Notes on the revised manuscript:

We are thankful for the constructive suggestions made by the reviewers, Professor Richards and Dr Groen, and the recommender, Dr Saloni Krishnan. All comments have been carefully considered and the manuscript is now revised accordingly. The modifications to the registered report are discussed below in detail. Reviewers' and recommender's comments are in *italics* and the authors' response in normal font.

1st reviewer (Todd Richards)

Reviewer's comment:

This pre-registered report has excellent development of background, hypotheses, and methods. Here are some suggestions to help improve the plan. First of all, great job on the plan to characterize the written language deficits!

Authors' response:

We would like to thank Prof. Richards for his kind and encouraging comment.

Reviewer's first comment:

Hypothesis 2 needs to be further explained and this reviewer recommends modifying the hypothesis to say that the correlation between writing competence and cerebral lateralization will be different between children with dyslexia and typically developing children. Specifically , the slopes of the correlations should be tested for differences between children with dyslexia and typically developing children.

Authors' response:

We thank the reviewer for suggesting a statistical test between the correlations of the two groups. We would like to point out that beyond our registered analyses, we aim to separately calculate slopes for each of the two groups separately and test for correlations. However, given the exploratory nature of these analyses, we decided not to register them. Further, we would like to point out that in order to calculate a Bayes Factor comparing the two slopes, Markov Chain Monte Carlo (or similar) simulation approaches will be necessary (see Schönbrodt & Wagenmakers, 2018; https://doi.org/10.3758/s13423-017-1230-y) because the models under consideration (correlation differences between children with dyslexia vs typically developing children) are not considered nested (i.e., the prior assumptions draw upon different -not conjugate- families of distributions). As such, this will complicate the scope of the current work, a fact that is counterintuitive to the aims of Stage 1 registrations, which suggest the registration of the most relevant -to the hypotheses- statistical tests.

Reviewer's second comment:

For statistical analysis of the data section of the plan, the authors should clarify how the Bayesian approach can be used to statistically compare the cerebral lateralization between the two groups.

Authors' response

Taking into consideration the reviewer's comment, we included the following sentences to clarify the applicability of the Bayesian approach for our analyses:

"Bayesian hypothesis testing is a probabilistic approach that enables the comparison of two models (i.e., two hypotheses; alternative compared to null). This comparison is reflected in the Bayes Factor, a calculation that indicates the model (or hypothesis) with the highest likelihood (e.g., Savage-Dickey method; Wagenmakers et al., 2010)." (page 32 lines 11-15)

In the analysis plan we are describing the dependent and independent variables and the priors that will be used for the *t*-tests, i.e. "The dependent variable that will be compared between groups will

be the LIdiff. The prior for the *t*-test will be described with a half Cauchy distribution (Rouder et al., 2009) and with a scale parameter of 0.6 (Schmalz et al., 2021), which corresponds to Cohen's *d* for the difference in language lateralization between adults with dyslexia and neurotypical adults in Illingworth and Bishop (2009)". (page 33 lines 4-8)

Reviewer's third comment:
This recommends recommends adding statistical methods on how to compare the correlation slopes between the two groups using statistical software R for Hypothesis 2 (R has a nice tutorial on how to make this comparison). There should be a description on generating a scatter correlation plot for hypothesis 2 complete with regression fit lines for each group.

Authors' response:
We thank the reviewer for suggesting a comparison of the two correlation slopes. As discussed in the response to the first comment of the reviewer, unfortunately, given that such an analysis requires comparisons between non-conjugate distributions, within the Bayesian framework, such an analysis will require specific modelling of the proposed hypothesis (whether the correlations differ between the two groups) with Markov Chain Monte Carlo simulations. Given the complication of registering such a Bayesian model for a Stage-1 RR, we decided to register only the most relative -to our hypothesis- analysis, as suggested by Stage-1 registrations. However, we would like to point that we are planning to compute and report different correlations for each of the groups, but we are considering such analyses exploratory and therefore we did not register them.

