Round #1
Author's Reply:
by Robert McIntosh, 14 May 2023 13:52
Manuscript: https://osf.io/7vtdk?view_only=59385b6256b5492791f6882705c20424 version 1
Invitation to revise Stage 1 RR

Two expert reviewers have now provided their scheduled review of the Stage 1 plan (Reviewer#1 has identified himself as Ed Hubbard). Both reviewers express some enthusiasm about the proposed study but also list some substantive (conceptual and methodological concerns) that should be addressed before IPA could be considered. IPA is not guaranteed but will depend upon the adequacy of the responses and revisions as assessed at a further round of review.

Ed Hubbard lists a number of concerns that all contribute towards a concern that the study may be in danger of Type II errors (effectively, that it is underpowered). In this regard, it may also be relevant to consider Reviewer#2's concern that no outcome-neutral quality checks have been proposed for this study, to establish the basic adequacy of the fNIRS setup and processing pathway to detect expected effects that do not bear on the main hypotheses. In addition, I would emphasise that the statement of hypotheses does not make it clear how the overall conclusions for each main hypothesis (1 and 2) will be informed by the combination of outcomes across the sub-hypotheses that have been stated. This logic should be made evident.

We look forward to seeing a revised version of the plan, along with responses to all of the reviewer comments, if you decide to take on this challenge.

Yours sincerely,

Rob McIntosh, PCI Recommender

Dear Prof. McIntosh,
Thank you for giving us the opportunity to submit the revised manuscript of our Stage 1 Registered Report. We appreciate the time and effort that you and the reviewers dedicated to providing constructive feedback on our manuscript that significantly improved our study. We have incorporated the suggestions made by you and the reviewers. Those changes are highlighted within the manuscript.
Before providing our in bold point-by-point responses to the reviewers' comments and concerns, we would also like to address the concerns that you raised as the editor.
The danger of underpowering the study:
After assessing the risk of underpowering the study, we decided to change the design and implement a paired t-test (comparing two experimental conditions) with covariates of age and CP status (CP-knowers and subset-knowers) instead of an independent two-sample t-test. The decision was made based on the following reasons:
i) Running a well-powered independent t-test with small effect size (given the novelty of the study) is not feasible. According to Brysbaert's suggestions (2019

https://doi.org/10.5334/joc.72), to run a two-sample t-test with the suggested effect size of 0.4 and BF > 10, one would need over 200 participants in total (after excluding outliers/incomplete data/very noisy records/etc). That, however, is a hardly achievable goal when collecting neuroimaging data in preschoolers, especially in a single-lab project. The paired t-test design, however, would require a smaller sample due to a doubled number of observations per participant (Brysbaert, 2019).
ii) Switching to this design allows us to increase the power of our study even more and minimise the chance of running into type II error. As suggested by many studies, running a paired t-test with covariates increases the power of the study without increasing the sample size due to explaining more sampling variance (e.g., Hedberg & Ayers, 2015 https://doi.org/10.1016/j.ssresearch.2014.12.004; Safarkhani & Moerbeek, 2013 https://doi.org/10.3102/1076998612461832; Wu & Gagnon-Bartsch, 2021 https://doi.org/10.3102/107699862094146).
Accordingly, we recalculated our sample size using the Bayes Factor Design Analysis (BFDA) with sequential n design (Schönbrodt & Wagenmakers, 2018 https://doi.org/10.3758/s13423-017-1230-y). Calculation of sample size involved running Monte Carlo simulations and relied on the BFDA package in R (R Core Team, 2022; Schönbrodt, 2016; see https://github.com/nicebread/BFDA). The simulations were specified for a paired t-test, with unlimited BF, d = 0.35, Nmin =70 , Nmax =110. The simulations of our design under H1 showed that the probability of our study to gain BF10 = 10 before reaching Nmax is 77.8%, with the average stopping at a total of N = 84 participants. This agrees with Visser et al. (2023 https://doi.org/10.1002/icd.2412) guidelines on sample size calculations, and is realistic following Brysbaert's suggestions (2019).

The lack of quality checks:
We would like to highlight established outcome-neutral quality checks for this study, to ensure the basic adequacy of the fNIRS setup and of the processing pathway. As mentioned in the preprocessing section of our manuscript, the channel quality will be assessed using the QT-NIRS toolbox using SCI threshold = 0.6, Q threshold = 0.5 and PSP threshold = 0.1 (Hernandez & Pollonini, 2020 https://doi.org/10.1364/BRAIN.2020.BM2C.5). This toolbox is probably the most established, measurable, transparent, and replicable preprocessing step in the field.

The unclear relationship between overall conclusion for the hypotheses and outcomes of the sub-hypotheses:
We apologise for not being clear in phrasing our hypotheses and sub-hypotheses. We made changes to our hypotheses in hopes of both simplification and a power increase. Namely, we decided to leave the hypotheses that involve the frontal area as exploratory for Stage 2. In the previous analysis model, we planned to take into account the fronto-parietal activation and connectivity to investigate whether bilateral/unilateral frontal area is relied on when the numbers are well known (as in the case for CP-knowers) vs. when they are only starting to be processed in parietal area (as in the case for subset-knowers). While we will investigate this question in our exploratory analysis, we limit our confirmatory hypotheses and analyses to parietal regions as the regions of interest. This would further reduce the risk of type II error. Here are the updated version of the hypotheses along with their sub-hypotheses:

Hypothesis 1: CP-knowers will exhibit higher left parietal activation, defined by increased HbO/decreased HHb, in both conditions relative to subset-knowers because they are more advanced in their conceptual knowledge about the meaning of number words.
This difference is expected in the contrast of deviant auditory number word 'eight' minus control (Hypothesis 1a), and in the contrast of deviant auditory number word 'four' minus control (Hypothesis 1b).

Hypothesis 2: CP-knowers will exhibit higher bilateral parietal functional connectivity, defined by functional co-activation between these two regions, in both conditions relative to subset-knowers because they have developed a more solid link between number words in the left parietal area and intuitive representations of quantity in the right parietal area.
This difference is expected in the contrast of deviant auditory number word 'eight' minus control (Hypothesis 2a) and the contrast of deviant auditory number word 'four' minus control (Hypothesis 2b).
Please note that in both hypotheses, we expect smaller group differences for number word 'four' than for number word 'eight'. Smaller group difference for number word 'four' is expected because while both CP- and subset-knowers understand number word 'four' and are expected to have a strong parietal response, CP-knowers are more advanced, so their parietal response will be even stronger as compared to subset-knowers. However, larger group difference for number word 'eight' is expected because subset-knowers do not understand semantic meaning of number word 'eight' yet, thus, they are not expected to have a strong parietal response. CP-knowers, however, have a much better understanding of number word 'eight', and are expected to have a much stronger parietal response when compared to subset-knowers.
In both cases, the main hypotheses, in order to be confirmed, need to be informed by the combination of outcomes across the sub-hypotheses. We hope that we have managed to bring more clarity to our work and are looking forward to hearing from you.

