Finally, at this point we have no prior knowledge of the effect size(s) associated with this research (as this is the first research study to investigate the pedagogical effectiveness of preregistration), but we do know that the median effect size within effect size within social psychological research is typically moderate: see Lovakov & Agadullina, 2021), which we aim to have power to detect. A moderate effect would be meaningful evidence to suggest that preregistration has pedagogical value and the effect sizes found will also help researchers to plan appropriate sample sizes in future studies on this topic.

References:

Lakens, D., & Caldwell, A. R. (2021). Simulation-Based Power Analysis for Factorial Analysis of Variance Designs. *Advances in Methods and Practices in Psychological Science*, 4(1), 2515245920951503

Lovakov, A., & Agadullina, E. R. (2021). Empirically derived guidelines for effect size interpretation in social psychology. *European Journal of Social Psychology*.

<https://doi.org/10.1002/ejsp.2752>

Lakens, D. (2021). Sample size justification.

<https://psyarxiv.com/9d3yf/download?format=pdf>

8. I very much appreciate that hypotheses and tests are specified in a table, but I think the interpretation, alternative outcomes, and theory portions of the table should be more explicit. For instance, the final column applies broadly to all hypotheses and tests, stating “This may call into question various situational, contextual, and personal factors that impact how preregistration may be useful to students in this context “ Ideally, this section would include straightforward, concrete predictions that are guided by theory, for each hypothesis, and specific implications if findings do or don’t match expectations.

Response: For clarity of this section, we now remove the sentence that you have flagged and end concretely and clearly with: “Theoretically, the notion that preregistration confers a tangible, pedagogical benefit to students in their dissertation process could be (un)supported by all of our proposed analyses.” We have also added: “Explanations for all results will be presented in the discussion.”

9. The “Limitations” section acknowledges that pre-registrations may differ in their rigor, but states that this issue is beyond the scope of the current study. This strikes me as an incredibly consequential issue, if we truly want to understand the impact of pre-registration on student outcomes. I understand why a common, plain-language definition of preregistration is provided, but I think there needs to be some further confirmation that the participants actually conducted a valid preregistration. For instance, is it acceptable if students just saved a personal pdf with their “plan,” or do they need to have submitted it to an accepted repository? There is then WIDE variability in the detail with which a pre-registration is specified. OSF alone provides 8 possible options for preregistration. There is then also WIDE variability in adherence to the registered methods/analyses. What if a large number of students

submit a bare minimum, superficial pre-registration, just because it's required of their program? Many scientists have speculated that this may become the norm, and I worry that it could set up this study up for failure to find any differences between groups. Why not ask whether the participants actually followed through with the registration, or to what extent they deviated?

Response: This is an important point that we have grappled with throughout the design of this study. While we cannot 'check' participants' preregistration, and thus much of this relies on a) self-report and b) appropriate supervision for students in completing useful preregistration, we now have included a question in the materials whereby students must self-report the extent to which they followed their preregistration plan (1 = not at all, 3 = entirely). We will use this data descriptively, rather than inferentially, and will use it to guide our discussion and propose follow-up areas of research in this area. We have also added more detail to this point in the revised manuscript (p. 31)

Indeed, there is emerging literature to suggest that the specificity of preregistrations differs between researchers (Bakker et al., 2020). However, it is beyond the scope of this research to assess each preregistration for quality and rigour. Similarly, adherence to preregistration protocols is another indicator of preregistration value (i.e., if researchers do not strictly adhere to their analysis plan, it may not be useful in reducing QRPs or, in our context, improving statistics attitudes). XX (XX%) of participants in our sample indicated that they did not follow their preregistration plan in their dissertation, which suggests that more research is needed into the implementation of preregistration in a pedagogical context. Practical reasons for this may also be informed by our qualitative data here, which reports perceived (dis)advantages to preregistration. Therefore, future work, depending on our findings, may wish to establish the extent to which preregistration quality impacts on the core outcomes of interest in this work.