2nd reviewer (Margriet Groen)

Major issues
Reviewer's introductory comment :

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:(...)

Authors' response:
We would like to thank the reviewer for thoroughly considering and suggesting an alternative, potentially more focused, design for our introduction. However, we have to admit that we were surprised to discover such a discrepancy between the opinions of the two reviewers regarding the structure and the content of the introduction (Reviewer 1, Prof. Richards, stated that "*This pre-registered report has excellent development of background, hypotheses, and methods.")*. Having said that, we did carefully consider each of the suggestions made by Dr Groen, as follows:

Reviewer's comment part (a):
(...)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);(...)

Authors' response
We agree that this study focuses on the writing difficulties in children with dyslexia and we do discuss relevant evidence (page 5 line 17 to page 6 line 13), as the reviewer suggests. However, given the high comorbidity of dyslexia and dysgraphia and their common traits, we believe that it is

appropriate to also include the following three sentences on dysgraphia: “Moreover, 30% to 47% of individuals with dyslexia are also diagnosed with dysgraphia (Ashraf & Najam, 2020; Berninger et al., 2008), an SLD related to impairments in the handwriting component of written language in spite of typical motor function (Berninger, 2009, 2004). Both SLD share the feature of compromised phonological awareness (Döhla & Heim, 2016; Vlachos & Avramidis, 2020). They differ in that dysgraphia specifically affects handwriting skills, while dyslexia has an impact on all aspects of writing (Chung et al., 2020; Döhla & Heim, 2016; Vlachos & Avramidis, 2020)..” (page 6 lines 2-8)

Reviewer’s comment part (b):
(...)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; (...)

Authors’ response:
Secondly, the reviewer suggests discussing the findings on the cerebral lateralization for written language in neurotypical individuals focusing on fMRI and fTCD evidence. We do discuss fMRI evidence in the manuscript (page 6 line 8 to page 7 line 5). We have now also added the only two fTCD studies that have to date studied cerebral lateralization for written language (Kondyli et al., 2017; Papadatou-Pastou et al. 2022). Of note, these studies do discuss the dissociation of the linguistic and motor component of writing as well as handedness differences. We believe this addition is making the manuscript stronger and we thank the reviewer for pointing out that we have overlooked reporting these studies.
More specifically, we have included the following paragraph (page 12 line 4 to page 13 line 1):

“The usefulness of fTCD in studying cerebral lateralization in individuals with reading difficulties (Illingworth & Bishop, 2009) and in studying the cerebral lateralization of written language (Kondyli et al., 2017; Papadatou-Pastou et al., 2022) is illustrated by three studies that have informed this pre-registration.. Illingworth and Bishop (2009) is the only study that has employed fTCD to compare cerebral lateralization for (oral) language production in adults with dyslexia and neurotypical adults. In their study, adults with dyslexia (*n*=30) showed more atypical lateralization for oral language compared to the control group (*n*=30) and, in fact, this difference resulted from a decrease in left-lateralization rather than an increase in right-lateralization. FTCD data on the cerebral lateralization for written language are available only for neurotypical adults from Kondyli et al. (2017; 30 left- and 30 right-handers) and Papadatou-Pastou et al. (2022; 23 left- and 31 right-handers). The former study compared cerebral lateralization for written language using oral word generation as a control condition, therefore it employed a linguistic rather than a motor control task. Papadatou-Pastou et al. (2022) was the first attempt to dissociate the linguistic and the motor component of written language in terms of cerebral lateralization by using a motor control task (symbol copying). However, a limitation of this study was that the presentation and drawing of symbols is not directly comparable to writing, because symbols are less used in communication, and hence drawing symbols could require higher attentional demands compared to writing. In order to overcome this limitation, in the present study we will use a more ecologically relevant motor control to writing, letter copying. Our main hypothesis is the following:”

Reviewer's comment part (c):
(...) 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); (...)

