Thank you for the opportunity to review the manuscript for the Stage 1 Registered Report titled “Cerebral lateralization of writing in students at risk of dyslexia using functional Transcranial Doppler ultrasonography”. This is study has the potential to test whether children (aged 7-9 years) at-risk for developmental dyslexia are less lateralised for writing than their typically developing peers, but I have identified multiple issues with the manuscript and the study in its current form. None of these should be impossible to address, but they will require considerable rewriting (mainly of the introduction) and reconsideration of the analyses proposed.

Major issues

1. The introduction lacks focus, accuracy and depth in its discussion of relevant literature.
	1. In terms of focus, the authors currently discuss a lack of lateralisation for oral language in dyslexia in a rather hap-hazard way, followed by a minimal discussion of empirical work that has investigated lateralisation of written language production. The introduction should discuss the following in a detailed, but focussed way: a) evidence for writing difficulties in children with dyslexia. Currently, the authors discuss similarities and differences between dyslexia and dysgraphia to considerable extent. However, this discussion does not seem to feed into the study design as group classification will exclusively be based on reading performance. Therefore, it seems more appropriate to focus in the introduction on writing difficulties in dyslexia (and leave a discussion of dysgraphia for the discussion); b) evidence for cerebral lateralisation for written language production in typical readers focussing on fMRI and fTCD. This should include a discussion of lateralisation of motor activity associated with writing and how this is (or is not) dissociable from cerebral lateralisation for written language production. A discussion of how either/both are related to hand preference also seems appropriate; c) evidence for cerebral lateralisation for language production (assuming that evidence for cerebral lateralisation for written language production in dyslexia does not exist). In this discussion it would be helpful to be as specific as possible with regard to the types of studies that are discussed and focus on language production (rather than reading/language comprehension/speech processing); d) an overview of the present study and its hypotheses. Please note that the introduction in its current form does not provide the reader with sufficient background to understand why the authors hypothesise that writing competence will not be correlated with cerebral lateralization of the linguistic component of written language (H2; p11, l13-15).
	2. With regard to accuracy and depth, the introduction in its current form lacks detailed discussion of relevant empirical work and seems, in places, inaccurate in the references that are cited to support certain statements. See below for examples of this.
		1. p3, l8-12: The authors state that atypical lateralisation for oral language is a common trait in individuals with dyslexia or dysgraphia. This is presented as if this is a well-established fact. The references cited don’t justify the certainty with which this is presented. Araújo et al. (2012) report an ERP study in which a left-lateralised response for one component was only observed in the controls. However, statistically correct testing of topography differences is not straightforward and given the indirect relationship between electrode location and anatomical origin of the observed ERP, this does not present convincing evidence. Other references refer to anatomical differences (Filipek, 1996), a DTI study involving a passive phoneme task with written materials (Richards and Berninger, 20008), two narrative reviews (Démonet et al., 2004 and Vlachos and Avramidis, 2020) and a study on developmental language disorder (Wilson and Bishop, 2018). Either alternative references that convincingly test oral language lateralisation needs to be included or the statement needs to be toned down considerably.
		2. p4, l22-24: The authors refer to Archer et al. (2020)—a paper that presents a theory/hypothesis on the potential role of abnormal neural oscillations in response to visual in dyslexia. Given that the statement is about spoken syllables this seems odd. Also, how is ‘representation’ different from ‘storage’?
		3. p5, l15: The authors state that spelling and writing difficulties in individuals with dyslexia are harder to overcome than their reading difficulties. Although I agree that spelling and writing difficulties in dyslexia have been less thoroughly researched, I don’t see how the references support the statement or why it is necessary to pitch reading against writing difficulties. Even in ‘compensated’ adult dyslexic readers, their reading remains less accurate, slower and more effortful.
		4. p6, l9: Another statement about the cerebral lateralization of language with an odd choice of references as Ashburn investigates cerebellar function and Munzer et al. (2020) is a narrative review aimed at paediatricians. Richlan (2020) provides a narrative review of the literature on structural and functional brain activation abnormalities in dyslexia during reading and reading-related tasks (not oral language). Please clarify the main statement and justify with appropriate references, preferably to empirical studies.
		5. p6, l12-16: The characterisation of the oral language system seems very over-simplified and also unnecessary as there doesn’t seem to be any need to discuss within hemisphere differences (anterior vs. posterior systems) for a study that will not be able to differentiate between the two.
		6. p8, l18-20: Please clarify what can be concluded from the statement that “functional connectivity of the left hemisphere regions did not meet the significance criterion for any diagnostic group”.
		7. p8, l23-24: The authors support a big statement (that “early detection and intervention for children at risk for dyslexia has been associated with greater efficiency in the reorganization of the language network compared to older children”) with inadequate references to Démonet et al. (2004) and Munzer et al. (2020), both narrative reviews aimed medical professionals without empirical data.
		8. p10, l8: The authors refer to Kondyli et al., 2017 following a statement about fMRI, but that study employed fTCD.
		9. p11, l1-2: Please clarify how the tasks used in the references provided show that fTCD is a method “well suited for use in motor tasks, such as writing, as the measurement is unaffected by such movement”.
