Dear Recommender and Reviewers,

First of all, thanks for organizing this review and recommendation session. This kind of feedback is extremely valuable to us. And we particularly appreciate the quick turn around time and the constructive tone of the feedback. So thank you! We hope that we have captured in the revision all points raised.

We apologize for not making it clear in our previous submission that we had already begun data collection for this study, and that we would not be able to make changes to our experimental procedures. We hope that the inclusion of a Timeline at the end of the manuscript makes our agenda more clear now. The reviewers made several valid points on possible side effects and interactions of tDCS,TMS, and motor learning that we now discuss in the Caveats section. Our responses to the comments below are in green text and any new text since the previous submission is in blue.

Sincerely,

Gavin Hsu and Lucas Parra

Dear Lucas Parra,

Thank you for submitting your Registered Report on "Dose-response of tDCS effects on motor learning and cortical excitability" to PCI-RR. I received two detailed reviews that found your study of interest and raised a couple of points, such as fatigue, MEPs/TEPs, safety protocols, typing test, lateralization etc.. I would like you to address all issues raised by the reviewers in a revision.

Moreover, there were some confusions about pilot data and it was not clearly stated in the manuscript that data collection is ongoing. I found in the meta-data of your submission that your study is considered bias-control level 2. As this means that changes in the experimental design are not possible anymore, you need to discuss the points raised by the reviewers regarding design, add them as possible limitations, and be transparent about the current state of the study. By transparency I mean that you might add a section on "Proposed timeline" or however you want to name it, that is typically added to Registered Reports Stage 1, and add details about what was done when and what will be done when.

> Thank you for raising this point. We have added a timeline to clarify this point. We have at this point collected over 100 of 120 planned participants. We have now indicated that the design of data collection is frozen.
>
> We would appreciate your guidance on this classification of bias-control. We selected level 2, but are not entirely sure that is the correct classification. Some of the data have been accessed (by a blinded experimenter) to establish MEP and TEP processing, but we are and will remain blinded to the treatment variable until the end of data collection. Does this situation correspond to “bias-control level 2”?
>
> Either way, we have added a timeline to address these issues.

Thank you for considering PCI-RR! Looking forward to your revision.

Best regards, Christina Artemenko

*by **Christina Artemenko**, 07 Mar 2024 08:11*

Manuscript: **https://osf.io/nz2ja**

version: 1.1

Review by anonymous reviewer 1, 21 Feb 2024 13:25

In this pre-registered study, Hsu and authors plan to investigate a very important and interesting question that pertains to the field of brain stimulation: dose-response of tDCS effects with motor learning and cortical excitability. While this topic is of interest and the overall manuscript is well-written, I have some concerns about the complexity of the study design.

The incorporation of challenging physiological measures such as TEPS and MEPs from both hemispheres, alongside multiple motor learning assessments within a single experimental session, raises concerns about participant fatigue and its potential influence on motor performance and physiological measures. The time-consuming nature of tasks such as TMS-EEG setup and obtaining MEP measures for both hemispheres could lead to fatigue among experimenters and participants alike. Additionally, the extensive number of MEPs per time-point may be excessive for estimating cortical excitability.

> Thank you for pointing out the potential fatigue effect. In our previous study (Hsu et al., 2023) some subjects anecdotally reported that the final task after the break felt more difficult, possibly due to fatigue, and performance in the last sequence dropped compared to the first. The present study adds EEG setup and TMS prior to the tDCS interventions. In both, the participant is passively sitting on a char and resting. The preceding typing test only takes around 1 minute for most people. The second TMS session is during the pause and does not add any extra length. Nevertheless, we have added this as a caveat, and because the procedures in this study are longer than in the previous, we will interpret the exploratory follow-up results with caution:
>
> > Except for a short typing task, the present study is identical to our previous study[49] in terms of behavioral demands to the subject. Nevertheless, the procedures add about 60 minutes at the start with the participant at rest. Any overall drop in performance between the initial 12-minute training session and the follow-up tasks after a 60-minute pause could be the result of fatigue.
>
> We acknowledge that the number of MEPs is high, but this is not unusual in MEP studies that emphasize rigor. TMS-EEG studies also use a higher number of trials. We now include these arguments in the text:
>
> > Where MEP protocols can range from 20 to 100 TMS pulses[4,67,77], we opted for 60 pulses as suggested for reliability in TEP measurements[66]. The total number of TMS trials falls within historical safety guidelines[78] and the amount used for motor mapping[79].
>
> We believe that experimenter fatigue is not a major concern. Although the setup for TMS-EEG is time-consuming, with two experimenters doing it simultaneously, we are usually able to finish this within 20 to 30 minutes. With 5-second intervals between MEP

trials, 60 trials only take 5 minutes, and from our experience we find that this amount does not place excessive physical burden on the experimenter.