Authors' response:
Thirdly, the reviewer suggests that we present the evidence regarding cerebral lateralization during language production tasks in dyslexia and not to deviate the focus to data from other language tasks. We appreciate this argument, however, to our knowledge, most studies have assessed cerebral lateralization of language in dyslexia using silent reading (e.g., Schulz et al., 2008; Seki et al., 2001; Simos et al., 2000), speech perception (e.g., Heim et al., 2003), or phonological processing (e.g., Helenius et al., 2002; Shaywitz et al., 1998), which do not require participants to provide any oral response. Only two fTCD studies have assessed cerebral lateralization during oral word production (Illingworth & Bishop, 2009; Whitehouse & Bishop, 2008). Therefore, in order to be more inclusive regarding the literature on the cerebral lateralization of language in cases of dyslexia, we have introduced all of these studies, regardless of their focus on language production (page 7 line 18 to page 8 line 2):
"It is worth noting that most findings on cerebral lateralization for language in cases of dyslexia or reading difficulties come from studies assessing phonological processing (e.g., Helenius et al., 2002; Shaywitz et al., 1998), oral language perception (e.g., Heim et al., 2003), and covert oral language generation, manifested through silent reading (e.g., Schulz et al., 2008; Seki et al., 2001; Simos et al., 2000) or word generation (Whitehouse & Bishop, 2008; including adults with specific language impairment). Only Illingworth and Bishop (2009) have measured cerebral lateralization for language during overt oral word generation in adults."

Reviewer's comment part (d):
(...) 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). (...)

Authors' response:
Finally, the reviewer proposes to present the aims and the hypotheses of the present study, which are discussed in page 9 line 21 to page 10 line 18 and in page 13 lines 2 to 12.
We would like to clarify that we support the null hypothesis for our second Hypothesis because to our knowledge this is the first study to explore the correlation of writing competence and the cerebral lateralization of language and we could not rely on evidence from a previous study in order to form an alternative hypothesis nor to suggest a prior, which is required using the Bayesian approach. We have included the following text in order to support our rationale:
"Given that this will be the first study to investigate this correlation, there is no evidence that could allow us to support an alternative hypothesis, meaning the existence of a correlation between writing competence and the lateralization of the linguistic component of written language. Therefore, we put forward the null hypothesis" (page 13 lines 6-9)

Reviewer's introductory comment:
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. (...)

Authors' response:
We would like to thank the reviewer for her careful and in-depth observations. All of her comments helped us improve our manuscript in terms of accuracy. More specifically, we would like to address

separately each of the points that the reviewer made (page and line numbers correspond to the revised manuscript):

Reviewer's comment part (a):

(a) *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.*

Authors' response:

The reference list for the stated sentence (page 3 line 11) is changed from

"… (Araújo et al., 2012; Démonet et al., 2004; Filipek, 1996; Richards & Berninger, 2008; Vlachos & Avramidis, 2020; Wilson & Bishop, 2018)" to

"… (Illingworth & Bishop, 2009; Rumsey et al., 1997; Whitehouse & Bishop, 2008)"

Reviewer's comment part (b):

(b) *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'?*

Authors' response:

Following the reviewer's comment, the reference of Archer et al. (2020) was removed.

Reviewer's comment part (c):

(c) *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.*

Authors' response:

All the references that we used to support this argument discuss that longitudinal observation of individuals with dyslexia has shown that writing difficulties, which are understudied, are more difficult to compensate compared to reading difficulties, which are largely studied. We believe that stressing these points adds to the justification of our choice to study writing in individuals with dyslexia.

Reviewer's comment part (d):

(d) 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.

Authors' response:

The references used for supporting our main statement that the neurophysiological inquiry of cerebral lateralization has focused on oral language (page 6 lines 14 & 15) have changed from
"…(see Ashburn et al., 2020; Munzer et al., 2020; Richlan, 2020)" to
"…(e.g., Lohmann et al., 2005; Stroobant et al., 2011; Szaflarski et al., 2002).

