**Reviewer 1**

Title: Optimizing the Benefits of Mental Practice on Motor Acquisition and Consolidation with Moderate-Intensity Aerobic Exercise.

This article seems well built and brings evidence of a phenomenon not yet fully understood and that certainly deserves further study.

Major comments

Abstract: should report more details about the anthropometric data (all sample), unclear the exercise intensity used. The results are missing about the significance levels.

Methods: body mass and body height are missing please check

Several parts of the main document the aerobic term was used without a correct interpretation https://pubmed.ncbi.nlm.nih.gov/27747843/

Precision variable was not measured! Why?

Speed term is not appropriate for this investigation while

In several part of the main document the u.m. are missing (Table 1) please check

Movement speed (total number of sequences) should be “n” / therefore “Movement speed” is inappropiate for this variable

EMG data are missing

Line 122: please to include the load (watt) and normalization procedure for body weight for each subject; anyway, unclear the bike setting used (https://pubmed.ncbi.nlm.nih.gov/22183087/)

Line 180: sample power estimation is missing please check. https://pubmed.ncbi.nlm.nih.gov/32558628/

Minor comments:

Title: I suggest delete “Aerobic” term could be misleading for the readers  https://pubmed.ncbi.nlm.nih.gov/27747843/

Abstract: Line 23…unclear this sentence seems to be generic while should be more detailed in line with the aim of this study

After this first part of my revision on Introduction/discussion was not completed because this submission showed several flaws

**Reviewer 2**

This manuscript presents a study investigating moderate-intensity aerobic exercise's effects on motor acquisition and consolidation induced by mental practice using motor imagery. The study found that exercise after mental preparation enhanced motor consolidation, and the timing of aerobic exercise concerning practice sessions is critical in promoting motor consolidation. The study's results have implications for sports training, physical therapy, and other areas of motor skill development.

I have no expertise in sports medicine. Therefore, I appreciate the patience of the authors and the recommenders with the general nature of my comments that will focus on scientific methods.

The manuscript is well written. The rationale is clearly stated, and the method is straightforward and comprehensive. Importantly, this article has no critical design, analytical, or interpretation flaws. In addition, the authors have documented expertise in their field.

The article's structure is on point: the report's different sections are used appropriately, thus facilitating the reading and understanding of the authors' work. The Methods content is clear enough to allow for a verification study if needed. The research questions are well stated in the introduction and discussed at the end of the manuscript. The results are precise and well-stated. The discussion covers the results and additional relevant literature to elaborate on their findings.

I have a few minor suggestions, mainly to improve the manuscript for nonspecialist readers:

1. The authors could define aerobic exercise earlier in the introduction, as it has been used in daily lives in various contexts, while it has a defined meaning in their work.
2. The authors document the score of imagery vividness in their methods. However, it would be helpful to state clearly that groups do not statistically differ. Moreover, the reader might find it informative if a score of 40+ is above average or not concerning the demographics of the test subjects.
3. Even after the author's design justification, It is not clear why it would not be beneficial to standardize the mental and physical practice number of repetitions. However, the authors might consider reworking that section for a more general audience.
4. The authors might explain why their design does not test PP and aerobic exercise (prior or after) or how PP is a reasonable control for testing MP and aerobic training.
5. I wish to remind the authors that a given p-value does not preclude the effect size of the data they are analyzing. When a threshold of significance is stated, it should be respected. I would ask them to change this statement: "marginal difference between the MP-Exe group and the MP- Rest group (p=0.0501)".

Thank you for considering my comments.

**Reviewer 3**

The present manuscript examines whether mental practice benefits from aerobic activity. This follows up on recent results indicating that physical practice is improved by aerobic activity. Different groups completed physical practice alone, mental practice alone, mental practice preceded by exercise, or mental practice followed by exercise. Assessments of performance were conducted before, immediately after, and 24 hours after the intervention. Results suggest mental practice followed by exercise led to a beneficial effect on consolidation.

The research question is interesting and timely. The overall design of the study is appropriate to address the questions being examined, and inclusion of the groups performing exercise both before and after imagery is a positive as it allows examination of effects on consolidation. However, the small sample sizes for each group, coupled with analyses that are conducted in relation to changes in performance when absolute performance appears to differ between groups, make it difficult to interpret several results.

Major Points:

Sample sizes are relatively small with only n=10 per group. This is a concern as the raw data in table 1 suggest that the groups are not equivalent at baseline, and the variability in performance across groups is high. In the pre-test assessment, the raw measurements of movement speed and accuracy are both higher for the MP-rest group than others. While there is not a significant difference detected between the groups by the ANOVA, it’s difficult to say whether this is due to their truly not being a difference, or if there is an issue of sample size that prevents it from being detected. If the initial sample sizes were larger, we’d assume that the average baseline performance of the different groups would be closer, and this would become less of a concern.

As most comparisons in the analysis are based on percentage changes in performance, this makes it difficult to interpret several results. For example, the MP-Exe group are the lowest performers in baseline and at post-0h. While they make some improvements in speed and accuracy at the 24h follow up, they are still the worst-performing group observed overall. So when measured as a percentage change from the previous assessment, they make the largest change in performance. However, they also had the greatest potential to continue increasing, as the other groups are presumably much closer to their ceiling, so have less potential to improve.

Overall I would recommend increasing the sample size for each group. Performing equivalence tests/Bayesian analysis could also help to demonstrate whether the groups can truly be considered equivalent at baseline/at other points.

Data are presently analysed using separate one-way ANOVAS comparing the different groups, with separate anovas at each time point. In this situation the results would be stronger if preceded by a mixed-model ANOVA with within-factors of timepoint (Pre, Post 0h, Post 24h) and between factors of group (PP-Rest, MP-rest, Exe-MP, MP-Exe) would be stronger – see https://www.nature.com/articles/nn.2886

Minor points:

How was maximal theoretical cardiac frequency calculated (reference given, but useful to give the formula e.g. 220-age or some other calculation).

How was accuracy defined (e.g. was it possible to make multiple errors on the same sequence? What feedback was provided for errors?)