Author response to reviews of

Manuscript RIRS-2022-0094

The combination of reporting bias and underpowered study designs has substantially exaggerated the motor learning benefits of self-controlled practice and enhanced expectancies: A meta-analysis

submitted to *International Review of Sport and Exercise Psychology*

RC: *Reviewer Comment*   AR: Author Response   □ Manuscript text

To Dr. Coffee,

Thank you overseeing the review of our manuscript for publication at the *International Review of Sport and Exercise Psychology*. We appreciate the positive and helpful comments we received from this review process. We believe these comments have helped to strengthen our manuscript.

Based on comments from both Reviewers, the associate editor, and yourself, we have made some changes to the manuscript. Below we provide our point-by-point response (in normal font) to the Reviewers' comments (in **bold-italic**). All substantial changes in the revised manuscript are included in this response letter (in a text box) and appear in blue text in the revised manuscript. Thank you again for considering our work and overseeing the review process.

1. Reviewer #1

RC: *The current manuscript evaluates the evidence behind the two pillars of optimal theory: enhanced expectancies and autonomy support. The manuscript uses a state-of-the-art statistical methodology that is well-suited for the desired inferences. I think that the manuscript provides important and information as well as a cautionary tale for researchers interested in OPTIMAL theory. Most of my comments regard minor issues and clarification of some parts of the introduction section. I find it manuscript concise with a great description of the methods (especially Table 1) description and I like how the authors connect results from the publication bias-adjusted meta-analyses with recently published large sample size pre-registered studies. This provides converging evidence highlighting issues discussed by the authors.*

AR: Thank you for the positive comments.

RC: *I found the ordering of information in the second and third paragraphs slightly confusing. You cite McKay and colleagues for the g = 0.54 meta-analytic estimate of self-controlled practice in the second paragraph, and then reference Bacelar and colleagues for g = 0.54 meta-analytic estimate of manipulated expectancies. Are those the same two estimates with different names (I am surprised that two meta-analyses would lead to a very same results)? Or is the McKay study a re-analysis of the Bacelar study (if so, I would recommend first introducing the original study, and then the re-analysis)?*

AR: The meta-analysis by McKay and colleagues and the one by Bacelar and colleagues are separate, independent

projects focusing on different literatures. We agree that it is surprising that both analyses resulted in the same point estimate. We have added a comment to lines 67-68 clarifying that these are identical estimates to clarify they are based on a completely different set of studies.

> On Page 4, Lines 66-68:
>
> It is noteworthy that self-controlled practice studies and the studies included in the enhanced expectancies meta-analysis have much in common (in addition to identical random effects point estimates based on the published literature):

RC: *line 62: you refer to one of the estimates as self-control practice and the second one as a Bacelar' et al. meta-analysis. Please, keep the naming consistent (ideally refer to them with the name of the effect, if they are indeed different, as it is more descriptive).*

AR: We have taken this suggestion and refer to the meta-analyses with the name of the effect in the revised manuscript.

RC: *Then you mention that both include studies from Lohse and colleagues which further complicates the matter. I would recommend splitting the description of McKay study from the second paragraph, which seems to be primarily concerned with publication bias, and describing the literature in chronological order.*

AR: We see how this was confusing and have taken the Reviewer's suggestion. We have moved the description of the self-controlled practice meta-analysis to its own paragraph and inserted a summary sentence on lines 49-50. The paragraphs are in chronological order based on first appearance of preprint or publication.

> On page 3-4, lines 49-50
>
> Therefore, it is necessary to carefully examine the literature supporting OPTIMAL theory predictions to identify and potentially correct for selective reporting and publication.

To clarify that only a couple studies from the meta-analyses were included in the Lohse analysis we have added the following clarification:

> On Page 4, lines 70-71:
>
> b) both include a study or two examined in a meta-analysis by Lohse and colleagues (2016) that found evidence of low power, multiplicity, and bias

RC: *Are there any overlaps in the data sets?*

AR: No, there are no overlaps in data. However, a two-experiment paper by Ghorbani (2019) included one experiment that was analyzed in the enhanced expectancies meta-analysis and one experiment that was analyzed in the self-controlled practice meta-analysis.