Reviewer's comment part (e):

(e) *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.*

Authors' response:

The sentence that describes the anterior and posterior systems of the left hemisphere has been removed.

Reviewer's comment part (f):

(f) *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".*

Authors' response:

This sentence has been removed.

Reviewer's comment part (g):

(g) *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.*

Authors' response:

The reference list for the given phrase (page 9 line 18) has changed from:
"…(Démonet et al., 2004; Munzer et al., 2020)" to
"…(e.g., Connor et al., 2013; Lovett et al., 2017)"

Reviewer's comment part (h):

(h) *p10, l8: The authors refer to Kondyli et al., 2017 following a statement about fMRI, but that study employed fTCD.*

Authors' response:

The Kondyli et al. (2017) reference has been removed.

Reviewer's comment part (i):

(i) *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".*

Authors' response:

Following the reviewer's comment, we modified the sentence (page 11 lines 19-21) as follows:
From: "Moreover, this method is well suited for use in motor tasks, such as writing, as the measurement is unaffected by such movement (Badcock et al., 2012; Gutierrez-Sigut et al., 2015; Meyer et al., 2014; Petit et al., 2020; Whitehouse & Bishop, 2008; Woodhead et al., 2018)"
To: "Moreover, this method is well suited for use in motor tasks and specifically writing, as the measurement is unaffected by such movement (Kondyli et al., 2017; Papadatou-Pastou et al., 2022)"

Reviewer's introductory comment :

On a number of occasions, more methodological detail is required:

Authors' response:
We are grateful for these suggestions and we have taken them all into consideration. This allows us to provide the reader with a more comprehensive, transparent, and replicable methodology. We would like to respond to each of the questions raised above:

Reviewer's comment part (a):

A. *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.*

Authors' response:

Following the reviewer's comment, we have included the following sentences in our "Sampling plan" section:
Page 14 Lines 4-8:
"Of note, in Greece, children do not receive a formal diagnosis of dyslexia before the age of 9. Therefore, the 7-9 years old age-range ensures that children have reached the appropriate educational level in order to respond to the requirements of the language tasks (letter copying and written word generation) and at the same time they have not yet received a formal diagnosis."

Reviewer's comment part (b):

B. 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.

Authors' response:

In order to support our decision to initially sample 20 children in each group, we have included the following part:
Page 14 Lines 2-4:
"This sample size has been used in similar studies which report statistically significant findings when comparing laterality data derived from children with dyslexia and controls of that age range during language tasks (e.g., Mandke et al., 2022; Vandermosten et al., 2013)."

Reviewer's comment part (c):

C. p12 onwards: Please provide estimates of reliability for all proposed measures.

Authors' response:

Following the reviewer's suggestion, we have included estimates of internal consistency or reliability for all the tests for which it was applicable. More specifically, we have now submitted the following information for each test:

- Raven's: "The internal consistency of the Greek version of Raven's CPM is .82 (Sideridis et al., 2013)." (page 16 lines 13 & 14)
- Child Behaviour Checklist for school age: "(alpha coefficient = .93; Rescorla et al., 2006)" (page 17 line 8)
- Test for the Detection and Investigation of Reading Difficulties in Kindergarten and First and Second Grade: "The internal consistency of this test is higher than .8." (page 18 line 19)
- Test-A: "The internal consistency (Cronbach's alpha) of this test is .7 (Sideridis et al., 2015)." (page 19 lines 11 & 12)
- Reading Skills Assessment Test:
    - Word and pseudoword decoding component: "The internal consistency (omega coefficient) of these two components is .895 (η=1)." (page 20 lines 7 & 8)
    - Reading fluency component: Not applicable, given that it consists of a single item
    - Reading comprehension component: "The internal consistency (omega coefficient) for this component is .893 (η=.728)." (page 20 lines 16 & 17)
- Writing scale of the Luria-Nebraska neuropsychological test: "The Luria-Nebraska test has been successfully used in different cultures and was found to be reliable and valid (Golden et al., 1981a, b; Moses and Golden, 1979). The test-retest reliability for the writing scale of the standardized Greek version of Luria-Nebraska's neuropsychological battery is .93 and inter-scorer reliability varied between .85 and .92 (Vlachos & Bonoti, 2006)." (page 22 lines 17-20)
- Orthography test: "The test-retest reliability for the Orthography test is .88 and inter-scorer reliability is 1.0 (Vlachos & Papadimitriou, 2003). (page 23 lines 7 & 8)
- Edinburgh Handedness Inventory: "The test-retest reliability for the original EHI is .985 (Ransil & Schachter, 1994)." (page 24 lines 6 & 7)

