

Dear Thomas Evans,

Thank you very much for your thoughtful consideration of our manuscript and for providing such constructive feedback. We sincerely appreciate the time and effort you and the other reviewers dedicated to evaluating our work during both rounds of review.

We are delighted to hear that the revisions were effective in addressing the concerns raised by the reviewers and that you will be recommending the manuscript. We thank you for the opportunity to clear up the minor edits suggested by the second reviewer. We have also taken this opportunity to reflect further on the issue of sensitivity to heterogeneity raised by the first reviewer. We agree that preregistration of our approach at this stage is valuable, and we have amended the manuscript taking this into consideration. Please find our changes to each point addressed below and in the manuscript in color teal.

Once again, we sincerely appreciate your support and encouragement. We are thrilled by the prospect of your recommendation for our meta-analysis, and we look forward to the next steps!

Best regards,

Mariela Mihaylova for all authors

+++

Reviewer 1: My only remaining comment, for a future version, is regarding the publication bias tests. As indicated in my original review, the previously proposed methods are also sensitive to heterogeneity:

> "There seems to be an oversight here, because the included publication bias tests are also sensitive to heterogeneity. I include references below from which that follows"

The authors did not alter their methods substantively from what I gather, as the original publication bias section remains largely unchanged, save a few additions. I would be remiss to not point this out in an otherwise fantastic revision. I strongly encourage the

authors to revisit this as previously indicated - at the very least further discuss this in detail in the limitations if there is heterogeneity in the actual data.

We thank the reviewer for bringing up this important issue again. We acknowledge the importance of this issue and have revisited the publication bias section to address it. Additionally, we will expand more on the limitation of these methods in the discussion section once the work is conducted. We appreciate their thorough review and are committed to ensuring the robustness of our methodology.

We have added the below section on page 26 of the manuscript to reflect how we will deal with publication bias correction methods' sensitivity to underlying heterogeneity. We hope this addition is satisfactory.

“We also acknowledge that heterogeneity and publication bias are closely intertwined, and some measures of publication bias can be sensitive to underlying study heterogeneity (Harrer et al., 2021), which could affect the reliability and interpretation of our findings. One way in which we will disentangle the two in the current work involves conducting Egger's test to assess the presence of publication bias, while also evaluating heterogeneity using methods such as Cochran's Q or I^2 statistic, as outlined above. Significant heterogeneity may indicate that studies are estimating different underlying effects, whereas significant results from Egger's test could suggest publication bias. Moreover, we will perform sensitivity analysis using the leave-one-out method (Harrer et al., 2021) where effect sizes are recalculated with one study removed each time to assess the robustness of findings and identify potential outliers. Additionally, we will consider the sample size and quality of included studies when interpreting results, recognizing that small sample sizes and low-quality studies are more vulnerable to biases and spurious results (Brysbaert, 2019), which may influence our understanding of potential publication bias. Additionally, we will also assess study level power to check whether publication bias is likely (Quintana, 2023).”

Reviewer 2: page 9 Retrieval Practice main effect, the m and e of main effect could be capitalized.

We have revised accordingly: “Retrieval Practice Main Effect”

Page 9 first section, all sentences begin by “this” (but also last sentence page 8). Reading could be improved by a rephrasing.

We have rewritten this section by rephrasing the beginnings of sentences as per the below on both pages 8 and 9 (words changed in bold):

“However, other evidence from Clark and colleagues (2018) suggests a positive relationship between test anxiety and using retrieval practice when external incentives are applied, **suggesting memory** can be protected by retrieval practice in the face of stressors like test anxiety.

The above literature suggests mixed evidence for the protective effects of retrieval practice on memory following stress exposure. **Further investigation** is needed using a meta-analytic approach to determine the strength of the cumulative evidence for the protective effects of retrieval practice on memory following retrieval stress. **Addressing this** question is critical as it could imply that using non-invasive learning strategies might alleviate the memory impairment induced by stress. **Such an investigation** has the potential to challenge some of the major theories of stress, as it would suggest that there is a way to make memory less sensitive—and potentially protected—against what would be a stress-induced memory impairment. In terms of real-world value, the findings of this meta-analysis would additionally have major implications for designing learning-based interventions in applied settings such as schools and other learning environments.”

Page 14:

The date last searched was _____. At Stage 2, we reran the searches at least twice to ensure all literature was up-to-date. The date last searched was _____. The outcome was a total of YY prospective articles. After deletion of doubles, we had a total of XX articles (Figure 1).

I don't understand this sentence. You should conduct the search between stage 1 and stage 2 so why explicitly say that you will conduct the search at stage 2?

We agree this was unnecessary, however we took that wording from the original template. We have now revised to “We reran the searches at least twice to ensure all literature was up to date. The date last searched was _____. The outcome was a total of YY prospective articles. Following deletion of duplicates, we had a total of XX articles (Figure 1).”

Also I think it is better to say duplicates than doubles. You also began two following sentences by “after” and could find a synonym.

We replaced doubles with “duplicates” and we changed that sentence to: “Following deletion of duplicates” to avoid repetition with “after that” in the next sentence (as per above).

This enabled us to ensure full coverage : you can delete the “enabled us to”

We have deleted this.

I may propose you to rephrase “study leader” to “main contributor” in the text.

We replaced with “main contributor” throughout the paper on page 21.

Page 24 “we first illustrated” seems odd. You “display a funnel plot to check the existence of a publication bias.” Illustrated a publication bias would more mean that you draw a picture of a publication bias.

We agree this wording was unclear. We have now rephrased to “To assess small study effects, we first plotted the effect sizes and standard errors of each study, visually depicted in a funnel plot” on page 25.