Furthermore, while I believe your current design adequately addresses the question regarding the effect of tDCS dose on motor sequence learning, I am unclear about the integration of the concept of generalization into the project. It is essential to clarify the expected outcome and consider how the experimental procedures, including the measurement of MEPs and TEPs, following a break might influence aspects of motor consolidation and generalization.

We acknowledge that TMS could influence the subsequent task performance, and that the design does not rule this out. However, we are not aware of previous studies that found a significant effect of single-pulse TMS at 0.2Hz, and our follow-up experiments are exploratory and not the main focus of this study. We only include them to see if they replicate our previous results. We expanded on the caveat mentioned above as follows:

> Additionally, the TMS during the pause could in theory affect task performance. Thus, this study is not properly designed to test for lasting effects or carryover effects. A failure to replicate our previous results on this could be a consequence of the intervening TMS or fatigue.

We have also removed reference to generalization. Instead we write as follows:

> To test for any lasting effect on the unstimulated hemisphere, the task will then be repeated without stimulation on the right hand on a new sequence (R:S2). We will also test a new sequence on the left hand (L:S3) to see whether tDCS can boost subsequent learning on the stimulated hemisphere.
>
> Exploratory analyses include a test for lasting and carryover effects as observed in our previous study[46]. There we did not find a significant performance difference after 1 hour nor between the different hands or sequences.

Regarding the question of whether concurrent tDCS affects corticospinal excitability, I have reservations about the current design's ability to provide adequate testing. Combining tDCS with motor learning may saturate cortical excitability, making it challenging to isolate the effects of tDCS dosage on cortical excitability independent of motor learning.

We agree that this is a possible limitation of the study design, and have added the following discussion as a possible caveat:

> It is possible that the combination of tDCS with motor learning may saturate excitability. We should see in the control group whether the learning task on its own has any effect on MEPs and TEPs. In case of an effect in the control group, we would include this as a caveat that the study is not properly designed to isolate tDCS effects from motor learning alone. However, there is some evidence that sequence learning tasks alone do not affect MEPs[54,55], as opposed to repeated strength exercises[18,50–53].

To address these concerns, I suggest considering two separate experimental sessions: one focusing on recording physiological measures before and after tDCS administration (with groups separated by tDCS dose), and another session dedicated to testing the effects of motor learning. This approach would help mitigate issues related to fatigue and provide a clearer understanding of the independent effects of tDCS dosage on cortical excitability.

> Thank you for the suggestion and this is a great idea for a potential followup study regardless of whether we have a positive or negative outcome here. We apologize for the confusion and we have now made it clear with an inclusion of a timeline that we have already begun data collection and cannot change the experimental design in terms of data acquisition. Having said that, we still believe that a positive outcome in either behavioral and MEP results with this design would still provide conclusive novel evidence on effects of tDCS on motor learning that this field is currently lacking.

Review by Charlotte Wiltshire, 05 Mar 2024 17:12

This study aims to understand the neuromodulatory effects of HD tDCS at higher stimulation intensities on motor learning and cortical excitability. The design appears to be appropriate and well powered. Given the growing inconsistency in this field, this is a timely and interesting study! I do have a few comments and concerns. I would like to hear more about the safety protocols that are being used, particularly the interaction of tDCS and TMS protocols. I also have some concerns over the “pre-test” used to assign participants to the tasks. Please see comments below for details.

Rationale is good, though further discussion of the discrepancies between positive and negative findings would be useful. There is not much detail on what the scepticism around tDCS is based on, for example, and what this would mean for the field if there is no relationship between tDCS and cortical excitability. There are a couple of references to work that has not shown an effect of tDCS on motor cortex excitability but they are not discussed in detail and other work could be added to build a better picture of the current state of the field.