- Quantification of Hand Preference Task: “The test-retest reliability for the original Quantification of Hand Preference task is .78-.8 (Doyen & Carlier, 2002).” (page 24 lines 19 & 20)
- Peg-moving task: “The test-retest reliability for the Peg-moving task is .89 (Doyen & Carlier, 2002).” (page 26 lines 9 & 10)
- Picture name writing task: Not available

Reviewer’s comment part (d):

D. p12, l8-9: What will be the cut-off for ‘right-handed’ as measured on the EHI?

Authors’ response:

In order to clarify the threshold for the categorization of children as right-handed, we have included the following sentence:
Page 24 Lines 5 & 6:
“Children with scores higher than 50% will be considered as right-handed.”

Reviewer’s comment part (e):

E. 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?

Authors’ response:

We would like to thank the reviewer for pointing out this incorrect phrasing. We have now used the correct terminology:
“Intelligence: Of at least average intelligence, as indicated by a non-verbal IQ score higher than the 85 (see Mackenzie & Wonders, 2016) in Raven’s Coloured Progressive Matrices (Raven & Court, 1938; Sideridis et al., 2015, see bellow),”(page 14 lines 16 & 17)

Reviewer’s comment part (f):

F. p13, l15: Please clarify the cut-off of 25% or higher for typical readers was determined.

Authors’ response:

In order to clarify our rationale, we have included the following phrase:
“in order to exclude potentially at-risk children” (page 15 line 21)

Reviewer’s comment part (g):

G. 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?

Authors’ response:

We would like to clarify the process for each of the initial tests that will indicate the risk for dyslexia by including the following text in our manuscript:
For the Test for the Detection and Investigation of Reading Difficulties in Kindergarten and First and Second Grade (Porpodas, 2007): “The sum of the initial scores is converted into a standard score (T-score). Ten is an average performance, and the standard

deviation is three. The children who will have a T-score equal or less than 7, are the children characterized as high risk. The performance of T-score 7 and below is described as a low performance. So the cut-off is one standard deviation below the average performance. In fact, this means that these children are likely to receive the diagnosis of dyslexia." (page 18 lines 14-19)

For the Test-A (Panteliadou & Antoniou, 2007): "An initial score for the overall performance in each subscale of the test is computed. Then, the raw score is converted as appropriate, in terms of range and gender, to a standard score (T-score) following the Appendix of the Exam Guide Table and the percentage value (percentile) is calculated. The mean is the 50th percentile." (page 19 lines 8-11)

Reviewer's comment part (h):

H. *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?*

Authors' response:

Following the reviewer's comment we have moved the sentence that describes the types of errors in page 21 lines 10-14. Moreover, we state that there will be two raters that will evaluate the performance of children in the writing tasks. Regarding the assessment of inter-rater agreement, we have included the following sentence in the manuscript:
"Cohen's kappa will be used to assess the inter-rater agreement for the evaluation of writing competence (Shabankhani, 2020)" (page 22 lines 21 & 22 and page 23 lines 8-10)

Reviewer's comment part (i):

I. *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.*

Authors' response:

To clarify the procedure that children will follow during the Peg-moving task, the pegs will be moved from the top (filled) row to the bottom (empty) row with a right-to-left direction when using the right hand and with an opposite direction when using the left hand. Children will always start with the right hand and they will perform three trials with each hand. The latter information is included in the manuscript (page 26 lines 1 & 2). To make this procedure more transparent we have included the following sentence in our manuscript:

"When participants use the right hand, they will move the pegs with a right-to-left direction and when they use the left hand, they will move the pegs with a left-to-right direction." (page 25 line 13 to page 26 line 1)

Reviewer's comment part (j):

J. 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?

Authors' response:

The reviewer has asked to provide some examples of the items that will be presented for children to name. In order to delineate our word/picture choice, we have incorporated the following text in our manuscript:
"In general, the words that children will be requested to write are included in their year 1 and 2 textbooks. For instance, for the Greek letter «μ», the children will be asked to name a baby, an eye, an apple, and an ant (all starting with the letter «μ» in Greek)." (page 27 lines 9-12)

Reviewer's comment part (k):

K. 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?

Authors' response:

Besides the neurophysiological assessment (cerebral laterality) of children in the Space Portal task, we record the children's responses. Although we have not preregistered any assessment of these responses, we will take them into consideration in case they could provide an interesting parameter for secondary analysis. However, such an analysis is beyond the scope of this registered report.

Reviewer's comment part (l):

L. 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).

Authors' response:

Following the reviewer's comment the two figures have been merged in Figure 4 (page 29).

Reviewer's comment part (m):

M. 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.

Authors' response:

Considering the reviewer's comment, we have modified the wording to make it clearer:

"Values beyond -3 SD to 4 SD from the mean (see Bishop et al., 2014), affecting less than 5% of the data, will be corrected with a linear interpolation from 1.5 seconds either side of the extreme value."

was replaced by:

(Page 30 lines 11-14)
"Values beyond -3 SD to 4 SD from the mean (see Bishop et al., 2014), affecting less than 5% of the data, will be corrected with a linear interpolation from 1.5 seconds either side of the extreme value (if ≥ 5% if affected, no adjustment will be made)." and

"Epochs with cerebral blood flow velocity lower than 50% or higher than 150% of the mean velocity or with an absolute left-minus-right-channel difference of 20 times each child's inter-quartile, affecting 1% of the data, will be rejected."

was replaced by:

(Page 30 lines 15-18)
"Epochs with cerebral blood flow velocity lower than 50% or higher than 150% of the mean velocity or with an absolute left-minus-right-channel difference of 20 times each child's inter-quartile, affecting ≥ 1% of the data (differences affecting < 1% of the data will be treated as true differences), will be rejected."

Reviewer's comment :

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.

Authors' response:
We understand the concern of the reviewer for the reliability of an additional parameter that depends on the reliability of two other parameters. LI_{dif} has already been used in a recently published study comparing cerebral lateralization for written language in left- and right-handed neurotypical adults with a similar task including written word generation and symbol copying conditions (Papadatou-Pastou et al., 2022). Moreover, we have received in-principle acceptance for a Stage 1 Registered Report in the journal *Laterality: Asymmetries of Brain, Behaviour, and Cognition* using the same task in adult participants and we have already collected preliminary data. We have not acquired any pilot data for this study prior to receiving in-principle acceptance. However, taking into account the reviewer's comment, we decided to provide a repeated measures Cohen's d for each of our analyses which takes into consideration the correlation of repeated measures, thus providing a more appropriate estimate of the difference (see Lakens, 2013).

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00863

Minor issues

Reviewer's comment :

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

Authors' response:
Following the reviewer's comment, we modified the text (page 4 lines 3 & 4) as follows:
From "discrepancies in reading performance"
To "deviations from the typical reading performance"

Reviewer's comment:

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

Author's response:
To improve the wording, we have added the following phrase (page 5 line 10):
"the perception of…[visual stimuli]"

Reviewer's comment:

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

Authors' response:
The reviewer correctly identified that we missed the word "letter" from this phrase. Therefore we edited the text accordingly.

Reviewer's comment:

p5, l17: What is meant by 'orthographic' writing?