RC: *When describing the model types (line 135), I would recommend changing b) to something along the lines of: "homogeneous studies summarized with a common effect (fixed effects) or a distribution of heterogenous studies summarized with a mean effect and heterogeneity (random effects)" (which provides a much more precise description).*

RC: *Yes, this is a much more precise description and we have used it nearly exactly. Thank you for the constructive suggestion.*

AR:

> On Page 7, Lines 139-141
>
> b) homogeneous studies are summarized with a common effect (fixed effects) or a distribution of heterogeneous studies summarized with a mean effect and heterogeneity (random effects)

RC: *Description favorably contrasting the performance of RoBMA-PSMA should also include the limitation that RoBMA-PSMA might perform worse than the remaining methods under a combination of strong p-hacking and publication bias (Bartos et al., 2022)*

AR: Thank you for this suggestion. We have revised this section of the manuscript.

> On Page 10, Lines 206-211
>
> Simulations do suggest a simpler version of the RoBMA approach that includes only the PET-PEESE publication bias models (RoBMA-PP) is more effective in scenarios with strong p-hacking and publication bias. These strong performance indices combined with the ability to simultaneously model various plausible manifestations of reporting bias made RoBMA-PSMA (and RoBMA-PP) an attractive choice for re-analyzing the enhanced expectancies and self-controlled practice meta-analyses.

We have also included a footnote in the Results section about the RoBMA-PP models outperforming the RoBMA-PSMA in this situation.

> On Page 14
>
> We also fit RoBMA-PP models that perform better than RoBMA-PSMA models in the presence of strong p-hacking and publication bias. The RoBMA-PP models did not lead to substantively different conclusions for the analysis of self-controlled practice or enhanced expectancies.

RC: *Visualization of the prior distribution in Figure 2 does not seem correct – the spike at 0 seems to have much more mass than the continuous density. I would recommend using internal plotting functions provided by the RoBMA package directly for obtaining a ggplot object and for prior and posterior distribution plot).*

AR: The issues with the original visualization were a) the x-axis was truncated, so some of the continuous density was not shown, and b) the bin width of the histogram captured a small amount of the continuous density and included it with the spike at 0. We have fixed both issues by expanding the x-axis and creating a custom bin of width .001 at zero, resulting in a line that represents the spike null.

We have also created plots using the internal plotting functions suggested by the Reviewer and included them as supplementary material, along with figure captions to assist the reader. We believe the histograms will be more familiar and intuitive for most readers in our field, but readers with greater familiarity with Bayesian approaches may appreciate how the data are presented in the supplemental figures.

RC: *The two influential cases seem to have a huge effect on the results. Could you please provide some*

description of the studies? (I am not doubting their exclusion – the extreme increase in heterogeneity when including them provides additional justification for that.)

AR: We have added footnotes to provide some context about the two outliers in the previous self-controlled learning meta-analysis and the enhanced expectancies meta-analysis.

> On page 13, footnote 4
>
> The two outliers in the self-controlled practice were studies by Lemos and colleagues (2017) and Marques and colleagues (2017). Lemos et al. 2017 measured ballet movement form and tested the effect of choosing among video demonstrations while Marques et al. 2017 measured front crawl movement form and tested the effect of choosing which self-modeling video to watch during practice.

> On Page 13, footnote 5
>
> The two outliers in the enhanced expectancies meta-analysis were Goudini et al. (2018) and Navaee et al. (2018). The study by Goudini and colleagues measured linear tracing performance and tested the effect of feedback after good and bad trials. The study by Navaee and colleagues measured beanbag tossing performance and tested the effect of providing children on the autism spectrum with nonveridical feedback suggesting performance was 20% better than the average.