Reviews
Reviewed by Ed Hubbard, 13 May 2023 03:13

This is a preregistration study to study of the neural origins of number understanding in 3-4 year old children. The study team plans to use functional near infrared spectroscopy (fNIRS) to measure brain responses in parietal and frontal regions in children who either understand counting principles (CP-knowers) or who do not yet understand counting principles (subset-knowers). Knowledge of counting status will be assessed via the classic "give-a-number" task, and participants will be matched, as well as possible on other demographic and cognitive factors, especially non-verbal IQ and mean age.

Brain responses in these children will then be measured via fNIRS while the children engage in an adaptation design: Children will be presented with a repeated auditory number word ("six") and occasional deviants of "four" or "nine" (same ratio distance from the standard). For children who do not yet know the counting principles, the word "nine" is expected to be outside their semantic understanding of numbers, while “four is within their semantic range. For CP-knowers both numbers would be within their semantic range.

This study addresses an important and timely question and the study team is highly expert in both numerical cognition and fNIRS. Finally, the use of fNIRS makes it more feasible in the young (3-4 years old) children who would be tested here. The system the authors have chosen is a portable system (Brite, Artinis Medical Systems BV, The Netherlands). The use of fNIRS is an important choice, and it has several key advantages, specifically being more appropriate for kids in this age range due to its tolerance for motion, the ease of use, and lower cost. The use of a portable system will increase their ability to record data in various locations including preschools and other school buildings, but at the expense of having fewer optodes/channels. fNIRS in general has poorer spatial resolution than fMRI, and the limited number of channels could be an important limitation for multivariate analyses (see below).

The analysis plan seems appropriate (but, see below), and involves traditional univariate analyses of signal change (Oxy-Hb/HbO and Deoxy-Hb/HHb), functional connectivity analyses and (unspecified) multivariate analyses. Statistics will be carried out within a Bayesian framework, which will allow the authors to not only provide evidence in favor of differences, but also to measure the strength of evidence for the null hypothesis.

However, I have many concerns about the study as currently proposed, some theoretical, and some more methodological. As I see it, these concerns each make it more likely that the study team will fail to detect differences (Type II Errors), rather than increasing the possibility of spurious positive results (Type I Errors). I present these concerns here in the hopes that the study team will address these concerns prior to carrying out the study, and therefore increase their likelihood of success.

We thank Prof Ed Hubbard for the thorough and accurate overview of our proposed study and positive comments.
We also acknowledge the concerns regarding the multivariate analyses, as well as theoretical and methodological concerns. We have now majorly revised the study to address these concerns. Namely, we made changes to the procedure (fNIRS task and Give-a-Number task), and to the study design and sample size calculation to accommodate for the power of the study, the effect size and data collection feasibility. Please find below our thorough responses to each of the concerns.

CP-Knower Status = Semantic Understanding?

My first, conceptual/theoretical concern about this study is that the authors equate CP-knower status with semantic understanding of numbers. Although it is clear that young children know the count sequence in a rote manner prior to semantic understanding, it is not clear to me that successful performance on the give-a-number task is the only (or even the best) indication of semantic understanding of auditory number words. Success on the give-a-number task indexes not only semantic number knowledge but also executive function skills (maintaining task set, inhibiting the response to simply continue giving objects; Chu et al., 2019 https://doi.org/10.1016/j.jecp.2019.104668; Chen et al., 2022 https://doi.org/10.1111/bjdp.12439). Additionally, recent work has suggested that children may have “partial knowledge” (O’Rear et al., 2020 https://doi.org/10.1111/desc.12944) of number sequences, even prior to “full” success on the give-a-number task. These results suggest that children may have graded semantic representations of number words even prior to being coded as CP-knowers in the traditional analysis. For both of these reasons, semantic understanding that 9 is larger than 6 might be present, but weaker, even in children who are not yet CP-knowers. If so, we would predict that this partial semantic knowledge would lead to more similar activation patterns between the two groups. Concerns make it more likely to fail to detect differences between groups.

Response: We would like to thank the reviewer for pointing out these important issues. We agree that the Give-a-Number task involves some domain-general cognitive processes, but as the reviewer (as an expert in mathematical cognition) well knows, almost all numerical and mathematical tasks demand both domain-specific and domain-general processes. So, because of the following reasons, we decided to go with this task: i) We are not aware of any other cardinality task that does not involve domain-general cognitive processes, ii) this is the most established task in behavioural literature of cardinality assessment, iii) we will be able to cross compare our findings with other few existing neuroimaging studies of cardinality knowledge (e.g., Bugden et al., 2021 https://doi.org/10.1016/j.dcn.2021.101011, Pinhas et al., 2014 https://doi.org/10.1162/jocn_a_00631), iv) developing new tasks would probably need testing for reliability and validity in large-scale behavioural study first, which is out of our resources.

However, we acknowledge the outstanding debates about the limitations of the traditional categories of subset-knowers vs CP-knowers. In our revision, we would go even beyond one additional group of partial-knowers as the reviewer suggested. As recent literature suggested (O’Rear et al., 2020 https://doi.org/10.1111/desc.12944), we may need to discriminate within both subset knowers (e.g., 1-knowers vs 2-knowers, etc) and within CP-knowers (e.g., 5-knowers vs 6-knowers, etc.). We therefore plan to conduct additional exploratory analysis to look at the CP knowledge as a continuous variable, rather than a categorical variable, to test whether increased CP knowledge would be associated with increased brain activation in the left parietal region and bilateral parietal connectivity. As the editor previously suggested, we will include these exploratory analyses only in the Stage 2 of this Registered Report.

