Revision round #4
Decision for round #4 : Revision needed
Invitation to revise Stage 1 RR

Thank you for the clarifications and modifications. I think that we are getting towards a Stage 1 plan that can be recommended for In Principle Acceptance, but that further clarifications are still necessary. Your responses to the previous comments have helped me pinpoint the outstanding issues. I have also discussed with members of the PCI board, who have provided helpful advice.

1) Your analysis plan is now more coherent, but there is still a problem in that you have identified that you intend to target the standards of evidence stipulated by Cortex, whilst at the same time you have stated that you will interpret your BF's with respect to the scheme of Lee & Wagenmakers (2013). The problem here is that these are not compatible schemes. If you are targeting Cortex as a journal, then you are signing up to treat BF 6 as your decision threshold. According to this scheme, BF values lower than 6 must be treated as inconclusive. You should therefore remove your intention to refer to the interpretative scheme of Lee & Wagenmakers, and clarify that 6 is your decision threshold for supporting a hypothesis (or 1/6 for supporting the null). Reflect these changes in your design table.

2) Your BFDA is predicated on an effect size of .35, but then you go on to say that you expect the effect to be smaller for 'four' and larger for 'eight'. Your sample size calculation must be informed by the smallest effect you are targeting, so you need to make it clear that .35 is your effect size of interest for the 'four' condition, in which you expect the effect to be weakest (assuming that this is the case).

3) You have stated that you will require both sub-hypotheses (a) and (b) to be supported to accept H1 and H2. In this context, it is not really relevant to say that you these hypotheses are "expected to be mainly driven by strong evidence for the Hypotheses 1a and 2a (related to number ‘eight’), while Hypotheses 1b and 2b (related to number ‘four’) will probably provide a weaker evidence (in comparison to 1a and 2a)." If your experiment is designed to be sufficiently sensitive for the weaker sub-hypothesis, then you should simply state that you require both sub-hypotheses to be supported in order to conclude in favour of the over-arching hypothesis.

These clarifications should be reflected by appropriate changes to all relevant parts of the manuscript.

I also note a couple of other places where I think changes in wording would be beneficial:

"In addition, there is a considerable difference in signal-to-noise ratio for fMRI vs. fNIRS with different spatial and temporal resolutions, which can mean statistical power from an fMRI study may not be directly compared to that of fNIRS." >> It may be better to say that standardised effect sizes (rather than 'statistical power') may not be directly comparable.

"Given that even 92 participants would be considered as a large sample for a neuroimaging study in preschoolers, if the BF for either of the hypotheses does not reach 10 6 (whether due to the effect’s extreme weakness or the common in the field deficit in neuroimaging data

quality in preschoolers) even after recruiting maximum feasible sample size, we will consider this inconclusive result as an important message to the field." >> It is a part of the review process that the reviewers have accepted the study as sufficiently interesting to be worth reporting, regardless of outcome, but it seems odd to state in the manuscript that you will consider an inconclusive result to be an important message for the field. This statement should therefore be deleted (although if you wish you can reiterate that a failure to meet the threshold of evidence for both sub-hyptheses will mean that the evidence is inconclusive for the overarching hypothesis).

by Robert McIntosh, 01 Jul 2024 17:02
Manuscript: https://osf.io/7vtdk?view_only=59385b6256b5492791f6882705c20424
version: 7

Dear Prof. McIntosh,
Thank you for providing us with helpful advice on wording in this round of the revision, we appreciate the time and effort that you and your colleagues dedicated to our study. We have incorporated all your suggestions. Those changes are listed here, as well as marked as tracked changes within the manuscript.

1) Analysis plan:
We have removed the interpretative scheme of Lee & Wagenmakers, as well as clarified that 6 is our decision threshold for hypotheses testing. These changes have been reflected in the analysis plan (page 20 of the manuscript) and the design table (page 27 of the manuscript), which now reads as follows:
“BF10 of 6 for the difference in the left parietal region in CP-knowers compared to subset-knowers will be considered a decision threshold for supporting the hypothesis and the BF10 of 1/6 for the difference in the left parietal region in CP-knowers compared to subset-knowers will be considered a decision threshold for supporting the null hypothesis – see Analysis plan section in paper. BF10 between 1/6 and 6 will mean that the evidence is inconclusive”

2 & 3) Effect size and sub-hypotheses:
Following your comments, we have clarified that .35 is the effect size of interest for the 'four' condition, as well as that we require sub-hypotheses to be supported in order to conclude in favor of the overarching hypotheses (please see page 21 of the manuscript):
“Note that both Hypotheses 1 and 2 will require support for their respective sub-hypotheses, 1a and 1b, and 2a and 2b, to be accepted. The study is designed to detect an effect size of .35, which corresponds to the expected effect in the 'four' condition, where we anticipate the weakest effect. This is because both CP- and subset-knowers understand the semantic meaning of the number word 'four' and are expected to show a parietal response; however, CP-knowers are expected to exhibit a stronger parietal response than subset-knowers due to their more advanced understanding. In contrast, we anticipate a larger group difference for the number word 'eight' because subset-knowers, unlike CP-knowers, do not yet comprehend the semantic meaning of the number word 'eight' and are not expected to show a strong parietal response. Thus, the sample size and design are sufficiently sensitive to detect the expected effects across both conditions.”

We have also made changes to the wording on pages 9 and 10 of the manuscript in response to the editor's suggestions.