RC: *Authors should provide a greater discussion of the evidence for the presence of heterogeneity and its implications alongside the evidence for/against the effect and publication bias (lines 339). E.g., the presence of heterogeneity suggests that while there might be no evidence for the overall mean effect, there might be a distribution of effects and some of the studies might "work" (especially with tau = 0.49 estimate).*

AR: We have moved some information that was in the Limitations section of our original submission to connect directly to the discussion of heterogeneity, while expanding on this discussion. For clarity, the estimated tau with outliers removed was .35, not .49 (that was with outliers included) so we discuss the .35 estimate.

> On Pages 16-17, Lines 355-360
>
> The likely presence of heterogeneity suggests there is not one true effect, so individual studies may have been testing interventions with real benefits. However, this also means individual studies may have been testing interventions with real detriments as well, and we cannot discern which studies fall into which category. The estimated τ of .35 suggests the possibility of substantial variation in the effect of enhanced expectancies, highlighting the importance of cautious interpretation given the effect modifiers remain unknown.

RC: *lines 130, 224 should refer to Bartos et al. (2022) for RoBMA-PSMA (The Maier et al. 2022 article introduces the RoBMA methodology, before the extension with PET-PEESE models).*

AR: Thank you – we have updated this reference.

RC: *Line 185, please format "01" in "BF01" as a subscript*

AR: We have made this edit.

RC: *Line 202 should refer to Bartos and Schimmack (2022)*

AR: We have made this edit.

RC: *Line 245 should include a version of the RoBMA package*

AR: We have made this edit.

We thank Reviewer 1 for the constructive and thoughtful comments. Their feedback has improved the manuscript.

2. Reviewer #2

RC: *The authors utilized the RoBMA-PSMA technique to evaluate the effect of self-controlled practice and enhanced expectancies on motor learning. In addition, a z-curve analyses was used to estimate the statistical power of the enhanced expectancies studies. I do commend the authors for highlighting the need for better powered experimental designs.*

AR: Thank you for the positive comments.

RC: *Line(s) 30-31: It's not that Wulf and Lewthwaite (2016) argue that information-processing theories fail to account for such findings. Information-processing theories have accounted for motivation-related factors albeit a temporary influence (Holding, 1976; see Guidance Hypothesis). I suggest authors revise this sentence.*

AR: We now emphasize the distinction between temporary and more permanent influences of motivation to be as consistent as possible with the arguments forwarded by Wulf and Lewthwaite in the "Need for a new perspective" section of their paper (page 1383).

> On page 3, Lines 30-33
>
> Wulf and Lewthwaite (2016) argued that information processing theories account for only a temporary influence of motivation, but fail to capture the more permanent influences of motivation on motor learning suggested by these findings; this failure to account for more permanent motor learning effects created the need for a new theory.

RC: *Line(s) 39-40: Authors should either provide description in their response or within the introduction regarding what they mean by "motivational rather than informational mechanisms." While the OPTIMAL theory does provide mechanistic explanations, the intentional binding effect for example can also be explained through an informational mechanism.*

AR: We revised this part of the Introduction to provide a more inclusive summary of the putative contribution of OPTIMAL theory to advancing our understanding of motor learning.

> On Page 3, Lines 39-42
>
> While motivation had been deemphasized in previous motor learning theories, within OPTIMAL theory numerous lines of 21st century research may also be explained by motivational mechanisms, rather than or in addition to informational mechanisms.

RC: *Line(s) 66-68: I would reword this sentence as even Bacelar and colleagues state, "However, it is important*

> *to note that the present meta-analysis did not assess the gray literature, and, therefore, does not present direct evidence of reporting bias."*

AR: On Page 4 (Line 71), we have changed "reporting bias" to "small study effects".

RC: *Line(s) 69-88: This is a great explanation and I comment the authors for their clear and concise messaging.*

AR: Thank you for this positive comment.

RC: *Line 90: Please provide a reference for, "A challenge for most reporting bias models is large heterogeneity in true effects"*

AR: We have made this change and now reference Carter et al. (2019) here.

RC: *It is not clear what prior distributions of the effect parameter were assumed in the modeling process. The robustness of Bayesian methods cannot be taken for granted but is dependent on the "quality" of prior beliefs in terms of subjectivity of the prior information. I apologize if I missed this information, but if this can be added, that would be helpful.*

AR: Thank you for catching this. It was an oversight on our part as this information was previously only reported in the figure caption. We have updated the manuscript accordingly.