2. On a number of occasions, more methodological detail is required:
	1. p11, l18: It would be helpful if the authors could clarify why children aged 7-9 years would be considered at risk, rather than diagnosed. Typically, children ‘at-risk’ are younger (pre-reading). A brief overview of the education system, the reading and writing related curriculum (e.g., the age at which reading/writing instruction starts) and procedures for diagnosis of dyslexia in Greece would be really helpful to the reader.
	2. p11, l18 onwards: Please clarify how the sample size (n=20 in each group) was determined. In the study design template at the end of the manuscript the authors mention the effect size observed in Illingworth and Bishop (2009), but that is not currently explained in the methods.
	3. p12 onwards: Please provide estimates of reliability for all proposed measures.
	4. p12, l8-9: What will be the cut-off for ‘right-handed’ as measured on the EHI?
	5. p12, l11: Why the 85th percentile? How does that reflect ‘average’ intelligence. Wouldn’t average mean 50th percentile? Please clarify. Or is 85 a standard score?
	6. p13, l15: Please clarify the cut-off of 25% or higher for typical readers was determined.
	7. p15-16: Please clarify the type of score that will be derived from the reading tests used to identify signs of dyslexia. Will this be a standard score? If so, what is it’s mean and SD? Also, please clarify how the cut-off of ‘at least one standard deviation below the mean’ was decided upon and does ‘the mean’ refer to the norms or to the scores obtained from participants in this study?
	8. p18, l6/l13/l17: The authors refer to scores determined based on the number of errors observed on the writing tasks. Only on p19, it is explained what is meant by ‘an error’. It would be easier for the reader to follow, if the authors first discuss what will be considered ‘an error’ and then describe the scoring system. Relatedly, more context and detail regarding the scoring of errors on the writing tasks is needed. For instance, is the ‘typically taught manner’ homogenous across schools in Greece? How much ‘leaning towards the lines, written above or below the lines’ etc. will be considered incorrect? What is meant by ‘understandability’ (p19, l4)? Who will make these judgements? Will this be conducted by more than one person? If so, how will inter-rater reliability be assessed?
	9. p21, l12: Figure 3 suggests that pegs will be moved from the top row to the bottom row. In the text it is stated that “children will begin with the right hand by removing the pegs that are place ipsilaterally” suggested a movement from right to left. Please clarify the placement of the peg-board and the direction of movement for the pegs. Please also describe whether this will be the same for both hands and all trials.
	10. p23: With regard to the space portal task, please provide further information on the items that will be presented for written language trials. Just to prevent any issues with writing down the names of the items presented in the target picture, have these items been piloted to confirm that children of the relevant age are familiar with them and the pictures used unambiguously elicit the intended word?
	11. p23: With regard to the space portal task, please provide further information on any analyses carried out of the responses generated by the child. Will the letters/words be checked for accuracy and or number? Will task performance be taken into account either as an inclusion criteria or as an experimental measure?
	12. p24/26: It would be more informative and easier for the reader to follow if figures 4 and 5 were combined into one figure, with the timelines for stimulus presentation and periods for data analysis combined into one. Please also make sure you indicate the period that will act as the baseline, as well as the event marker that is mentioned in the text (p25, l17).
	13. p25, l20: Please clarify what is meant by “affecting less than 5% of the data”. Do you mean values beyond -3 SD to 4 SD are expected in less than 5% of the data? If that is what you mean, what is that based on? Or do you mean only if less than 5% of the data constitutes values beyond -3 SD and 4 SD will those data points be corrected with a linear interpolation from 1.5 seconds either side of the extreme value? In that case, please state explicitly what will happen if more than 5% of the data are affected by this. The same questions apply to p26, l1. Please clarify.
3. With regard to the analyses, the authors propose to compute a third laterality index to represent the lateralization for the linguistic component of writing as the difference between the laterality index for the written word generation and the letter copying conditions (p26, l8-11). Their main hypothesis (H1) depends on this measure, as they predict that this difference score LI of the group of children at-risk of dyslexia will be less lateralised than for the typical readers. I’m not sure this is a good idea for several reasons: a) the reliability of a difference score depends on the reliability of its component scores and the correlation between the component scores (see Trafimow, 2015 for a discussion). These reliabilities and the correlation might be sufficient to warrant the use of a difference score, but as this seems a relatively new paradigm, we might not know whether this is the case. Do the authors have any pilot data to confirm the use of a difference score is warranted? b) It could be that the groups do (or do not) differ in LI for the letter copying condition, but do (or do not) for the written word condition. A difference score will not necessarily provide a lot of insight into the origin of any difference between the groups. Therefore, it might be more prudent to compare LIs for the conditions separately (alternative I). If a group difference in LI is found for the written word condition, but not for the letter copying condition, this will provide evidence for a difference in lateralisation for what the authors refer to as the ‘linguistic component of writing’. It would also be an option (alternative II) to use the LI for the written word condition as a dependent variable in a regression model and use the LI for the letter copying condition as well as group (at-risk for dyslexia vs. typical reader) and their interaction as predictors. Both these analyses (alternative I and II) seem more informative and ‘safer’ to me.