10. In the end, if you find no differences between groups with these methods, would you confidently declare that preregistration has no benefits? If not, why not? If there is a difference, would we know it is due to the act of pre-registering, or could it be explained by some other factor? As in, do students in a pre-registration context just get more exposure to open science issues as part of their process, and could the same benefits therefore be achieved by briefly teaching students about QRPs and Open Science? The above concerns all boil down to ensuring that any result is meaningful (null or otherwise), which should be a primary goal of the design and RR

proposal. I know some of these concerns may seem persnickety, but they are aimed at the overarching goals of ensuring that any results can be confidently interpreted, and that the methods are precisely reproducible.

Response: Thank you for asking this important question. First of all, we would like to reiterate that we will evaluate the strength of evidence for our findings with Bayesian analyses to ensure our study is informative. As we see it, even if there are certain contextual confounds that mean participants in the preregistration group receive more open science training than the control group, this is not necessarily problematic in of itself. Preregistration does not happen in a vacuum, and therefore we expect that implementation of a preregistration model will likely be accompanied by more thorough coverage of other open science tools, practices, and behaviours (for example, a supervisor is likely to explain to students *why* preregistration is required given current concerns about the replication crisis/credibility revolution). However, given our emphasis on difference from time 1 to time 2, and how we collect data on supervisor attitudes, prior understanding about open science, and statistics attainment, we are confident that these results will directly inform the conversation about whether preregistration confers benefits to students. Finally, if we do not find differences between the groups for our planned hypotheses, we would be interested in following this up with some exploratory analyses to determine whether there may have been differences within the groups that predicted the differences in outcomes. This, coupled with our exploratory qualitative questions about perceived (dis)advantages of preregistration from a student perspective, would further inform the role of preregistration in conjunction with other factors.

Reviewer 5: Noémie Aubert Bonn

1. Thank you very much for sharing the Registered Report “Evaluating the pedagogical effectiveness of study preregistration in the undergraduate dissertation: A Registered Report” with me and for asking for feedback. The report presents a study to assess whether pre-registration of undergraduate Psychology dissertations may have benefits on students’ attitudes towards statistics and questionable research practices (QRP) and on their understanding of open science concepts. I read the registered report with great interest and believe that the project will help answer an interesting question which is still understudied at present. The report is clearly structured and well written, and my overall impressions are positive.

Despite my positive appraisal, there are a few points I noticed which I believe may help the authors strengthen their work. Three points are more important and I detail them further below. I then follow with minor points which I present in a list in the end of this review. Before starting with my comments, I would like to mention that I have not used Bayesian analyses and my statistical knowledge is unfortunately rather rusty, I therefore recommend that the authors do not use my absence of comments on the statistical analysis as a confirmation that the planned analysis is adequate.

The first major point I would like to raise concerns the research questions and the way in which the project is described. As I was reading through the project, I was

very confused by the comparison of the two groups, thinking to myself that the group who is likely to preregister their dissertation is also likely to study in a center where awareness to open science is markedly better. In fact, if the two groups were to be compared one to the other, I believe that the study would look at the influence of different institutions and research groups more than the influence of the pre-registration process itself. In an ideal scenario, randomly assigning a condition to each participant would avoid this problem, but I can understand that this may not be realistic, especially in a project on such a large scale. Nonetheless, even with the design kept as is, it is not until page 17 that I understood that the authors plan to compare the difference/interaction/progress between T1 and T2, thereby cancelling differences between prior knowledge and attitudes.

This distinction is very important and should be made clear throughout the report. For example, the hypotheses presented on page 9 fail to capture the distinction. The first hypothesis states that “Students who preregister their dissertation will have higher positive affect towards statistics, higher self-reported competence with statistics, higher perceived value of statistics, and less difficulty with statistics at T2 compared to students who do not preregister their dissertation.” In this phrasing, the hypothesis H1 appears to only compare both groups at T2, rather than to compare the groups on their improvement between T1 and T2. The same goes for H2, where the authors mention that “Students who preregister their undergraduate dissertations will have a reduced endorsement of QRPs compared with students who do not preregister their dissertation.” While it should state that they will show a greater reduction of their endorsement of QRPs between T1 and T2 than the control group. And so forth, also for H3. The authors should make sure that the manuscript clearly explains that the results will compare both groups the difference/improvement/interaction between T1 and T2 to avoid confusion and early criticism from readers.