> On Page 9, Lines 178-182
>
> Similarly, the prior distribution for the presence of an effect was divided evenly between a spike prior equal to zero effect (i.e., the null) and a normal distribution of plausible true effects with M= 0 and SD = 1. Likewise, the prior probabilities of a fixed effect and heterogenous effects were divided evenly, with plausible values of τ following an inverse gamma distribution with shape = 1 and scale = .15.

RC: *It is also unclear if the Bayesian method was tested for sensitivity regarding the author's prior assumptions. Inherently, this could have influenced the weighting functions. Please provide this information in the paper or in your response.*

AR: We use neutral priors that were dominated by the data for each posterior estimate. For example, the prior probability of no publication bias was .50. Among the publication bias models, the prior probability was divided evenly between the weight function selection models and the PET-PEESE regression models. The former modeled publication bias with thresholds while the latter considered continuous processes and we took a neutral stance on which process was more likely. Had we used an informed prior with respect to the probability of publication bias in our samples, we would have assumed it was more than 50% likely such bias exists based on our knowledge of the field and previous research (e.g., Lohse et al., 2016). Therefore, we consider the evidence in favor of bias to be somewhat conservative.

We have added the following (immediately following the additions from the previous comment).

> On Page 9, Lines 182-185
>
> We chose neutral priors to allow the data to dominate our posterior estimates and not impose our own opinions on the analysis. Admittedly, our prior beliefs were that publication bias is likely, so the analysis we report is somewhat conservative in this respect relative to our personal priors.

With respect to the effect estimates produced by our models, we considered the possibility that the normal

distribution with Mean = 0 and SD = 1 is too wide for plausible effects in this field. For example, a meta-meta-analysis of the social psychology field reported a median effect of d = .36 (Lovakov & Agadullina, 2021). Assuming a similar distribution of plausible effects for self-controlled practice and enhanced expectancies (both being somewhat social psychological phenomena), we might have a prior distribution of plausible effects with Mean = 0 and SD = .36. We tested the impact of using this prior distribution and the results were nearly identical. We have included the following footnote to advise the reader of this sensitivity analysis:

> On Page 9, Footnote 1
>
> We considered a more informed distribution of plausible effects based on the empirical benchmarks reported by Lovakov and Agadullina (2021) for social psychological phenomena. In this sensitivity analysis, we included a normal prior distribution with M = 0 and SD = .36. The results of this sensitivity analysis were nearly identical to the primary analysis.

RC: *The method is robust but is an additional tool to the frequentist meta-analysis that can help achieve a more reliable result. However, they are two different methods fundamentally. While each on their own may benefit the current state of the literature, the comparison (similar or not) is like comparing apples with oranges. I see value in this well written manuscript if presented independently and with explanations regarding results from one method to another.*

AR: We have revised this section to highlight that the present results vary from the previous findings based on the underlying assumptions of our methods and the type of analysis we have conducted.

> On Page 18, Lines 390-397
>
> Indeed, the present study uses new methods and different assumptions to correct for the bias that Bacelar and colleagues (2022) suspected in their previous study. Further, the Bayesian approach currently employed offers an updated belief in the likelihood of bias, heterogeneity, and the presence of an effect while considering plausible forms of selective reporting. Conversely, the frequentist approach employed previously offers a test of the null hypothesis that there is no effect of enhanced expectancies manipulations while assuming selective reporting is absent. The two studies offer valid answers to different questions.

Thank you again for the positive comment. We appreciate the thoughtful points raised by Reviewer 2 who has helped to strengthen our manuscript.

3. Associate Editor

RC: *Thank you for submitting your work to IRSEP. Two reviewers have now provided detailed comments on your manuscript. While both agree on the quality of the work, there remains issues regarding the clarity and interpretation of terms, statistics, and theoretical concepts that need addressing. I look forward to reading a revised manuscript in due course.*

AR: Thank you for the feedback and we hope we have effectively addressed all the concerns raised by the reviewers.