Dear recommender,

We thank you for the opportunity to revise and resubmit the Stage 1 registered report titled, '*Reading two languages simultaneously*' to Peer Community Initiative (PCI). We appreciate you and the reviewers for your thorough and valuable feedback. We have made major revisions to the We tried our best to address the suggestions and clarifications sought by the reviewers. We made crucial changes in the proposed study and have explained the reasons behind them. We divided our response into two portions: the summary of the revisions made, and detailed responses to reviewers. We refer to Maxine Schaefer as Reviewer 1 and the anonymous reviewer as Reviewer 2.

Summary of revision

We revised the title (based on the comment by Reviewer 2) as '*Reading and vocabulary knowledge in English-Meetei Mayek biliterates*'. We have rewritten the introduction for clarity, specified the hypotheses more explicitly, and revised the power analyses. Each a priori analysis is mapped to the specific research question, hence we have an updated sampling plan. In the *Supplementary File*, details of the measures adaptation pilot study- descriptive summary and alpha coefficients are now included.

The major change made for the proposed study is the selection of measures. In our previous submission, we intended to use the Raven's Progressive Matrices (RPM) as a measure of non-verbal intelligence, two measures of phonological awareness (PA), three measures of word reading, and two measures of vocabulary reading for both English and Meetei Mayek. We have decided to drop the RPM and the two phonological awareness tasks (Elision and Blending) for the reasons discussed below. Instead, working memory test will be administered. Working memory has been widely studied in literacy research (Kim et al., 2021; Lervåg et al., 2018; Peng et al., 2018) and we have included the relevant effect sizes in the *Supplementary File* (Page 2 and 3). Therefore, our revised measures include a digit span working memory test, three measures of word reading, and two measures of vocabulary reading for both the languages. Also, we will not be collecting parental data.

The first reason for the measures change is contextual political scenario. The site of the proposed study, Manipur (India), has been hit by one of the worst violence in our recent history. Schools have remained closed for most working days of 2023. Since students have lost a major chunk of their school hours during covid and the recent upheavals in Manipur, schools have expressed reluctance to give their time. Students' well-being and learning are the priorities, so we strive to minimize the time they would need to give us. This is one of the main reasons for revising the measures. Ethnic tension remains high, and the schools cannot share information of children's family or parents. Therefore, we have decided to forego using parental measures, and seeking their consent. We address the details of consent seeking in our response to reviewers and include the same in the main manuscript. In our original plan, we intended to gather parental information about their income, educational level and literacy resources. This exclusion will change the missing data handling strategy using auxiliary variables. Changes have been made in the *Supplementary File* (Page 6-7).

Based on a study we conducted with 118 Grade 3 sample from Manipur, both phonological awareness tests (Elision and Blending) of CTOPP showed Cronbach's alpha estimates around .3. We also learned that majority of our sample had little to no experience with phonics tasks. Therefore, these two measures were found to be the most difficult to answer (from students' point of view) and the most challenging to administer. Due to these reasons, we have decided to exclude these measures from the proposed study. Deacon et al. (2009) have also discussed the challenges of adapting and/or constructing reading measures for populations with markedly varying level and form of exposure to the languages. For example, children learn language A in formal schooling, but are prominently exposed to another language B through print media. They highlighted the perils of low reliability of a measure in making accurate inferences.

Moreover, a relevant conjecture based on the literature is that phonological awareness tasks are most conducive for readers whose phonological representations are more active and mature due to instruction and experience. For our target sample, we expect that a different format such as receptive phonological awareness task (e.g., odd-one-out tasks; see Jasińska et al., 2019) would be more suitable. However, we did not have the time and the resource to construct such a task for both languages. Also, this task is likely to have ceiling effect for older readers like our sample.

Paucity of time is related to the first reason we have explained for measures revision. Non-verbal intelligence test (RMP) takes an average of 45 minutes for completion. We tried using it for another study and realised it was not possible to include it for the sake of judicious usage of time. The working memory test we have chosen is quicker to administer and straightforward. Also, the two phonological tests were proven to be the most time consuming. Instructions had to be re-told repeatedly because children struggled to understand the task requirements. We have decided that excluding these measures would be the best strategy given our circumstances and resources. Moreover, these were not the key predictors in our planned analyses.

Detailed responses to reviewers

For some of the comments, we included only the abridged version of the longer sentences.

Reviewer 1

Suggestions for the Introduction/Literature Review section

"The introduction/literature review introduces the key ideas of the paper, with reference to important terms in the literacy field….. Specifically, Linguistic Interdependence, which is mentioned in research question 2, is not addressed sufficiently… detailed description .., with reference to some research examples would help the reader follow the rationale for research question 2"

Response: Thank you for pointing this out. We have elaborated the key terms introduced (e.g., lexical quality, lexical legacy), written details about the linguistic interdependence hypothesis, and cited more relevant studies.

Requesting more contextual information

"*I follow the recommendations, for example in the journal Infant and Child Development, that samples should be described in as much detail as possible.*
For example, what educational activities are children in this sample exposed to before formal schooling? Is there a policy for children to attend early childhood education centres, or do they stay at home until Reception/Kindergarten or Grade 1? What languages are used in this early education? Please also comment on the extent of use of English and Meetei Mayek in the communities the children live in. The manuscript clearly indicates that English is dominant in schooling and that is the medium of instruction, but how familiar are children with Meetei Mayek? Is it their first language? I noted that Meetei Mayek is a lingua franca, suggesting that children may also be speakers of other languages. If so, please provide some information. How familiar are children with English? Is Meetei Mayek used in society, e.g., in road signs, i.e., do children get incidental exposure to Meetei Mayek or they only ever see it in school?"