> When there is controversy, the review process can be treacherous. So we intentionally said the minimum possible, limiting ourselves mostly to citations. We have now added a brief discussion in the introduction that will hopefully not stir up too much controversy:
>
> > Some null results of tDCS on MEP may be explained as reversal of effects at higher intensities[28] leading to a “no man’s land” at intermediate intensities[29]. Our primary hypothesis is that of a monotonic effect of tDCS, but we remain open to the possibility of a reversal of the effects on MEP or motor skill learning. If the dose response of MEP reverses with increasing tDCS intensity, but is monotonic for skill learning (or vise versa), it would be more difficult to reconcile this with a common physiological substrate.

Experimental design, including researcher blinding is appropriate.

Will the time of day be set for all of the participants? Justification for/against? (Effects on GABA, wakefulness, caffeine, food).

We have fixed times for morning and afternoon sessions such that we have adequate time to run two subjects per day. Given the large sample size of this experiment, we can not afford to have single subjects per day. We have added these details to the Procedures section:

> Up to 2 experimental sessions will be conducted per day, starting either at 9:00 AM or 1:00 PM.

This is a good suggestion to consider these additional factors. Adding time of day as a covariate has the potential to increase power (if it has some predictive ability) or can also reduce power (if it is not related to the outcome). There are a multitude of such variables we could include (time of day, sex, years of schooling, caffeine, medication, the list goes on). Given that we have no specific hypothesis on these factors, we prefer not to include any of them as a covariates in the primary analysis. We will report the descriptive statistics for the following:

> We will report descriptive statistics for samples organized by time of day (morning vs. afternoon), age, sex, and stimulation condition.

If we find no effect of tDCS, we will include time-of-day as a binary covariate in a post-hoc exploratory analysis. If we do find an effect, there could be numerous followup studies to determine multiple other factors that modulate performance.

Caffeine intake prior to the experiment should be monitored for both participant safety and the effect on the experiment. I suggest keeping caffeine intake in line with the participant's normal dose, but no caffeine within 1-hr of testing.

This is a good suggestion that we had considered ourselves but decided not to include, as it is hard to ensure compliance. The only instruction given to the subjects is to be well rested. Caffeine will be one of the many possible sources of variation that will remain uncontrolled. Additionally, as we now make clear in the timeline, data collection already started, so we can no longer monitor or restrict caffeine.

The typing test:

I'm a little confused about this task. In the procedures section, it describes the task as typing real (English) words. In the "Predicting motor learning ability" section, it describes it as a series of letters. This distinction is important and relates to my next comments.

We apologize for the confusion and thank you for pointing this out. The description of the task was duplicated. It is the same in the main experiment, as in the pilot we did to establish a correlation with learning performance. We have consolidated and simplified the description now at the beginning of procedures as follows:

> Subjects will be asked to correctly type a series of 63 common English words as quickly as possible using both hands. The words will appear in a sequence of 22 characters on the screen (including spaces), with the text scrolling along leftward as the subject correctly types out characters. A stationary cursor will highlight the leftmost character to be typed next. The task will continue until all characters have been pressed correctly. There will be no time limit on this task, but we estimate that it can typically be completed in approximately one minute.

The typing test seems weighted towards those who have learned to touch-type, rather than capturing a general measure of de-novo sequence learning or baseline sequence speed. This may also be affected by the languages spoken by the participant. This would need consideration of the number of languages spoken and the proficiency of English, though I would prefer to see a test that is not 1) linked with language use and 2) linked with a skill that the participants may or may not have previously learned (touch typing). Similarly, performance on this task may rely on touch-typing abilities which are not easily transferable to number typing i.e., participants may have automatic, well learned behaviours for typing that are not apparent for pressing numbers on a response pad.

This also applies for the analysis plan. Including this as a co-variate in the model seems like you would capture variation in touch-typing proficiency rather than any baseline level of sequence execution.

After reading until the end, I see more justification for using this typing measure based on previous work. If this task is used, this justification needs to come earlier in the manuscript but my above concerns for this task still stand.