We have also decided to adjust the Give-a-Number task according to our study specifics. Namely, we switched to the titrated version of the task instead of non-titrated as previously planned. This change in the protocol seemed appropriate because of two reasons: i) titrated version is highly reliable and is in high degree of concordance with non-titrated (Marchard et al., 2022 https://doi.org/10.1016/j.cognition.2021.104998), ii) non-titrated method is commonly used in older preschoolers (e.g., between 4-5 years old), and often, as the only task in the experiment (e.g., Krajcsi, 2021 https://doi.org/10.1016/j.cogdev.2020.100968). Not only is our sample younger and has less tolerance for repetitive tasks, they also have to complete other tasks throughout the testing session, the most important of which is the fNIRS task that includes preparation as well. Based on our experience in previous studies, making the subset-knowers repeatedly struggle with the trials beyond their knowledge level will cause them to become tired and frustrated. As a consequence, they are much more likely to disengage, which will lead to either great loss of fNIRS data due to increased movement/non-compliance or complete rejection to do the fNIRS task in the first place. Overall, our preference for the titrated version

agrees with the Marchard et al. (2022 https://doi.org/10.1016/j.cognition.2021.104998) recommendation to favour titrated Give-a-Number when the experiment includes other tasks and participants are too young to be likely to score high in the Give-a-Number. Similarly to the classic titrated version of the Give-a-Number task, the task gradually increased the level of complexity and its stopping point was dependent solely on the child's responses. However, we did not want to use a linear structure of the titrated Give-a-Number task, to avoid some of the participants understanding the predictive nature of an increasing series (similarly to Krajcsi, 2021 https://doi.org/10.1016/j.cogdev.2020.100968). To keep the gradual, albeit not completely linear increase in task complexity, we split the task into three blocks: i) 1, 3, 2, ii) 6, 4, 5, iii) 9, 7, 8. Within each block, each number would be asked twice, and if the accuracy for that number is 50% (e.g., first response is wrong, but second is correct), the number would be asked the third time. If a child fails two or three numbers in the block (for example, gives correct answers for number 4, but not for 6 and 5), the next block will not be asked. For the main analysis, children who can correctly give the experimenter 1-4 objects, but not more, are grouped as subset-knowers, while children who can correctly give 5 and more objects are grouped as CP-knowers. For the exploratory analysis, children are defined as an N-knower, if they respond correctly on at least two out of the three times when N is requested.

This strategy allows us to score children the same way as in the titrated Give-a-Number, while i) avoiding attempts to predict the rest of the trials, ii) shortening the time spent on task, iii) being able to group children as CP-knowers and subset-knowers for the main analysis while also having data to look at the CP knowledge status as a continuous variable, as mentioned above, iv) recording any potentially meaningful inconsistent responses. Inconsistent responses are characterised by the child giving the correct response only once for a certain number. Since the procedure is titrated, inconsistent responses can only be given next to consistently correct responses (e.g., 1-knower can have inconsistent responses to number 2 or 3, but never to number 9). Based on our experience, this type of inconsistent responses may be due to accident, loss of interest or genuine attempt to give the correct answer (e.g., trying to count all three times but being correct only once due to failure to keep attention), or child is in the transition to the next level, but not there yet (i.e., partial knower). While this coding will not affect the conservative grouping strategy into either CP/subset-knowers, it will be taken into account in our exploratory analysis, as this might suggest that the child possesses partial knowledge about the number.

Adaptation Paradigm

I am very concerned about the design of the adaptation paradigm, in which only one deviant type is presented per block (that is, after adapting to "six" only "four" is presented for all the deviants in a block). Given the presence of only one deviant type per block, even young children might recognize this constancy, and would presumably pay less attention to the deviants during the course of each run. The authors argue that they have chosen this design to avoid task switching, which is difficult for children. However, there is no active task on the deviants (the only task is to detect occasional winks of the smiley face at fixation) so having different types of deviants does not introduce additional task switching demands. As adaptation effects reflect a mix of bottom-up and top-down processes (e.g., Summerfield et al., 2008 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2747248/; Larsson & Smith, 2012l https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3278317/) reducing the top-down attentional components that are present in other adaptation studies may reduce the ability to detect differences between the conditions. Concern makes it more likely to fail to detect differences between conditions.

Response: We acknowledge the concerns about the paradigm that have been highlighted. Additionally, we had the opportunity to present our Registered Report during the MCLS conference in June 2023 and receive further feedback. Accordingly, we updated the task as we explain below. We used block design because our sample is very young, with most of them not being able to tolerate fNIRS cap for even 10 minutes. Our experiment thus must be very short, and block design should be prioritised over ER to get a better signal-to-noise ratio in young children (Li et al., 2017 https://doi.org/10.1016/j.dcn.2016.07.002, Perlman et al., 2014 https://doi.org/10.1016/j.neuroimage.2013.04.057). Furthermore, our experiment must keep only one condition per run to additionally increase the signal-to-noise ratio in accordance with Rahimpour et al. (2023 https://doi.org/10.1038/s41598-023-33780-1) study, as data of young children has higher variance and lower signal to noise ratio (Gemignani & Gervain, 2021 https://doi.org/10.1016/j.dcn.2021.100943).

This block design paradigm of non-symbolic numerical adaptation task with only one deviant in each block has been successfully reported in fNIRS studies in infant (Edwards et al., 2016 https://doi.org/10.1111/desc.12333; Hyde et al., 2010 https://doi.org/10.1016/j.neuroimage.2010.06.030).

Accordingly, we kept the paradigm as block design with only one deviant per block because i) we wanted to consider the ratio rather than numerical distance as the literature has shown that the brain responses would be more sensitive to the former (Goffin & Ansari, 2019 https://doi.org/10.1111/mbe.12206). Therefore, we did not want to use different deviants for each condition (e.g., using 5, 6, 7, 8 for the difficult condition that only CP-knowers would respond to), ii) mixing our deviants in one block would mix up the conditions as we aim to distinguish between simple condition (with a deviant processed by both groups) and complex condition (with a deviant expected to be processed by only CP-konwers).

We have updated the paradigm as follows. Children will keep hearing the adaptation auditory number word during both experimental block and the inter-block rests, while in the experimental blocks the adaptation auditory number word will be interchanged with deviant auditory number word. While we can distinguish between the experimental blocks and the rest, it will not be understandable for children because they constantly hear the adaptation auditory number word with some deviants during different times (as they are not aware of the timing of those short blocks). We believe that the balance of the bottom-up and top-down attentional systems exists in our design (see above successful block design of numerical adaptation studies in infancy).

We also modified the adaptation and deviant auditory numbers words. Deviant auditory number words include the monosyllabic number words ‘four’ and ‘eight’, which differ by ½ and ¼ ratios from ‘two’ as the adaptation auditory number word (i.e., 2/4 and 2/8). The following ratio has been chosen following the adaptation signal recovery that suggests the increased ratio (when the ratio is calculated by dividing

the larger number by the smaller number) is linked to better signal recovery in adaptation task in the brain areas associated with symbolic processing (Ansari, 2008 https://doi.org/10.1038/nrn2334; Nieder & Dehaene, 2009 https://doi.org/10.1146/annurev.neuro.051508.135550; Vogel et al., 2017 https://doi.org/10.1016/j.neuroimage.2017.03.048). The only deviant auditory non-number word is 'rin'.