Response: We have now reworded our hypotheses entirely to respond to this concern (see p. 10-11). Specifically, our hypotheses now make it clear that we are interested in the *interaction* between time and preregistration group. We detail this above in our response to Reviewer 4. That is, we are interested in how preregistration may impact students’ outcomes in ways *above and beyond* the kinds of differences that occur anyway after completing a dissertation (i.e, time differences from Time 1 to Time 2). While we appreciate that contextual factors may be factors in whether students preregister or not, we have mitigated risk of confounds as much as feasibly possible. Most notably, capturing data on students’ supervisor approach and statistics performance, to use as covariates, should mitigate some of these concerns. We will, however, discuss the need for follow-up studies that take into account contextual factors (e.g, institution, research group culture, future career goals etc) in our discussion.

As we see it, even if there are certain contextual confounds that mean participants in the preregistration group receive more open science training than the control group, this is not necessarily problematic in of itself. Preregistration does not happen in a vacuum, and therefore we expect that implementation of a preregistration model will likely be accompanied by more thorough coverage of other open science tools, practices, and

behaviours (for example, a supervisor is likely to explain to students *why* preregistration is required given current concerns about the replication crisis/credibility revolution). However, given our emphasis on difference from time 1 to time 2, and how we collect data on supervisor attitudes, prior understanding about open science, and statistics attainment, we are confident that these results will directly inform the conversation about whether preregistration confers benefits to students.

2. Also related to this point, not only is it possible that the groups differ from the baseline, it is also possible that, if the experimental group rates higher on the different variables at T1, they may show less improvement between T1 and T2 than a group who started at a lower point but is nonetheless in the most research-intensive learning-period of their degree. While I do not have a recommendation for this potential issue, I think that the authors should consider this possibility and maybe discuss whether the analysis, as it currently stands, may risk masking the full effect of preregistered reports by leaving less room for improvement overall in the experimental group. This is all hypothetical, but I thought that it should be addressed, or at least reflected upon before starting data collection. The authors should clearly explain the limitations of this non-random assignment in the 'Risk and mitigation'.

Response: This is a very valid concern. We agree that we cannot mitigate this risk, but we do now flag it explicitly in our 'Risks and Mitigation' section of the revised manuscript (see p. 21). If this is the case in our data, we could conduct exploratory analyses looking at the mean difference between the preregistration and non-preregistration group (e.g., by subtract Time 1 scores from Time 2), but this would be purely exploratory and in response to this risk.

Specifically, we expect that despite any differences between groups at Time 1, there will be a significant interaction indicating that **engaging with** the preregistration process has an *additive* effect on students' attitudes, behaviours, and perceptions of Open Science (*i.e.*, **it improves scores beyond improvement that occurs due to differences in time point**). It could also be feasible for ceiling effects to occur in the preregistration group at Time 1, particularly **given the aforementioned concern about contextual factors that impact students' knowledge of Open Science and QRPs. This could mean that differences from Time 1 to Time 2 are 'masked' due to high scores at Time 1 for the preregistration group. Whilst we cannot methodologically mitigate this concern, we will discuss it in detail following data collection and use this to guide interpretation of our results.**

3. A second important point concerns the COM-B measure and its role in answering the research question. From the report, I did not find it clear what the COM-B will contribute nor how it will be used in interpreting the data. The authors mention that the COM-B results will be used to compare groups, but I am unsure what this finding would mean about the preregistration process per se. If I understood properly, in most cases, the decision to preregister a dissertation project comes from the research laboratory or the supervisor. The COM-B could then provide information on how prepared and motivated those planning on pre-registering their study really feel, but I am unsure whether a group comparison is truly helpful. I also feel that, for the control group or those who are not planning on performing a preregistration, the COM-B questions will be highly abstract (in fact I found most COM-B questions very abstract, but I detail this point later on when discussing the 'Study materials'.

In this regard, I was surprised that the authors did not plan to use the COM-B to assess how prepared students feel about their final dissertation project (rather than about the preregistration), in which case they could do a T1-T2 comparison and see whether the preregistration helps improve the COM-B scores of the experimental group.