> This is all true, and that there are other behaviors that can contribute to better performance on the task. Alternative measures are possible, such as the well-known serial reaction time task, but we wanted to avoid tasks too similar to our main skill learning task so as to minimally interfere with that. We expect minimal interference with learning as typing English in our cohort (mostly college students) is a fully trained behavior. Using the typing test also saves the additional time required to explain an unfamiliar task to the subjects. Importantly, the pilot shows that a significant variance in learning is captured by this standard metric. We added the following clarification at the beginning of procedures:
>
> > The subject will first be asked to complete a baseline task to determine their baseline dexterous skill. We elected to use a simple typing test, as keyboard typing has been associated with better manual dexterity, visuomotor coordination, and motor speed[69–71]. This task is broadly recognizable as most adults are already experienced to some extent, making it easy to administer with minimal instruction. It is also sufficiently different from the main task, so as to minimize interference with the actual sequence learning task. Importantly, we found in a pilot study that there is a considerable positive correlation (r = 0.58) between baseline typing speed and baseline sequence learning performance (see below: “Predicting motor learning ability”).
>
> We also now refer to this task throughout the manuscript as “baseline typing speed”, instead of “baseline learning ability”, which could have caused some confusion.

A clearer description of the benefits of this neuro-navigation procedure would be useful for those not familiar with this set-up. For example, this study does not use subject-level anatomy (MRI structural scan), and so is relying on hotspots to find the hand representation on the motor cortex. There’s no problem with this, but a little more explanation may be useful. The main benefit of neuro-navigation comes from being able to monitor the position of the coil (make sure

it is consistent throughout the study in terms of angle and location) and being able to move the coil back to exactly the same place before and after training.

> Thank you for this suggestion and we agree that it would be beneficial to elaborate on these decisions. We have added the following:
>
> > Neuronavigation will ensure precise placement of the coil across multiple trials and sessions such that the same cortical location can be stimulated consistently[47,48]. For this purpose, individual anatomical magnetic resonance images (MRIs) are not required, as instead we will rely on "hotspotting" to locate the FDI representation in M1.

Why are 60 pulses used to assess the baseline MEP level? Standard protocols use ~20-30 MEPs. This saves time but is also important for participant safety. The number of pulses (from the very start of the hotspot/thresholding procedure) until the final pulse should be monitored and kept within safety operating procedures. Please state what these are for your institution/protocol. My reading is that you will have 60 MEPs for LH (Right FDI), 60 for RH (Left FDI), plus ~50 for "hotspotting" each side. How is this expected to interact with the particularly high doses of tDCS (up to 4 electrodes with 1.5mA)?

> We agree that the number of TMS trials is more than what is used in some MEP studies. Our IRB does not stipulate limits on the number of total MEP pulses. We believe that the total amount we use throughout the experiment is well within limits of tolerability and safety:
>
> > Where MEP protocols can range from 20 to 100 TMS pulses[4,67,77], we opted for 60 pulses as suggested for reliability in TEP measurements[66]. The total number of TMS trials falls within historical safety guidelines[78] and the amount used for motor mapping[79].
>
> We agree that interactions between the high number of TMS pulses and high tDCS intensity cannot be ruled out, even though we do not expect it. We have added this as a caveat:
>
> > We are not aware of reports on lasting effects of 0.2 Hz single-pulse TMS, despite decades of research using this modality. We therefore do not not expect any interactions with the sequence learning task nor tDCS. At the same time, we cannot in theory rule out such interaction effects.

1cm steps for locating the hotspot seems large. Smaller increments may be necessary to find the best hotspot. Any variability in this between participants (e.g due to not being accurate on the hotspot) will lead to significant variability in MEP.

> We agree, and in practice we do narrow down to smaller increments <5mm as we approach the hotspot. We have corrected the text to reflect what we are already doing:
>
> > The increments will be lowered to less than 5 mm as we approach the "hotspot", continuing until the MEP amplitude appears to stop increasing.

Please expand on the use of both left and right hemispheres. Presumably this is a control condition, but I don't see this included in the analysis plan or explicitly mentioned in the design of the experiment. Apologies if I have missed this.

> Thank you for pointing this out. The right-hand tests are all exploratory and we have added the following to our list of exploratory measures:
>
> > MEP and TEP effects on the right hand will be exploratory. Although they may serve as a within-subject control, there may be behavioral carryover effects, and the high tDCS intensities may cause parts of the left hemisphere to be stimulated.

Blinding of the participants is not expected. Some more detail on the implications of this is important. Will participants have any prior experience with tDCS? If not, what evidence is there that the ramping procedure will not be an effective sham?

