Dear recommender and reviewers,

Thank you for this opportunity to revise and resubmit the Stage 1 Registered Report titled, "Reading and vocabulary knowledge in English-Meetei Mayek biliterates". We have revised the manuscript again as per the feedback given by the reviewers.
We describe the details of our revisions based on the comments by the two reviewers.

Maxine Schaefer (Peer review of Stage 1 Revised Manuscript)
Reading and vocabulary knowledge in English-Meetei Mayek biliterates

The revised Stage 1 manuscript is now titled Reading and vocabulary knowledge in English-Meetei Mayek biliterates which more accurately describes the study. The research study aims to answer two research questions: 1) Are the relationships among measures of word reading and vocabulary knowledge similar or different in English and Meetei Mayek? 2) Is there linguistic interdependence (English to Meetei Mayek) in word reading and vocabulary knowledge?
I commend the authors on the clarity of the revised stage 1 report, and for the detailed feedback on the reviewers' comments. All my comments raised in the initial review have now been addressed and I am satisfied to support this stage 1 report.
The literature review has been revised for clarity and new sources have been added that provide more context to the research problem. The research questions clearly follow from the presentation of the literature, and the proposed analytic choices can address the questions. The analytic choices are clearly defined, which enables replication of the choices made. The authors have also specified how they would interpret different results I relation to their hypotheses.
The sample size for each analysis is sufficient to provide informative results, based on their SESOI and associated power analyses. I was able to reproduce the power analyses. When replicating in R, I did get some warning messages that some functions in TOSTER had been deprecated and replaced with new function names. Nevertheless, I found the same results as included in the report.
The pilot results and analysis of piloted instruments is presented in detail in the supplementary material. The results therein demonstrate that the authors have created valid and reliable instruments that can be used in the present study to answer the research questions.
The authors have sufficiently addressed how they will check data quality, deal with missing data, and sensitivity analyses. The proposed study seems to be undertaken in line with ethical standards in the field, with ethical clearance from the university, and consent obtained from school, class teacher and students. The authors have also considered the political scenario and reduced the time requirements related to participating in the study.
I am very much looking forward to the results of this research. The research will add to our understanding of reading development for multilingual readers in underexplored contexts.

Response: Your valuable feedback was instrumental in revising the manuscript. We are grateful that you ran the analyses and were able to reproduce them. Eventually, we aim to share on the OSF page a *renv.lock* file that details all the package versions. We hope to be able to run the study and share the findings with you soon.

Review by anonymous reviewer 1

First of all, let me say that I am sorry to hear that your research has been impacted by the current political situation. Second, I found the manuscript much improved (and I like the new title), but still have some suggestions for improvement.

Response: Thank you for your understanding. Your timely reviews made the revised title and the improvised manuscript possible.

- At several points in the manuscript, it is stated that you will be recruiting children in Grades 3 and 4 (e.g., abstract, section at the end of the instruction), but in your response to reviewers, you stated that you decided to concentrate on Grade 3. Assuming the latter information is correct, please edit the manuscript to match your intended sampling.

Response: Grade 4 was incorrectly mentioned twice, we apologise for this oversight. We have corrected it. Our target sample is indeed only Grade 3.

- Thank you for the changes that you made to the introduction; I found it much clearer.

Response: Thank you for your comment, and for acknowledging the revision made.

- Can you motivate better in the manuscript, why you decided to use three different measures for word reading? There are two timed measures (one for words, one for nonwords) for each language. In addition, there is also the untimed measure. Can you make clearer what the untimed measure is adding?

Response: We have added a sentence, 'We use multiple measures to capture the different dimensions of the same construct' (Page 21). We also described that the TOWRE (timed) measures are tests of word reading efficiency. On the other hand, the untimed measure, that is, the Word Identification subtest of the Woodcock Reading Mastery Test is a test of word reading accuracy. Under the Measures section, these have been made clearer (Page 21).

- As I understand it, the only exclusion criterion is that participants have more than 30% missing data. Are there really no other exclusion criteria (e.g., uncorrected vision or hearing issues, diagnosed neurological difficulties)?

Response: As per our ethics review application and approval, we will not be recruiting any participant with diagnosed or suspected neurological difficulties. We have now explicitly written this under the Participants heading (Page 19). We will corroborate this with the respective class teachers before we start data collection.

- Can you explain the averaging of the scores more? In one of the vocabulary tests, the maximum score is 17, in the other one it is 30. By simply taking the average of these two numbers, you would essentially give more weight to the test with the higher maximum number, correct? Or are you standardizing the scores first and then averaging them? Please elaborate.