Response: Capability, opportunity and motivation are defined by the COM-B model as the core components required to enact any behaviour. Our inclusion of COM-B items to our study is an exploratory set of measures that tests whether students feel they have the appropriate capability, opportunity, and motivation to engage in the behaviour of preregistration. This will provide further information about potential barriers and facilitators of Open Science practices, which has been unexplored in this population.

Indeed, there is a growing emphasis on approaching open science behaviours through a lens of behaviour change in order to enact researcher change (e.g., Norris & O'Connor, 2019). All actors within the research process enact behaviours related to research, whether students or supervisors. In instances where students will be encouraged by their supervisors to preregister, this intuitively may have influence on the Social Opportunity aspect of COM-B: reflecting influences from senior staff members. However, this is only one of six sub-components of the COM-B. We need to ask questions on the various influences on students' science-related behaviours in order to understand and change them. By not asking students about the influences on their preregistration behaviours, we would be enforcing our own expectations on their dissertation experiences.

The design of this study allows for a unique exploration into how capability, opportunity, and motivation may impact students' preregistration processes. As noted in the manuscript, we will treat the COM-B analyses as exploratory, and discuss in the final manuscript the need for more targeted investigations into how behaviour change models may 'fit' within conversations surrounding Open Science uptake.

4. Finally, the last major point that I noticed in the design is the lack of knowledge about the type of dissertations that students are undertaking. It is possible that some students focus on exploratory qualitative studies or literature studies and

therefore do not feel able to do a preregistration, but also do not strengthen their statistical skills and confidence in the same way as other students who conduct a more quantitative study. A few more details about the specific study type, whether the dissertation is fit for preregistration, and the training acquired between T1 and T2 (integrity, statistical, open science, etc.) would be important to capture to exclude possible confounding factors and biases.

Response: We recognise that the appropriateness of preregistration to research paradigms that differ from hypothesis-testing, positivist, quantitative research has been called into question in the literature. Therefore, in our study we will only recruit students who are planning an empirical *quantitative* dissertation. This is included in the study materials and we have a question to check for this. In the manuscript, we now make this clearer on p. 11-12. Also, it is worth noting that the British Psychological Society dictates that students must complete an *empirical* piece of work during their final year, and literature reviews do not currently count as empirical contributions:

To be eligible for inclusion, participants are required to confirm that they are a final-year undergraduate student, studying Psychology at a UK institution **and planning an empirical quantitative undergraduate dissertation. Participants must have not already pre-registered their proposed undergraduate study at Time 1, and must confirm this in the beginning of the study.**

We also note the following in the discussion (p. 31)

Future work could also focus on how preregistration may be useful for different types of dissertation, including qualitative studies and analyses of secondary data.

Smaller points that are easier to address:

Note: For the sake of clarity, I added continuous line numbers to the document available in the OSF. I will refer to these page and line numbers (Pxlx) where relevant in the following points.

5. The abstract should mention the fact that participants are not randomly assigned to the groups, but self-assigned based on their completion of a preregistration.

Response: This has been added into the abstract.

6. P4I64 compat should read combat.

Response: Thank you for spotting this. This has been amended.

7. P4I72 From the beginning, it is not very clear what the 'Attitude towards statistics' means. A few examples of points may be helpful to avoid confusion early on.

Response: We now add a few examples to the constructs that come under ‘attitudes towards statistics’ when we first mention the term on page 6.

8. P6I109 The first sentence of this paragraph would benefit from a context setting to mention that this is in the UK since this differs in different settings.

Response: This has been added.

9. P7I142 not clear what this sentence means. Should ‘report worries’ be ‘students worry’?

Response: Yes - this has been amended.

10. P7I147 The sentence starting with “Indeed, an undergraduate publication...” should be moved one sentence ahead, it is out of context where it stands.

Response: This has been moved.

11. P9I186 The term ‘utility’ used in this paragraph is not entirely clear, especially in the first sentence of ‘utility in.... dissertation provision’. A better term may increase readability.

Response: This has been amended to ‘the pedagogical effectiveness’.

12. P9I188 “to improve students’ ... endorsement of QRPs” doesn’t sound quite right. I believe the authors mean the opposite.

Response: This has been changed to ‘awareness of QRPs’.