Authors' response:
Following the reviewer's comment, we have corrected the wording (page 5 line 20):
From "orthographic writing"
To "orthographic processing"

Reviewer's comment:

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.

Authors' response:
Following the reviewer's comment and in order to make the text more comprehensive, we completed the sentence as follows:
Page 6 Lines 10 & 11:
"…compared to reading difficulties which can be apparent earlier than school age"

Reviewer's comment:

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.

Authors' response:
We have renamed the task to "picture name writing task"

Reviewer's comment:

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

Authors' response:
To clarify this methodological issue, we would like to explain that all children of the same age/grade will be tested using the same texts.

Reviewer's comment:

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.

Authors' response:

The additional handedness data will be collected in order to be included in the raw data files of this study that will be uploaded in osf.io of the group, for future analyses by our group or other groups. Large scale studies (e.g., Cornish & McManus, 1996; DeLisi et al., 2002; Groen et al., 2013) and meta-analyses (e.g., Papadatou-Pastou et al., 2020) have suggested measuring and reporting both hand preference and hand skill when studying handedness, given that there is a debate on which is the optimal handedness measure. Therefore, providing raw data of different handedness measures would allow comparability between different studies. When it comes to the present registered report, no handedness data will be analysed other than those confirming the inclusion of participants.

Reviewer's comment:

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.

Authors' response:
We would like to thank the reviewer for detecting this inconsistency. Following her comment, we have now corrected it, using the term lateralization throughout our manuscript.

Reviewer's comment:

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

Authors' response:
We would like to apologise for this typographic error. It is corrected now.

Reviewer's comment:

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

Authors' response:
Following the reviewer's comment we have performed the following modification:
"which can be applied to studies with children as proposed by M. Annett (1985)" (page 25 lines 6 & 7)

Reviewer's comment:

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

Authors' response:
Following the reviewer's comment we have removed the parenthesis.

Reviewer's comment:

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

Authors' response:

Following the reviewer's comment, all the instances where we referred to a study using "according to" were changed to "following". More specifically,
Page 27 Line 12
From "…according to Kondyli et al. (2017)" to "…following Kondyli et al. (2017)" and
Page 30 Line 11
From "…according to Badcock et al. (2018)" to "…following Badcock et al. (2018)"

Reviewer's comment:

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

Authors' response:
We appreciate the reviewer's comment. However, failure to secure a signal can be due to a number of reasons, so we intended to be more specific and state that the reason for exclusion will be the failure of the ultrasound beam to penetrate the skull [please also see Papadatou-Pastou et al. (2022) where this exclusion criterion is used in verbatim].

Reviewer's comment:

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?

Authors' response:
We would like to confirm that the plural tense was used to signify the two conditions. We have specified that in text as follows:
"The periods of interest (POI) for the two conditions of the Space Portal task will be 2 to 58 secs, that is, after the appearance of the stimulus until the stimulus disappears" (page 30 line 22 to page 31 line 1)

Reviewer's comment:

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.

Authors' response:
Following the reviewer's comment, the wording (page 31 lines 5 & 6) was modified as follows:
From "The validity of the measurements will be secured…"
To "The reliability of the measurements will be secured…"

Reviewer's comment:

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

Authors' response:
Following the reviewer's comment, the text (page 31 lines 20 & 21) was modified as follows:
From "…the child provides their oral consent…"
To "…the child provides their oral assent…"

Reviewer's comment:

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

Authors' response:
We would like to illustrate that the guardian will be waiting right outside the room during the task and they will be free to get in the testing room if the child asks for them. We have included the following sentence in the manuscript:
"The guardian will be asked to stay right outside of the testing room, but they will remain free enter the room as long as the child requests it." (page 31 line 22 to page 32 line 2)

Reviewer's comment:

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).

Authors' response:
We would like to confirm that the test is indeed directional. In a Bayesian *t*-test, a directional hypothesis is represented by a half Cauchy distribution, rather than a Cauchy distribution.