Response: Thank you for alerting us to this oversight. We will standardize (z-score transformation) and then average them. Under the '*Data Quality Checks*' heading (Page 23), we have edited the last paragraph to include this information of the composite scores.

- For hypothesis 2, will Bayesian analyses only be run in case there is a null effect in the frequentist analyses or will they be run regardless? If the latter the case, one possibility could be to only use Bayesian analyses to start with for interpretation.

Response: We will be running Bayesian analyses as well as frequentist analyses. In case of null effect, we believe that it helps us interpret the results better. If not, it gives us another source of evidence and we can be more confident about our findings. Also, it might help us design future experimental studies and be a starting point for a priori power analysis.

- I am not familiar with some of the R packages that are being suggested for supplementary analyses (e.g., sensemakr) – would it be possible to include more detail as to what they are doing (given that, to my knowledge, these are not standard approaches)?
Response: We are motivated to use the sensemakr package because it was based on the omitted variable bias framework (Cinrlli & Hazlett, 2020). It directly pertains to the important comment made by you in the first round of revision about '*how can you exclude the possibility of a third factor*' (Page 8, response to reviewers round 1). We acknowledge that it is predominantly used by social scientists interested in causal inference for observational data. We are cognizant of the fact that this methodological approach has just begun to receive attention in Psychology and ardent proposals have been increasingly made for it (for details, see Grosz et al., 2020; Rohrer, 2018; Wysocki et al, 2022). Tracing the citations tells us that most studies using this package are published in Sociology and Economics journals. The closest we could find to our study was a recent paper on children's academic achievement (Jensen, 2023).
sensemakr output gives us the percentage of variance that the unobserved or unmeasured confounding variable must explain for the obtained results to be considered not robust. We will focus on the variance reported as the robustness value (RV). Besides the quantitative estimates, it gives us a clear explanation and interpretation in the output itself. We have clarified these points on Page 26 of the manuscript.

Minor comments:
"For the child population whose main literacy input is through school books, word reading and vocabulary development across ages (Newman et al., 2016) might be different from those living in a rich literacy resources environment, particularly when formal schooling does not begin in the first language (L1) but instead in a second language (L2) or even the third, and oral exposure to the dominant literacy language does not precede that of the written language." This is a very long sentence, please split it in two for readability.
Response: Thank you, again, for helping us improve the writing for readability. We have re-written this sentence in the introduction (Page 1).
- Incomplete sentence: "For English, N = 267 (< 354) and for Meetei Mayek, N = 1078 (> 354)"
Response: It was supposed to be "The results were $N = 267$ (< 354) for English, and $N = 1078$ (> 354) for Meetei Mayek" (page 20). We have re-written it.
- "As explicitly stated in the DAGs" – what does this abbreviation stand for? It is not used anywhere else in the manuscript.
Response: Thank you for pointing this out. This sentence is from Page 26 (Data Analytical Strategy). We wrote about the Directed Acyclic Graphs on Page 19, but had forgotten to include the abbreviation, (DAGs). We have fixed that. As mentioned in the previous round of revision, more details about the DAGs as formalization of the Research Question 2 and justification for the statistical analyses are provided in the Supplementary File (Page 1). We have also addressed this in the previous response to the reviewers (Page 8).

References

Cinelli, C., & Hazlett, C. (2020). Making sense of sensitivity: Extending omitted variable bias. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 82(1), 39-67. https://doi.org/10.1111/rssb.12348

Grosz, M. P., Rohrer, J. M., & Thoemmes, F. (2020). The Taboo Against Explicit Causal Inference in Nonexperimental Psychology. *Perspectives on Psychological Science*, *15*(5), 1243-1255. https://doi.org/10.1177/1745691620921521

Jensen, S. S. (2023). The timing of parental unemployment and children's academic achievement. *Advances in Life Course Research*, 100557. https://doi.org/10.1016/j.alcr.2023.100557

Rohrer JM. (2018). Thinking Clearly About Correlations and Causation: Graphical Causal Models for Observational Data. *Advances in Methods and Practices in Psychological Science, 1*(1), 27-42. doi:10.1177/2515245917745629

Wysocki, A. C., Lawson, K. M., & Rhemtulla, M. (2022). Statistical control requires causal justification. *Advances in Methods and Practices in Psychological Science, 5*(2), 25152459221095823.