> We apologize for the confusion from our statements around the blinding and have removed the phrase, "even though we do not expect blinding". To clarify, we do expect blinding within the individual participants. What we meant to say was:
>
> > We do not expect to achieve comparable levels of placebo effects from skin sensation alone, since participants under the sham condition will experience significantly lower levels. Nonetheless, as discussed below in the Pilot Data section, we do not expect a correlation between sensation and motor performance.
>
> We exclude subjects with prior experience of tDCS. Subjects are therefore expected to believe that they are being stimulated. We have now added:
>
> > Sham stimulation at 1 mA may not be effective in a within-subject design[82–85], but in this case the subjects will not have multiple stimulation conditions to compare across. Therefore, we expect all subjects to perceive some level of stimulation.
>
> And in the section on Subjects we write:
>
> > Subjects with prior experience of tDCS will be excluded from the study to facilitate blinding.

What is the rationale for using effect sizes in line with previous studies for H1 and then being more conservative for H2?

> We believe that the effect size we referenced in the table for H2 is exceptionally high, while some others have reported null effects on MEPs. We would like to estimate something in between with a more moderate effect size, and we do not expect to achieve the levels reported by Ahn and Frohlich. On the other hand, our estimate for H1 is based on our own data, collected using methods very similar to the current ones.

Will any data cleaning be performed on the MEP data. For example, if there is no clear sign of an MEP for a trial, will this be included or excluded? Will you check that participants are at rest (EMG signal) during the MEP recordings? Will trials with a excess tension be excluded?

> We only exclude trials where there is no clear MEP. We added the following clarification:
>
> > MEP amplitudes for each subject will be calculated by averaging the raw amplitudes across 60 EMG epochs triggered by the TMS pulses. The average will only include trials with a detectable biphasic MEP.
>
> On test data we currently use the following definition for biphasic pulse: We use the matlab findpeaks() function to detect peaks. To be included, an MEP has to consist of a positive peak with a minimum Peak Prominence and maximum Peak Width, followed by a negative valley with a minimum Peak Prominence and minimum Peak Width. The peak has to occur within a time window after the TMS trigger (approximately 20-50ms). There may be outliers that require refining these specific criteria.

More details of the current modelling would be useful. What was used to model current flow and density?

> Thank you for pointing this out. We have added more details of the current flow modeling:
>
> > Current flow modeling using ROAST[90] (realistic volumetric approach to simulate transcranial electric stimulation) was conducted a priori in consideration of safety. Simulations were run on the same 10 anatomical MRIs we used in our previous study to formulate the electrode montage[49]. Applying the 1.5 mA per electrode pair under our previous configuration for a total of 6 mA, we found that the maximal electric field achieved on the surface of the brain was approximately 1.8 V/m, which corresponds to an estimated current density of 0.23-0.50 A/m^2 in the brain[91].

Only later on is it mentioned that a very similar study has been run (Pilot data). Is this a replication study or does this build on this work in some way? Perhaps putting this “pilot” study into the introduction would be appropriate. It seems as though more emphasis should be put on this prior study and the fact that this is a replication (or part-replication?).

> Yes, one important aspect of this study is a partial replication of the previous work. Thank you for the suggestion and we now make this clearer in the Introduction:
>
> > Finally, this study may serve as a partial replication of our previous study at 4 mA where we also tested whether tDCS effects outlasted the period of stimulation and whether they were specific to the stimulated hemisphere and task. The experimental design therefore closely matches the previous study[49].

What is the maximum for the sensation rating score used in the pilot experiment? This would help to interpret graph 3B.

Thank you for pointing this out. We have now added the following details to the Pilot Data section (as well as the figure caption):

> 6 mA tDCS was well tolerated at a rating of around 4 on a visual analog scale from 0 to 10[81] (Fig. 4b)

Are there any checks in place to make sure that participants tolerate the stimulation to an acceptable level? Participants will, of course, be told that they can stop at any time, but given the potential for adverse effects with higher doses, will there be any additional measures taken?