With regards to the attention problems in the task, the key advantage of this task is that it demands minimum engagement of children. As very crucial for this age, the task should either be non-monotonous, dynamic and interactive or the task should allow the child to pay minimum attention to it. However, another common confound in neuroimaging studies in preschoolers is frequent movement artefacts. The common solution to this is watching silent videos while presenting auditory stimuli. Many studies have commonly adapted this strategy, and comparisons of study conditions with/without a silent video have shown that the data quality is improved with using silent video (for a comparison study, see Welke & Vessel, 2022 https://doi.org/10.1016/j.neuroimage.2022.119218, and for other examples of experimental paradigm where children were watching silent videos, please see Hoyniak et al., 2018 https://doi.org/10.1111/desc.12608, Ni et al., 2021 https://doi.org/10.1016/j.heares.2021.108211, Orekhova et al., 2009 https://doi.org/10.1016/j.clinph.2008.12.034). Therefore, the other update in the task is presenting silent nonnumerical videos while children are listening to the auditory stimuli.

Please find a more detailed description in the revised methods section. This approach allows both better adaptation and increased number of deviants.

Sample Size/Power Analysis

I am concerned that the authors can only draw on one (quite different) study to estimate effect sizes for the power analysis. The paradigm that the authors intend to use is based on Vogel et al., 2017, while the effect size estimate comes from Holloway et al., 2013. Holloway et al. found an effect size (d) of approximately 0.73. However, that study used fMRI to measure adaptation in bilingual (Chinese-English) adults, while the proposed study will use fNIRS to measure adaptation only to auditory number words in 3-4 year old children. Additionally, although the paradigms are both adaptation paradigms and both use at least some auditory stimuli, there is very little else that is similar between the paradigms.

Response: We apologise for the misleading phrases in the manuscript. While we refer to Holloway et al. (2013 https://doi.org/10.1162/jocn_a_00323) study, because of the methodological and sample differences that we have mentioned in the manuscript (please see paragraph 2 of the power calculation section) and the reviewer pointed out as well, our sample size calculation was not based on this study. We did our initial calculation based on a medium effect size of 0.5 and not the rather large effect size of 0.73 (Holloway et al., 2013 https://doi.org/10.1162/jocn_a_00323). However, given the current lack of studies on the neuroimaging differences between subset-knowers and CP-knowers, it is sensible to decrease the effect size even lower as an extra-precaution step from medium to low (e.g., d = 0.35).

This raises several questions:

First, is it reasonable to expect similar effect sizes in children and adults? The general pattern of weaker responses in children (including many studies by the co-authors of this proposal; e.g., Ansari & Dhital, 2006 https://pubmed.ncbi.nlm.nih.gov/17069473/) suggest that the answer here would be no.

Response: As mentioned above, we did not rely on the adult study by Holloway et al. (2013 https://doi.org/10.1162/jocn_a_00323) given the difference in neuroimaging methods, age of our sample and the existing bias towards publishing only significant results. Please see above for details.

Second, it is reasonable to expect similar effect sizes for fMRI and fMRI? Again, the answer would appear to be no. In a particularly relevant study, Cui et al. (2011 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3021967/) collected fMRI and fNIRS data simultaneously in a wide range of tasks targeting the exact parietal-frontal networks that will be targeted here. On a positive note, Cui et al. found that activation measures were positively correlated across fMRI and fNIRS, especially when looking at the most strongly activated regions (voxels and channels, respectively). However, they found that the contrast-to-noise ratio for fNIRS was less than half that of fMRI when looking at all locations, and still 50% higher for fMRI when examining the most strongly activated locations (see Figure 10 of Cui et al., 2011).

Response: as mentioned above, we did not rely on the fMRI by Halloway et al. (2013). Please see above for details. We agree that the expectation of a similar effect size between fMRI and fNIRS is not reasonable, similarly to the expectation of a similar effect size between adults and children. We used a small-to-medium effect size to accommodate for these expectations. We would also like to note that our new small-to-medium effect size is smaller than effect size found in other neuroimaging studies of numeracy, e.g. the ERP study of the processing of number words in preschoolers by Pinhas et. al (2014 https://doi.org/10.1162/jocn_a_00631) that revealed large effect size for the same ROIs as in our study, and, as far as our knowledge goes, similar effect sizes are expected between EEG and fNIRS (in fact, the effect size can even be larger in fNIRS than EEG when recorded simultaneously, see our mathematical study in children, Soltanlou et al., 2018, https://doi.org/10.1038/s41598-018-20007-x).

Third, given the differences between the previous paradigms and the proposed paradigm, would we even expect the effect size to be similar between the two studies?

Taken together, it seems that the ostensibly conservative “medium” effect size (d = .5) that the authors consider for their power analysis might, in fact, still be wildly inflated, which would lead to a significant underestimate of the sample size needed for a reasonable power. Concern makes it more likely that the study would be underpowered, leading to a failure to detect differences between groups/conditions.

Response: as mentioned above, we agree that the effect size would not be similar between the two studies, and suggest decreasing the effect size to d = 0.35 as a sensible measure. Furthermore, after assessing the risk of underpowering the study, we decided to change the design and implement a paired t-test (comparing two

experimental conditions) with covariates of age and CP status (CP-knowers and subset-knowers) instead of an independent two-sample t-test. The decision was made based on the following reasons:
i) Running a well-powered independent t-test with small effect size (given the novelty of the study) is not feasible. According to Brysbaert's suggestions (2019 https://doi.org/10.5334/joc.72), to run a two-sample t-test with the suggested effect size of 0.4 and BF > 10, one would need over 200 participants in total (after excluding outliers/incomplete data/very noisy records/etc). That, however, is a hardly achievable goal when collecting neuroimaging data in preschoolers, especially in a single-lab project. The paired t-test design, however, would require a smaller sample due to a doubled number of observations per participant (Brysbaert, 2019).
ii) Switching to this design allows us to increase the power of our study even more and minimise the chance of running into type II error. As suggested by many studies, running a paired t-test with covariates increases the power of the study without increasing the sample size due to explaining more sampling variance (e.g., Hedberg & Ayers, 2015 https://doi.org/10.1016/j.ssresearch.2014.12.004; Safarkhani & Moerbeek, 2013 https://doi.org/10.3102/1076998612461832; Wu & Gagnon-Bartsch, 2021 https://doi.org/10.3102/107699862094146).
Accordingly, we recalculated our sample size using the Bayes Factor Design Analysis (BFDA) with sequential n design (Schönbrodt & Wagenmakers, 2018 https://doi.org/10.3758/s13423-017-1230-y). Calculation of sample size involved running Monte Carlo simulations and relied on the BFDA package in R (R Core Team, 2022; Schönbrodt, 2016; see https://github.com/nicebread/BFDA). The simulations were specified for a paired t-test, with unlimited BF, d = 0.35, Nmin =70 , Nmax =110. The simulations of our design under H1 showed that the probability of our study to gain BF10 = 10 before reaching Nmax is 77.8%, with the average stopping at a total of N = 84 participants. This agrees with Visser et al. (2023 https://doi.org/10.1002/icd.2412) guidelines on sample size calculations, and is realistic following Brysbaert's suggestions (2019).
We would also like to mention that this sample is several times larger than the existing literature in early development. Following our recent systematic review (Soltanlou, Patil, Nemati, Ansari, in preparation), there are a total of 25 neuroimaging studies of numerical cognition in preschoolers, and almost all of them have a sample of below 30 children. Additionally, considering our resources, it will not be possible for us (and probably most of other developmental labs) to calculate the sample based on a small effect size that will demand over 200 children, and might be more suitable for behavioural studies in adults. This might be possible in a multi-lab project, which is still quite difficult due to the methodology (e.g., the recording devices need to be the same, etc).