13. P9I190 It should be clear that T1 is always pre-preregistration. It becomes clear later on but I noted it as a question at this point in the manuscript.

Response: We have made amendments to make this clearer.

14. P9I193 the term ‘affect of statistics’ appears to invert the roles tested, i.e., to look at how statistics impact students rather than how the student perceives statistics.

Response: We now make it clear (p. 11) that this is about students’ positive approach towards statistics. The term ‘affect’ is used in the measure we are employing: The Survey of Attitudes Towards Statistics (SATS-28) which is why we use this term, for consistency.

15. P9I195 Is “self-reported competence” with statistics different from “less difficulty with statistics”?

Response: We have now made it clearer that these refer to the four distinct subscales of the within the Survey of Attitudes Toward Statistics (SATS-28). See page 11 on the revised manuscript for this more clearer.

16. P9I198 Endorsement seems like the test is about self-reported behaviours, where the questions are mostly about acceptance/tolerance with or perception of QPR.

Response: We have clarified in the hypothesis that 'endorsement' may also mean "or acceptance" for clarity.

17. P9I200 confidence in terminology sounds like faith in. Even if wordier, it may be clearer to say confidence in their understanding of OS terminology.

Response: We have now added "in their understanding of" after "confidence" throughout.

P10I223 Remuneration or compensation?

Response: We have amended this.

P11 maybe link to the power calculation available on the OSF?

Response: We have signposted this more explicitly now.

18. P11I238 name the university granting the ethics approval

Response: We have added this - University of Leeds School of Psychology.

19. P11I241 "ensuring they meet the inclusion criteria" → How could this be ensured?

Was a registry of students consulted? Was any assurance granted?

Response: We will include the inclusion criteria in recruitment materials and ask participants to confirm that they match the criteria via check-list boxes in the study. We now detail this explicitly on p. 12.

20. P13I268 The authors should mention that the grades are self-reported. The questionnaire also allows respondents to select 'Prefer not to answer'. What happens in this case? Will the data be included or excluded?

Response: This 'prefer not to answer' option has been removed from the materials, in the interest of data completeness.

21. P13I278 From what I understood, respondents mention whether they plan to perform a preregistration at T1, and mention whether they did it at T2. Are sub-groups created for controlling the answers of participants who planned it but in the end did not do it and those who did not plan it but in the end did it? At T1 are participants allowed to answer that they do not know whether they will preregister their study or not?

Response: Participants are allowed to say that they do not know whether they will preregister, as this item will not categorise participants into groups. Time 2 reporting of preregistration practices (i.e., those who have vs those who have not) will be the mechanism used to group participants for analyses. We include participants plans at Time 1 to keep tabs on recruitment of participant per group. We now make this clearer in of the revised paper. (see p. 13)

We will use ‘Cross Logic Quota’ sampling within Qualtrics (see Qualtrics, <https://www.qualtrics.com/support/survey-platform/survey-module/survey-tools/quotas/>) to roughly monitor group allocation at Time 1, although this will be done using the preregistration *plan* questions (see below), which could differ from the final preregistration group allocation at Time 2 (i.e., some participants could plan to preregister but do not actually preregister at Time 2).

22. P14I297 This may be a standard questionnaire, but I find the term ‘sensible’ very ambiguous in this test and potentially problematic for non-native speakers (i.e., it is often mixed with ‘sensitive’, and a ‘sensitive issue could approximate ‘problematic’).

Response: We have chosen to retain the wording ‘sensible’ because these items were taken from another published study which investigated student’s attitudes towards QRPS (Krishna & Peter, 2018; <https://doi.org/10.1371/journal.pone.0203470>). We want to accurately reflect their items as closely as possible, so as not to introduce new confounds in wording.

23. P14I310 maybe ass a few words to mention what an attention check is. I understood later, but it is not a term I heard very often.

Response: We have now added this on p. 15: “(i.e., to ensure that participants are actively paying attention to the survey materials and to prevent spam/bot respondents)”.

24. P15I335 “The same sample of students will be asked to complete the above measures again at Time 2” unclear that the COM-B is not included then.

Response: COM-B is a measure of planned behaviour; therefore, it is not appropriate to include it at Time 2 when behaviour is in retrospect. Participants will only complete COM-B at an exploratory measure at Time 1.