Thank you for checking this. The measures below are taken, which are now listed in the Methods section. With the the description of tDCS procedures we write:

> The tDCS device constantly monitors impedance during stimulation to ensure that every channel is below 10 kΩ, thus reducing the likelihood of an adverse reaction on the scalp. Once the current reaches full intensity, we will ask the subject whether the sensation is acceptable, and if so, whether they would like to proceed to the task. Throughout the experiment the participant is reminded that they can ask to end the stimulation at any point, for both TMS and tDCS.

Additionally, we now include the following with our description of sensation reporting:

> The IRB protocol for TMS and tDCS includes an adverse event reporting form which asks the participant to report any of the following: headache, neck pain, scalp pain, tingling, burning sensation, skin redness, sleepiness, trouble concentrating, acute mood changes, and nausea/lightheadedness/dizziness. We will report these ratings as secondary safety outcomes in this study.

Assuming proper application of the conductive gel, we expect the current density below the electrodes to be tolerable and not cause adverse effects. We elaborate under the Pilot Data section:

> Additionally, the base area of the gel in the Soterix HD1 Holder is approximately 4.5 cm^2, through which a current of 1.5 mA would yield approximately 0.33 mA/cm^2 current density on the skin per electrode. Since this is below the upper tolerability limit of 0.5 mA/cm^2 commonly cited in iontophoresis literature[96], we expect this level of stimulation to be tolerable.

The analysis plan suggests that a few different null effects are predicted (e.g. a saturating effect; there will be no difference between the 4mA and 6mA conditions). Further investigation of this effect using Bayesian analysis may help to provide support in the event that there is a null finding. My impression of this work is that Bayesian Hierarchical regression modelling could be implemented in the same way as the linear models.

Thank you for this suggestion. We are not as confident in the MEP outcomes, but we believe we can be more thorough with the behavioral outcomes. We have updated the planned analysis for the behavioral results:

> Because we have prior data on behavior, we will apply a more rigorous analysis pipeline for this outcome (Fig. 3). If a significant effect is found in the initial linear model with intensity as a graded variable (and typing speed as a covariate), intensity would be interpreted as a monotonic effect, since we do not expect to resolve a significant difference between 4 and 6 mA. In the case where no significant effect is found at the first stage, we would test a follow-up linear model with intensity as a categorical variable. A significant finding at this second stage would be followed by a Tukey HSD pairwise comparison between 4 and 6mA. A significant difference between the two groups would be interpreted as a "reversing" effect, whereas a lack of a difference would be interpreted as a "saturating" effect, with linearity ruled out. A non-significant finding at the second stage would be interpreted as a lack of effect from tDCS intensity. Post-hoc analyses in the second and third stages will be Bonferroni-corrected accordingly. In the case of a null finding resulting in a saturating dose response or no effect, we will use Bayes factor analysis to measure the evidence in support of the corresponding null hypothesis using an established MATLAB toolbox[89].

The new Figure 3 visualizes the analysis pipeline:

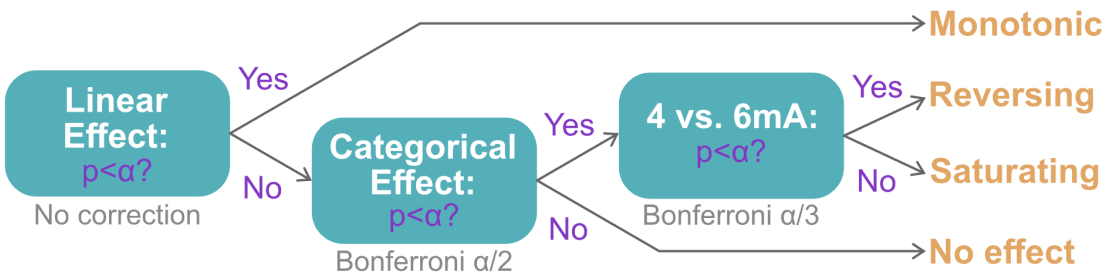

Figure 3. Analysis pipeline for primary behavioral outcome. An initial linear model at the first stage of analysis will use intensity as a graded variable and typing speed as a covariate. If no significant effect is found, a second stage analysis will use intensity as a categorical effect. If a significant effect is found, a third stage analysis will apply a Tukey HSD pairwise comparison between 4 and 6 mA. Second and third stage analyses will be Bonferroni-corrected accordingly, with $\alpha = 0.05$. The interpretations of the dose effect given these possible outcomes are "monotonic", "reversing", "saturating", and "no effect".