Minor issues

Introduction:

p4, l2: What is meant by ‘discrepancies’? Discrepancies between what and what? Or do the authors mean ‘deficiencies’?

p5, l7: Part of sentence seems ungrammatical “, and of visual stimuli …”

p5, l10: Do the authors mean ‘learning letter names’?

p5, l17: What is meant by ‘orthographic’ writing?

p6, l5: When saying ‘delayed’, do the authors mean that difficulties in writing might be observed later because writing is acquired later than reading by all children? Please clarify.

p9, l5: The authors refer to their task as a ‘word generation task’. This is slightly misleading as an often used task in the field that is referred to as a ‘word generation task’ is markedly different from the task the authors propose to use. Specifically, the classic word generation task provides the participant with a letter and the participant is asked to generate as many words as possible starting with that letter. The task the authors propose to use presents the participant with a picture that contains a number of items that all start with the same letter in Greek. The participant is asked to write down the names of the items. To prevent confusion in the literature, it would be desirable to use a different name for the task the authors propose to use.

Methods

p17, l9-13: Will the ‘age-appropriate narrative and expository texts’ be the same for all children in the study? Please clarify.

p19, l18 onwards: The authors mention they will assess handedness using a child-friendly version of the EHI (used as an inclusion criterion), but also using two additional tasks ‘for completeness reasons’. Please clarify how those data will (or won’t) be used.

p20, l8/l20: On line 8, the authors refer to a ‘laterality index’, whereas on line 20, they refer to a ‘lateralization index’. Please use one term consistently.

p20, l14: The ‘degrees’ symbol should be upper script.

p21, l6: “that M. Annett (1985) proposed that it can also be applied” seems ungrammatical. Please reformulate.

p22, l12: Text in brackets “(Cerebral language lateralization)” seems unnecessary and potentially confusing. Would take it out.

p25, l19 and various other occasions throughout the manuscript: ‘following Badcock et al. (2018)’ is probably better than “according to Badcock et al. (2018).

p25, l11: “penetrating the skull” sounds a bit dramatic, maybe rephrase to say “failure to secure a signal”.

p26, l5: The authors mention “The periods of interest …”, but go on to only specify one time period (2-58 seconds). Do you mean the period of interest for both conditions?

p26, l11: It says that “The validity of the measurements will be secured with a split-half reliability test …”. Split-half reliability is a measurement of reliability not of validity.

p27, l12: I think a child usually gives ‘assent’ rather than ‘consent’.

p27: Please clarify whether the guardian will accompany the child into the room where the testing takes place.

p27-28: I don’t have any experience with Bayesian analyses, so do not feel component to comment on their appropriateness. With regard to hypothesis 1 though, I wondered whether the proposed tests are directional. After all, the authors have a clear directional hypothesis (LI is smaller, i.e., behaviour is less lateralised, for the at-risk group compared to the controls).