Predictions/Framework
Overall, I found the presentation of the specific empirical predictions to be confusing and poorly motivated. The key concern is that, under certain circumstances, increased signal is associated with greater skill/more mature performance, while under other circumstances, increased signal is associated with poorer skill/less mature performance. Although these predictions are sometimes in opposite directions for different regions (frontal vs. parietal/right vs. left hemisphere) the integration of all these moving parts is lacking. For example, the review of the functional connectivity fMRI literature on p.6 seems contradictory:

Emerson & Cantlon (2011) -> greater skill associated with greater fc in frontoparietal networks

Hyde (2021) -> younger children show greater frontoparietal connectivity, and then shifts to parietal (?)

Perhaps this just explained in a confusing way, and the authors need to present diagrams showing the developmental model that leads to their specific empirical predictions more explicitly. Concern makes it more difficult to interpret any differences that are observed.

Response: we apologise for the confusing statements of our hypotheses. We agree that the visualisation will make them more comprehensive, so we have added the diagram on page 8. Please note that the diagram does not reflect the predictions regarding frontal area, as in the revised manuscript, we concentrate on bilateral parietal regions as our regions of interest and will report the findings of the prefrontal regions as exploratory analyses in Stage 2. Here are our confirmatory hypotheses:
Hypothesis 1: CP-knowers will exhibit higher left parietal activation, defined by increased HbO/decreased HHb, in both conditions relative to subset-knowers because they are more advanced in their conceptual knowledge about the meaning of number words.
This difference is expected in the contrast of deviant auditory number word 'eight' minus control (Hypothesis 1a), and in the contrast of deviant auditory number word 'four' minus control (Hypothesis 1b).

Hypothesis 2: CP-knowers will exhibit higher bilateral parietal functional connectivity, defined by functional co-activation between these two regions, in both conditions relative to subset-knowers because they have developed a more solid link between number words in the left parietal area and intuitive representations of quantity in the right parietal area.
This difference is expected in the contrast of deviant auditory number word 'eight' minus control (Hypothesis 2a) and the contrast of deviant auditory number word 'four' minus control (Hypothesis 2b).
Please note that in both hypotheses, we expect smaller group differences for number word 'four' than for number word 'eight'. Smaller group difference for number word 'four' is expected because while both CP- and subset-knowers understand number word 'four' and are expected to have a strong parietal response, CP-knowers are more advanced, so their parietal response will be even stronger as compared to subset-knowers. However, larger group difference for number word 'eight' is expected because subset-knowers do not understand semantic meaning of number word 'eight' yet, thus, they are not expected to have a strong parietal response. CP-knowers, however, have a much better understanding of number word 'eight', and are expected to have a much stronger parietal response when compared to subset-knowers.

Please find below the diagrams added to the manuscript:

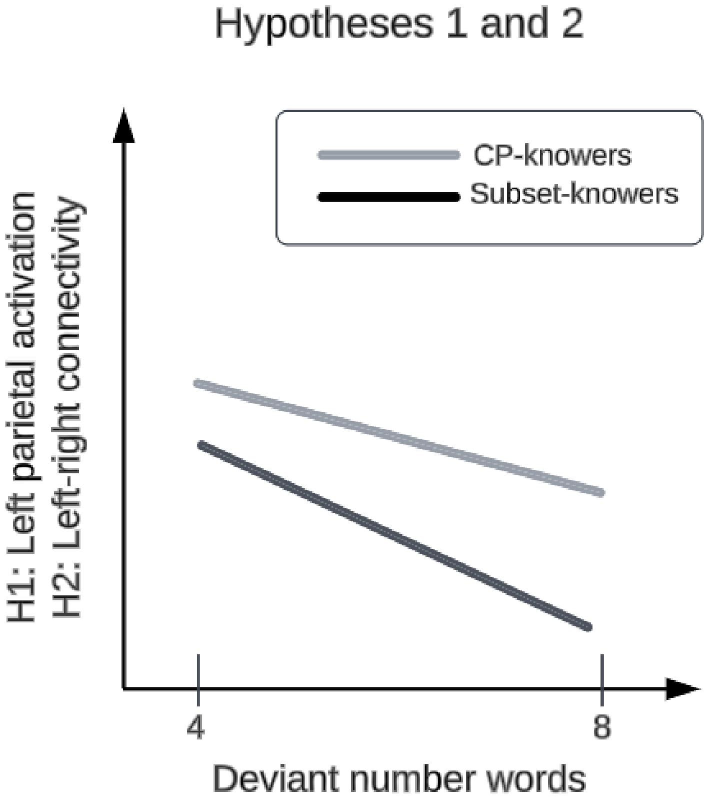

Figure 1: conceptual visualisation of the hypotheses for this study. According to Hypothesis 1, CP-knowers will exhibit higher left parietal activation, defined by increased HbO/decreased HHb, in both conditions relative to subset-knowers because they are more advanced in their conceptual knowledge about the meaning of number words. Similarly, according to Hypothesis 2, CP-knowers will exhibit higher bilateral parietal functional connectivity, defined by functional co-activation between these two regions. Please note that in both hypotheses, we expect smaller group differences for number word 'four' than for number word 'eight'. Smaller group difference for number word 'four' is expected because while both CP- and subset-knowers understand number word 'four' and are expected to have a strong parietal response, CP-knowers are more advanced, so their parietal response will be stronger as compared to subset-knowers. However, larger group difference for number word 'eight' is expected because subset-knowers do not understand number word 'eight' yet, and are thus not expected to have a strong parietal response. CP-knowers, however, have a much better understanding of number word 'eight', and are expected to have a strong parietal response. Please also note that the lines' angle of inclination serves only illustrative purposes, namely, showing which group and which condition is suggested to have highest activation (Hypothesis 1) or strongest connectivity (Hypothesis 2).

