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

We thank the reviewer #1 for these comments. However, we did not collect anthropometric data from the sample, since we did not consider as variables that could influence our main outcome (accuracy and speed on a finger-sequence task). We agree that this information is usually reported when focusing on physical exercise. Nonetheless, we took a certain percentage of the theorical cardiac frequency to quantify moderate-intensity exercise. The participants were recruited from the Sport Science faculty (we added this information in the text).

We added the statistical p-values in the abstract.

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

We now report the physical effort as moderate-intensity exercise throughout the document.

Precision variable was not measured! Why?

We measured accuracy as the number of correct sequences performed in 30 seconds.

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

We agree with the reviewer that the international unit of speed is m/s. However, in our article, we used movement speed related to the total number of sequences since the duration of the trial (30 seconds) was similar for all participants. If one participant performed a greater number of sequences in comparison to another, we can say that he/she was faster.

EMG data are missing

We thank the reviewer #1 for pointing out this omission. We now present the EMG data per block in Table 2, at the end of the result section.

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/)

We thank the reviewer for this comment. However, we did not normalize the load relative to body weight, but we normalized between participants, asking them to cycle at 65-70% of their maximal theoretical cardiac frequency, based on Riebe et al., 2018, ACSM’s Guidelines for Exercise Testing and Prescription. American College of Sports Medicine.

The resistance of the bike was set individually to reach this percentage.

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

Since the task we used in the current study was not similar to the ones developed in previous studies, we determined our sample size based on the number of participants recruited in previous studies: Statton et al., 2015 (3 groups of 8 participants); Truong et al., 2022 (4 groups of 11-12 participants); Freitas et al., 2020 (4 groups of 12 participants).

Statton MA, Encarnacion M, Celnik P, Bastian AJ. A Single Bout of Moderate Aerobic Exercise Improves Motor Skill Acquisition. PLoS One. 2015 Oct 27;10(10):e0141393. doi: 10.1371/journal.pone.0141393. PMID: 26506413; PMCID: PMC4624775.

Truong C, Hilt PM, Bouguila F, Bove M, Lebon F, Papaxanthis C, Ruffino C. Time-of-day effects on skill acquisition and consolidation after physical and mental practices. Sci Rep. 2022 Apr 8;12(1):5933. doi: 10.1038/s41598-022-09749-x. PMID: 35396365; PMCID: PMC8993858.

Freitas E, Saimpont A, Blache Y, Debarnot U. Acquisition and consolidation of sequential footstep movements with physical and motor imagery practice. Scand J Med Sci Sports. 2020 Dec;30(12):2477-2484. doi: 10.1111/sms.13799. Epub 2020 Sep 19. PMID: 32777113.

Minor comments:

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

We now refer to moderate-intensity exercise

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

We change the sentence as follows: “We hypothesized that exercise before MP would further increase speed and accuracy at a finger-sequence task measured right after MP (potentiation of motor acquisition), whereas exercise after MP would further increase speed and accuracy the day after MP (promotion of motor consolidation)”

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.

We thank the reviewer for these positive remarks.

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.

In response to Reviewer #1, we deleted the term “aerobic” and changed it to “moderate-intensity” exercise. This precludes to the confusion between daily life activities and moderate-intensity exercise, such as biking, as in the context of this experiment.

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.

We added the information that the MIQ-R score of the MI groups did not differ ($F_{2,27}$=0.182, p=0.83). We also added the reference of Guillot et al., (2008), who showed that healthy young participants (such as in our study) considered as good imagers have an average score of 45.7 ±2.7 while poor imagers had 38.6 ±5.2. We now changed the sentence “In the current study, the average scores suggest average to good motor imagery ability”. To note that initial imagery ability assessed with the MIQ-R score does not correlate with the performance gain following mental practice (Ruffino et al., 2017).

Guillot A, Collet C, Nguyen VA, Malouin F, Richards C, Doyon J. Functional neuroanatomical networks associated with expertise in motor imagery. Neuroimage. 2008 Jul 15;41(4):1471-83. doi: 10.1016/j.neuroimage.2008.03.042. Epub 2008 Apr 4. PMID: 18479943.

Ruffino C, Papaxanthis C, Lebon F. The influence of imagery capacity in motor performance improvement. Exp Brain Res. 2017 Oct;235(10):3049-3057. doi: 10.1007/s00221-017-5039-8. Epub 2017 Jul 21. PMID: 28733754.

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.

Indeed, we could have matched the physical practice session to the mental practice session, i.e. three blocks of ten 10-sec trials. Instead, we matched the physical practice to the trials performed during the tests, i.e. 30-sec trials. Anyhow, the total amount of mental and physical practice was similar: 300 seconds of imagined or actual movements, distributed in trials/blocks. We changed the description of this section, as follows:

After Pre-Test, all groups (physically or mentally) repeated the movements for a total amount of practice of three hundred seconds. For the PP-Rest group, the practice session consisted of ten 30-second trials with the same instructions as in the Pre-Test. Each actual trial was followed by thirty seconds of rest. For the participants of the MP groups, the practice session consisted of three blocks of ten 10-second imagined trials of the sequence. We chose 10-second trials for MP to minimize the deleterious effects of long trials on motor imagery clarity (Rozand et al., 2015). Each imagined trial was followed by ten seconds of rest. Blocks were separated by 1 minute of rest.