25. P16I347 I feel that asking whether participants plan on publishing in open access may also be relevant here, although an option to mention that they do not for financial reason may be needed then.

26.

Response: We have decided not to include this item, because most undergraduate dissertations are not published, and supervisors likely have the ultimate say over publication routes in light of, as you mention, access to funding and institutional support.

27. P17I382 ‘additive’ effect is not so clear. I would also say that going through the preregistration process has an effect, rather than the process itself.

Response: We have amended on p. 21. We have also added more detail here about our use of the term additive:

Specifically, we expect that despite any differences between groups at Time 1, there will be a significant interaction indicating that **engaging with** the preregistration process has an *additive* effect on students' attitudes, behaviours, and perceptions of Open Science (**i.e., it improves scores beyond improvement that occurs due to differences in time point**).

28. P18I395 reregistration à preregistration

Response: This has been amended.

29. P23I439 “Unlike with statistics attitudes...” This sentence contains many negatives and could be improved if rephrased slightly.

Response: We have now amended this sentence for readability.

30. P24I442 The section on the qualitative analysis is not clear. What will this analysis be used for? Which question will it answer or complement?

Response: We now add more detail to this section and align it with an exploratory research question.

**Note that revised study materials can be accessed here: <https://osf.io/7t9yd/>
For transparency, we have retained the previous version of materials on the OSF too.**

31. Demographic question 8 is not discussed in the manuscript. How will this information be used?

Response: We will use this information descriptively to describe our sample (see page 12 of the revised manuscript).

32. Study Design: “Collecting more data in order achieve significance” à Missing a ‘to’ and would be more accurate if phrased as “Collecting more data than planned...”
Reporting & analysis QRP: “selectively reporting studies” is jargon and may be very difficult to understand by psychology undergraduate students.

Response: We have amended the wording of the Study Design QRP. We have changed the latter to “Only reporting studies that support the results in a paper”.

33. What are the descriptors used in the scale 1-7

Response: We have now included this as - 1 (*Not at all confident*) to 7 (*Entirely confident*).

Brief COM-B measure

34. I would recommend to avoid 0s in a scale for ease of calculation without having to turn the values in log. These 6 questions are extremely difficult to grasp. The 'definition' section should be rephrased using the second pronoun and using simpler terms and examples. As it currently stand I can anticipate that respondents will guess more than answer knowingly to these highly abstract conceptual questions. As stated above, maybe consider using this scale about the dissertation itself.

Response: We will retain the 0, as this is in line with the conventions of the COM-B model.. Our inclusion of COM-B items to our study is an exploratory set of measures that tests whether students feel they have the appropriate capability, opportunity, and motivation to engage in the behaviour of preregistration. This will provide further information about potential barriers and facilitators of Open Science practices, which has been unexplored in this population.

Indeed, there is a growing emphasis on approaching open science behaviours through a lens of behaviour change in order to enact researcher change (e.g., Norris & O'Connor, 2019). All actors within the research process enact behaviours related to research, whether students or supervisors. In instances where students will be encouraged by their supervisors to preregister, this intuitively may have influence on the Social Opportunity aspect of COM-B: reflecting influences from senior staff members. However, this is only one of six sub-components of the COM-B. We need to ask questions on the various influences on students' science-related behaviours in order to understand and change them. By not asking students about the influences on their preregistration behaviours, we would be enforcing our own expectations on their dissertation experiences.

The design of this study allows for a unique exploration into how capability, opportunity, and motivation may impact students' preregistration processes. As noted in the manuscript, we will treat the COM-B analyses as exploratory, and discuss in the final manuscript the need for more targeted investigations into how behaviour change models may 'fit' within conversations surrounding Open Science uptake.

Attention check

35. Are there multiple choices to this question to make sure the participant reads through?

Response: Yes - this item will be embedded within the COM-B measures. We have updated the materials to reflect this.

Post-only questions

36. The question starting with "If yes" should come directly under the first question, not under the question about groups.

Response: This has been moved.

Value of pre-registration

37. I would highly encourage authors to repeat the definition of preregistration at this point, and maybe every now and then in the questionnaire.

Response: We now repeat the definition at this stage.