Finally, we would like to mention another update in the manuscript that we did not have a chance to mention above. As we have access to both English and Russian speaking children, we invite the children who speak English or Russian as their primary language. In the case where the child speaks both languages equally fluently, the language of study will be defined based on the child's preferences. Note that English and Russian languages have similar grammar structure regarding numerals, and native speakers of both languages develop numeracy skills within the same timeline (Sarnecka et al., 2007 https://doi.org/10.1016/j.cogpsych.2006.09.001), hence, the language will not be a grouping factor in the analyses.

We updated the inclusion and exclusion criteria, that are read now as follows: "Inclusion criteria consist of (i) age between 2 years 9 months and 4 years 3 months at the time of measurement, (ii) no history or diagnosis of neuropsychological impairment including developmental disorder, (iii) no chronic disease such as diabetes mellitus, renal failure, and high blood pressure, (iv) no use of particular medication at the time of measurement, (v) normal or corrected-to-normal vision and hearing, (vi) ability to speak and understand English and/or Russian, (vii) a minimum of one clean channel per ROI and per participant for inclusion. Exclusion criteria consist of (i) non-compliance with the wearing fNIRS cap (declining to wear the cap completely), (ii) non-completion of the Give-a-Number task, (iii) extreme contamination of fNIRS data (rejection of all channels during preprocessing), (iv) experimental error during Give-a-Number task and fNIRS task".

Reviewed by anonymous reviewer, 05 May 2023 15:40
The study proposed by Ivanova at al. aims to tackle an important question in the area of cognitive development, namely children's acquisition of cardinality principle knowledge.

The manuscript is well written and include enough methodological details to allow the replication of the experiment and analysis. The proposed study is feasible, and its rationale is supported by the needs to fill some gaps in developmental literature as summarized in the introduction. The task choice and the age range of the children is particularly very well explained and justified. The experimental procedure and the analysis that the authors intend to pursue are well described in the manuscript. However, I believe that additional clarifications are needed, particularly concerning some of the methodological details of the fNIRS experiment, that would improve the quality and structure of the manuscript and would make it easier to follow and understand. My comments and suggestions are listed below.

1. Unless I misinterpreted the text, the authors made hypotheses only on HbO2 changes. Why are no predictions concerning deoxyHb? I think the hypothesis needs clarifying as the best practice would be to include results for both HbO2 and HHb (even if not significant).

**Response: We appreciate the reviewer's feedback regarding our hypotheses. We followed this suggestion and the recent recommendations for fNIRS studies (Yucel et al., 2020 https://doi.org/10.1117/1.NPh.8.1.012101) to report both HbO and HHb to provide a complete picture of data. Our revised hypotheses are now read as :
Hypothesis 1: CP-knowers will exhibit higher left parietal activation, defined by increased HbO/decreased HHb, in both conditions relative to subset-knowers because they are more advanced in their conceptual knowledge about the meaning of number words.
This difference is expected in the contrast of deviant auditory number word 'eight' minus control (Hypothesis 1a), and in the contrast of deviant auditory number word 'four' minus control (Hypothesis 1b).**

Hypothesis 2: CP-knowers will exhibit higher bilateral parietal functional connectivity, defined by functional co-activation between these two regions, in both conditions relative to subset-knowers because they have developed a more solid link between

number words in the left parietal area and intuitive representations of quantity in the right parietal area.
This difference is expected in the contrast of deviant auditory number word 'eight' minus control (Hypothesis 2a).
the contrast of deviant auditory number word 'four' minus control (Hypothesis 2b).

2. Relatedly, how will the authors deal with situations where there is a positive result for oxy or deoxy but not both for a given ROI?

Response: The reviewer is raising a valid concern that theoretically, we expect simultaneous increased HbO and decreased HHb would neurophysiologically represent brain activation (Pinti et al., 2020 https://doi.org/10.1111/nyas.13948; Sholkmann et al., 2014 https://doi.org/10.1016/j.neuroimage.2013.05.004). However, it is quite often that we observe significant changes in one but not in the other (e.g., Rahimpour et al., 2023 https://doi.org/10.1038/s41598-023-33780-1). This is mostly the case of HHb which is ignored as its changes are much smaller than HbO (Sholkmann et al., 2014 https://doi.org/10.1016/j.neuroimage.2013.05.004). For instance, it is quite common that significantly increased HbO is observed with non-significant changes in HHb. We therefore suggest that including such channels is acceptable. However, the signals should be excluded when they behave either very similarly (i.e., both significantly increase or decrease) or completely opposite by mirroring each other, typically caused by extreme motion artefacts.

3. I think quality checks are missing from the hypothesis and data analysis plan. The authors should include that they expect to find a significant difference in the hemodynamic response between stimuli and silence (which I believe is their baseline condition) over the parietal and frontal arrays. The results would show a significant increase in HbO2 and/or significant decrease in HHb in response to stimuli compared to baseline.

Response: We thank the reviewer for highlighting this point. Indeed, we will calculate GLM for each condition vs rest, and we expect to see parietal activation during the two numerical conditions vs rest. Additionally, we are contrasting it with the control condition, as mentioned in our manuscript. That is done in order to ensure that the brain activity we are aiming at is not only related to domain-general attention; as previous research long suggests, parietal area does show activations for bottom-up attentional capture, but these activations are dissociated from specific parietal response to numeracy (Ciaramelli & Moscovitch, 2020 https://doi.org/10.1016/j.neuropsychologia.2020.107551; Downar et al., 2001 https://doi.org/10.1006/nimg.2001.0946; Edwards, Wagner, Hyde, 2015 https://doi.org/10.1111/desc.12333), which is a possibility in an adaptation paradigm. We would also like to note that we have altered the experimental paradigm, which we discuss in the experimental paradigm in our response to the next comment as reporting more details seems to be more suitable there. Please also find a detailed description in the revised methods section as well.
We have also altered the hypotheses: we decided to leave the hypotheses that involve the frontal area as exploratory for Stage 2. In the previous analysis model, we planned to take into account the fronto-parietal activation and connectivity to investigate whether bilateral/unilateral frontal area is relied on when the numbers are well known

(as in the case for CP-knowers) vs. when they are only starting to be processed in parietal area (as in the case for subset-knowers). While we will investigate this question in our exploratory analysis, we limit our confirmatory hypotheses and analyses to parietal regions as the regions of interest. This would further reduce the risk of type II error.
Please find updated hypotheses above, in oure response to question 1.