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.

We did not include PP combined with moderate-intensity exercise as this was not the prime rationale of the study. However, this could be an interesting follow-up to compare the percentage of increase between PP and MP, when combined with exercise. In the psychological and sport literature, there is often the question of the benefits induced by PP and MP. Usually, MP increases motor performance, but to a lesser extent in comparison to PP alone. In the current study, we considered PP as a relevant control to determine how close/far is the performance improvement following MP associated with exercise in comparison to PP alone.

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)".

The sentence has been changed as follows: “The difference between the MP-Exe group and the MP-Rest group was not significant (p=0.0501), but with a moderate effect (Cohen’s d=0,71)”.

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

We understand the general concerns of Reviewer #3. We responded in a single block as they mainly relate to statistical analysis and interpretation.

We performed an ANOVA to test any difference between groups at Pre-Test, and we did not find any: line 202 "All groups showed similar performances in the Pre-Test as ANOVAs did not yield differences between groups for both movement speed ($F_{3,36}=0.93$, $p=0.43$) and accuracy ($F_{3,36}=0.82$, $p=0.49$)". In order to focus on performance improvement and to take into account the inter-subject variability, we decided in first instance to perform analyses on gain for acquisition and for consolidation, instead of on raw data.

We then performed 2 separate ANOVAs, as our fundamental questions were separate: one for acquisition and one for consolidation. In order to avoid false negative, we did not perform a mixed-model ANOVA as suggested by reviewer #3. We like the paper by Nieuwenhuis et al. in Nature Neuroscience, showing the errors in statistical analysis. In the current study, we normalized post-test data by pre-test measure to reduce variability between participants, then we performed an ANOVA to test for between-group differences on ratio, and not on raw data at post-test to avoid such type of errors as mentioned in the article by Nieuwenhuis et al.

The number of participants looks small regarding the number of groups. However, previous studies focusing on similar topics recruited between 8 to 12 participants per group. Since the task we used in the current study was not similar to the ones developed in previous studies, we could not perform an a priori estimation of sample size but we determined our sample size based on the number of participants recruited in these previous studies: Statton et al., 2015 (3 groups of 8 participants); Truong et al., 2022 (4 groups of 11-12 participants); Freitas et al., 2020 (4 groups of 12 participants).

We also added Cohen's d for pairwise comparisons to further provide information about differences (or not) between groups, even if the samples were small. This gives further details on the statistical analysis and reinforces the current results (with moderate to strong effects).

To check whether there was a ceiling effect, we asked 5 new participants to perform 20 actual trials (2 blocks of ten 30-sec trials, with 30-sec rest between trials and 1-min rest between the 2 blocks). The performance at the beginning of the trials was 14.3 $\pm$2.2 sequences for speed, and at the end of the practice 21.6 $\pm$3.2. The gain for acquisition was 52.3% in average. For comparison, the MP-Rest group started at 15.3 $\pm$4.6 sequences and ended at 20.1 $\pm$5.3 sequences (with a gain of 24.6%). Independent t-tests showed significant difference between MP-Rest group and the Actual group who performed 20 actual trials for raw performance at Post-test 0 ($p=0.04$) and gain during acquisition ($p=0.004$). For this specific task, it seems that all groups included in the study did not reach the performance ceiling.

Statton MA, Encarnacion M, Celnik P, Bastian AJ. A Single Bout of Moderate Aerobic Exercise Improves Motor Skill Acquisition. PLoS One. 2015 Oct 27;10(10):e0141393. doi: 10.1371/journal.pone.0141393. PMID: 26506413; PMCID: PMC4624775.

Truong C, Hilt PM, Bouguila F, Bove M, Lebon F, Papaxanthis C, Ruffino C. Time-of-day effects on skill acquisition and consolidation after physical and mental practices. Sci Rep. 2022 Apr 8;12(1):5933. doi: 10.1038/s41598-022-09749-x. PMID: 35396365; PMCID: PMC8993858.

Freitas E, Saimpont A, Blache Y, Debarnot U. Acquisition and consolidation of sequential footstep movements with physical and motor imagery practice. Scand J Med Sci Sports. 2020 Dec;30(12):2477-2484. doi: 10.1111/sms.13799. Epub 2020 Sep 19. PMID: 32777113.

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).

The calculation for theoretical maximal cardiac frequency was indeed 220 – age. The reference has been added: Fox S. M., III., Naughton J. P., Haskell W. L. (1971). Physical activity and the prevention of coronary heart disease. Ann. Clin. Res. 3, 404–432

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

Yes, it was possible to make multiple errors on the same sequence, but this was not taken into account. Once the participant makes an error within the sequence, this latter is subtracted from the total number of sequences to measure accuracy.