4. I am not sure the fNIRS experimental paradigm is clearly presented. I think this might be due to terminology. Could the authors clarify that what they are contrasting is the hemodynamic response to the block (33.6 sec) vs jittered inter-block interval (mean of 16 sec)?

Response: As mentioned above and in our manuscript, we will be contrasting each condition vs rest.
Please note that updated the task after we had the opportunity to present our RR during the MCLS conference in June 2023 and received further feedback. Accordingly, we have updated the paradigm as follows. Children will keep hearing the adaptation auditory number word during both experimental block and the inter-block rests, while in the experimental blocks the adaptation auditory number word will be interchanged with deviant auditory number word. We also modified the adaptation and deviant auditory numbers words, which now consist of the monosyllabic number words ‘four’ and ‘eight’, which differ by ½ and ¼ ratios from ‘two’ as the adaptation auditory number word (i.e., 2/4 and 2/8). The following ratio has been chosen following the adaptation signal recovery that suggests the increased ratio (when the ratio is calculated by dividing the larger number by the smaller number) is linked to better signal recovery in adaptation task in the brain areas associated with symbolic processing (Ansari, 2008 https://doi.org/10.1038/nrn2334; Nieder & Dehaene, 2009 https://doi.org/10.1146/annurev.neuro.051508.135550; Vogel et al., 2017 https://doi.org/10.1016/j.neuroimage.2017.03.048). The only deviant auditory non-number word is ‘rin’.
Please find a more detailed description in the revised methods section.

5. Do the authors expect significant differences in brain hemodynamics to the different comparisons in particular channels or across the whole fNIRS probes (within their 4 ROIs)?

Response: This is a very good question. While we hypothesised group differences in the defined ROIs, we may not observe those differences in all the channels. Following our previous fNIRS studies in children, this is because of both anatomical reasons (e.g., head size and shape differences) and technical reasons (e.g., noisy channels and light penetration). We therefore will choose the highest beta coefficient of HbO (and its corresponding HHb, if it meets our inclusion criteria of negative correlation with HbO, as explained above) for each condition and each ROI for the analysis. The highest coefficient of HbO for each condition on each of the regions of interest (ROIs; bilateral parietal for main hypotheses and frontal regions for exploratory analysis) will be used.

6. Procedure: are the authors planning to counterbalance the order of presentation of the tasks/test (IQ, verbal counting, etc)? If not, why so?

Response: As mentioned in the methods section of our manuscript, we have counterbalanced the order of runs in the fNIRS task using a Latin square design across participants. However, the order of behavioural tasks will be adjusted individually for every child based on their willingness, readiness and other factors. This decision is motivated by the unpredictable nature of children in this age and is a common practice in preschooler studies (Fishburn et al., 2019, https://doi.org10.1016/j.neuroimage.2018.09.023). While some may feel more shy in the beginning, others will have greater energy at the start, and become tired and unfocused over time. Pushing a child to perform a certain task first will likely result in participant's discomfort and the consequential loss of mental resources to continue with other tasks, which would lead to a huge unnecessary drop out. The comfort of the participants is one of our utmost priorities. This strategy would most likely cancel out any order effect, but please note that it is only about our behavioural tasks which are mainly used to understand the characteristics of the sample, and do not directly affect our hypotheses.

7. I am not clear about the exclusion criteria for the study. For example, how will you ensure task compliance during the fNIRS experiment? Will you manually exclude trials in which the children are not attending to the smiley face and/or due to external (e.g. parental) interference? Importantly, how are you planning to monitor children's attentiveness is not clear to me. In page 12 it is mentioned that this will be done by using speakers to produce the sounds, but it's unclear how the use of speakers can ensure stimuli attendance. Additionally, will you further analyse the fNIRS data from participants with missing data points on the behavioural tests/tasks? What about signal quality problems (e.g. do you have an objective criterion for the % of channels being excluded)? I also assume another standard exclusion criteria pertains to experimental error. Unless I missed this information, please include a list of exclusion criteria in the manuscript.

Response: Thank you for raising these important questions. In order to minimise common difficulties with task compliance specific to this age group (e.g., reluctance to wear a cap, boredom from a monotonous task, increased movement, see Perlman et al., 2014, https://doi.org10.1016/j.neuroimage.2013.04.057), we decided to change the procedure: instead of having to react to smiley faces, participants will be watching a silent cartoon of their choice, while passively listening to the auditory stimuli.
We would also like to avoid manual trial exclusion of the data, as that will increase the subjectivity of the data preprocessing and reduce transparency and replicability of data trimming. Instead, we will be excluding data based on the results of quality checks (QT-NIRS, Hernandez & Pollonini, 2020 https://doi.org/10.1364/BRAIN.2020.BM2C.5), in order to keep the same level of consistency across all preprocessing steps.
We apologise for being unclear about the use of speakers to control attentiveness to stimuli. The reviewer is correct that the use of speakers alone cannot ensure attentiveness during the experiment. While the participant's response towards changes in audio (pause in between runs, appearance of deviant auditory words) can give a hint about their attention level, the task will not require active listening or engagement now. As mentioned above, the participants will be watching silent cartoons throughout the task, which should keep them concentrated.

We also thank you for highlighting the need for clear exclusion criteria. We will analyse the fNIRS data of participants with incomplete behavioural data as long as they have done the Give-a-Number task, as this task is considered crucial for our purposes. Without this task, we can not understand the association between cardinality knowledge and its neurocognitive mechanisms. Therefore, to be included in the initial analysis, the participant must complete the fNIRS task and Give-a-Number task. And as rightfully noticed by the reviewer, completion of these two tasks depends not only on the participant's compliance, but also on the lack of experimental error. As we have made changes to the manuscript based on your comment, please find the added excerpt of the updated inclusion and exclusion criteria here:
"Inclusion criteria consist of (i) age between 2 years 9 months and 4 years 3 months at the time of measurement, (ii) no history or diagnosis of neuropsychological impairment including developmental disorder, (iii) no chronic disease such as diabetes mellitus, renal failure, and high blood pressure, (iv) no use of particular medication at the time of measurement, (v) normal or corrected-to-normal vision and hearing, (vi) ability to speak and understand English and/or Russian, (vii) a minimum of one clean channel per ROI and per participant for inclusion. Exclusion criteria consist of (i) non-compliance with the wearing fNIRS cap (declining to wear the cap completely), (ii) non-completion of the Give-a-Number task, (iii) extreme contamination of fNIRS data (rejection of all channels during preprocessing), (iv) experimental error during Give-a-Number task and fNIRS task". As we have access to both English and Russian speaking children, we invite the children who speak English or Russian as their primary language. In the case where the child speaks both languages equally fluently, the language of study will be defined based on the child's preferences. Note that English and Russian languages have similar grammar structure regarding numerals, and native speakers of both languages develop numeracy skills within the same timeline (Sarnecka et al., 2007 https://doi.org/10.1016/j.cogpsych.2006.09.001), hence, the language will not be a grouping factor in the analyses.
Finally, with regards to the signal quality problems, we are interested in having at least one channel per each ROI in the parietal area for the main hypotheses, and at least one channel per each ROI in the frontal and parietal areas for the exploratory analysis. This decision is informed by the expectation that collecting neuroimaging data in young children has many challenges that require additional flexibility with data analysis: i) We expect that the sample of our age will periodically demonstrate non-compliance with the procedure (Quiñones‐Camacho et al., 2020, https://doi.org10.1111/jcpp.13165), for example trying to take the cap off, crying or moving, which will make the data severely contaminated. And that, in return, will likely result in having a very small number of participants accepted for the statistical analysis, which will cause a decrease in power, effect size and validity of the results. ii) At this age, the differences in head's shape and size are large and visible will inadvertently influence the position of the channels (despite using standard 10-10 system caps). Thus, in each participant, one channel may be closer to the actual activation region within ROI than another.

8. Relatedly, what's the minimum number of trials required to carry out the GLM analyses?

Response: For guidance on the minimum number of trials for the GLM analyses, we relied on the study by Filippetti et al. (2023, https://doi.org/10.1016/j.neuroimage.2022.119756) that compared the feasibility of applying GLM and multivariate pattern analysis (MVPA) based approaches to infant fNIRS data collected from a block design study. According to the study, a total minimum number of 6 long trials of 8-10 seconds (3 trials per condition) is required to carry out GLM analyses, and a total of 16 trials (8 per condition) would suffice MVPA. Other fNIRS infant studies have used the engagement with the minimum 6 trials per condition as an exclusion criteria as well (Lloyd-Fox et al., 2014 https://doi.org/10.1117/1.NPh.1.2.025006, Southgate et al., 2014 https://doi.org/10.1016/j.neuroimage.2013.08.043). That being said, as of now there are currently no rule of thumb on number of blocks for auditory tasks in fNIRS research (Zhang et al., 2022 https://doi.org/10.1016/j.heares.2022.108593), with most of the researchers utilising having 3 or 4 blocks per condition (Sutoko et al., 2018 https://doi.org/10.1117/1.NPh.5.4.045001).With regards to the length of the blocks, it is suggested that for auditory paradigms, the length of 15 seconds is the most optimal one (Zhang et al., 2022 https://doi.org/10.1016/j.heares.2022.108593). In our case, each condition has 40 trials that make 4 blocks of approximately 16 seconds, thus leaving enough room for both excluding noisy trials and running GLM.

9. Are the authors planning to counterbalance gender in their sample?

Response: This is a valuable question, but we may not be able to counterbalance our sample by gender for two reasons. First, previous neuroimaging studies of numerical cognition in preschoolers did not observe gender differences (e.g., Kersey, et al., 2017 https://doi.org/10.1038/s41539-018-0028-7), hence we do not expect this difference in our study and do not have a ground for any related confirmatory hypothesis. Secondly, considering the challenges for sample size calculation and feasibility, we do not wish to introduce a new factor in our analysis. The resulting increase in sample size would directly influence the practicality, as we cannot restrict ourselves further. However, we will be able to run an exploratory analysis in Stage 2 to investigate potential gender differences if we have an adequate number of participants per gender.

10. I am not sure I fully understand the approach of using the highest coefficient of HbO2 for each condition and on each of the 4 ROIs. Could the authors clarify the rationale behind their choice? I would have thought that with GLM you can obtain beta parameters for each of the regressors and for each child, which can then be used to calculate a contrast between the conditions of interest for each infant.

Response: As the reviewer rightly explained, we will first obtain beta coefficients for HbO and HHb for each channel (that successfully passes the quality check using qt-nirs), each condition and each child. Out of them, we will choose the highest beta coefficient of HbO for each condition and each ROI. This method has been used within fNIRS literature (Chen et al., 2015 https://doi.org/10.1007/s10548-015-0424-8, Levin et al., 2022 https://doi.org/10.3390/app122312063). We have explained this reasoning in a response to question 7, by describing challenges associated with data collection in young children and how it can affect the data quality.

11. Probe locations, co-registration: A major weakness in the current proposal has to do with determining the spatial location of the probe. The authors propose to place their probes bilaterally over the parietal and frontal lobes and these locations are well-motivated by the literature. Do the authors intend to engage in a co-registration procedure themselves or somehow use the information from previous work? If the authors do not plan to do co-registration themselves, it's not clear how they will determine that the probe will be placed in particular cortical regions or even consistently across the children. This is both extremely important for data quality and interpretation and not trivial if a very specific protocol is not put in place. I think at minimum, the authors need a very clear procedure for placing the cap such that the channels can be localised to, say, 10-10 locations on the scalp and then a method of determining for each child whether the cap adhered to that protocol.

Response: We would like to thank our reviewer for raising such an important issue. For spatial registration, we relied both on the information from our previous fNIRS studies in children (e.g., Artemenko et al., 2018 https://doi.org/10.1186/s12993-018-0137-8; Soltanlou et al., 2017 https://doi.org/10.1038/s41598-018-20007-x, 2019 https://doi.org/10.1111/mbe.12225, 2022 https://doi.org/10.1007/s00429-022-02470-5), and on the designated software (fOLD software (Zimeo Morais et al., 2018 https://doi.org/10.1038/s41598-018-21716-z) and the template constructor in-built in OxySoft a software for fNIRS (Artinis Medical Systems BV, Netherlands)).
Furthermore, both head caps sizes (both produced by Artinis Medical Systems BV, Netherlands) we used (50 cm and 52 cm) had printed optode marks and 10-10 system indications, which allowed us to ensure appropriate location of the optodes on each of the caps. The caps have been placed using surface anatomical landmarks. The decision to follow this protocol has been informed by previous fNIRS studies done in preschoolers (e.g., Perlman et al., 2014, https://doi.org/10.1016/j.neuroimage.2013.04.0578). Please also see the excerpt from our manuscript, page 18 where we list the corresponding channels for our emitters: "The left- and right-hemisphere emitters will be respectively placed on P1, P3, P5, CP3, CP5 and P2, P4, P6, CP4, CP6 for measuring the left and the right parietal regions respectively, and on FC1, FC3, F3, F5 and FC2, FC4, F4, F6 for measuring the left and the right frontal regions respectively, following the international 10-10 system (Chatrian et al., 1985 https://doi.org/10.1080/00029238.1985.11080163)"