Response: We agree that information is crucial to careful interpretation of the results, thank you for your comment. We included details about practices of formal education, schooling and overall literacy rate in Manipur (Page 14-15, manuscript). We explained the differences in print and oral exposure to the two languages.
Our sample will be recruited only from private schools in Manipur. We did the same for the pilot studies. Most children in Manipur are enrolled in private schools (we provide the estimates in the manuscript on Page 14). This is a stark contrast from the all-India average pattern of higher enrolment in public/government schools (Table 1). We believe this to be a major departure from existing studies conducted in other parts of India. Discussing the reasons for the differences is beyond the scope of our response and the manuscript. However, the major difference in socio-cultural structures (e.g., relative lack of caste-based stratification, unique land and property law yielding nieghbourhoods that are not income-based) of Manipur could be attributed to the prevalence of affordable and accessible private schools across the state, and the north-east India in general. It must be noted that this is not necessarily a recent phenomenon.

Table 1: Government school enrolment of children ages 6-14 (Pratham, 2022; The Annual Status of Education Report)

Year	**All India**	**Manipur**
2010	71.8%	32.1%
2014	64.9%	24.4%
2018	65.6%	28%
2022	72.9%	32.8%

1A. The scientific validity of the research question(s).
"research question (RQ) 1 says Are the relationships among measures of word reading and vocabulary knowledge similar or different in English and Meetei Mayek? It is not explicitly stated whether the within language or across-language correlations are of interest here.

Secondly, the power analysis addresses a different question. The power analysis calculates the sample size needed to identify r = .50 between word-reading and vocabulary. However, the research question compares the magnitude of the within language correlation for these variables in English, compared to their within language correlation in Meetei Mayek. To address the first research question as proposed, the authors would need to specify what difference in the two correlations would be taken as evidence that they are sufficiently different. An equivalence test can be used to determine whether the difference is larger than this smallest effect."

Response: The research hypotheses have been rewritten for clarity. We have edited the research question to specify the correlation of interest is the within-language association. To interpret the differences in the correlation, we will do direct comparison (average *Z*-procedure and confidence intervals) and equivalence testing (Page 21, manuscript). We specified equivalence bounds for English with SESOI= .4, Δ_L =.2 and Δ_U = .5, and Meetei Mayek with SESOI= .2, Δ_L =.1 and Δ_U = .6. We added a priori power analyses for these equivalence tests (Page 2-3, *Supplementary File*).

"The second research question asks, is there linguistic interdependence (English to Meetei Mayek) in word reading and vocabulary knowledge? I ask the authors to more explicitly specify what result they will refer to as evidence for linguistic interdependence (will you refer to variance explained, statistical significance of the model comparison of models 3 and 4, or the size of the beta coefficient of the English variable?). The power analysis addresses variance explained in the outcome. Other researchers will say there is evidence for interdependence if English word reading is a significant (p< .05) predictor of Meetei Mayek word reading after controlling for other variables. Other researchers may refer to the beta coefficient being at least a certain value. Currently you only propose to examine the effect of English on Meetei Mayek, but you could also explore the same analyses with English variables as the outcomes. This result could be informative especially if language skills are stronger in Meetei Mayek."

Response: Thank you for pointing out the error we made. As Reviewer 1 and the recommender have accurately identified, there was a mismatch between the previous power analysis and the research questions. We have rectified this with different power analyses mapped to each question, and the different outputs are considered to inform the final selection of the sample size. The results of the a priori analyses for the required sample size are as follows: 354 (Hypothesis 1), 58 (Hypothesis 2i), and 88 (Hypothesis 2ii). We ran additional power analyses for equivalence testing (Hypothesis 1) (Page 2, *Supplementary File*). We have decided to choose the N= 354 as the minimum target for our new sampling plan. If this requirement is not met, we would still know beforehand the precautions we need to take about an underpowered study and make more principled interpretations of the results.
Following the suggestion of reviewer 1**,** an exploratory question is added to explore the role of Meitei Mayek reading on English word reading and vocabulary. The analyses are not

preregistered for the exploratory question, but we will run the same analyses with English variables as the outcomes.

1B. The logic, rationale, and plausibility of the proposed hypotheses, as applicable.
"*..The first hypothesis ..I recommend that the authors explicitly specify the difference in correlations that will be interpreted as a meaningful effect…The second hypothesis states word reading and vocabulary knowledge in English would positively contribute to their equivalent performance in to Meetei Mayek (MM). I agree that this is a plausible hypothesis given the literature on cross-linguistic transfer. I would ask the authors to more specifically state the interpretation of possible outcomes… I am not sure how I would interpret a null effect, where adding English to the model results in a non-significant p value, and the model comparison is not significant. I look forward to reading the authors' more detailed views on this hypothesis.*"

Response: Thank you for this insightful comment. We have described how we will interpret the specific outcomes as positive, negative or null based on beta coefficient and significance testing (Page 22-23Manuscript). Since we are no longer measuring phonological awareness, the number of covariates is reduced. Therefore, the number of hierarchical regression models for both word reading and vocabulary knowledge are also lesser. We attempted to make a constrained and cogent interpretation of null or negligible effect. Bayes factor results will be the main basis for the null effect interpretation (this has also been mentioned in the response to reviewer 2).

1C. The soundness and feasibility of the methodology and analysis pipeline
I would encourage the authors to share the Meetei Mayek tests for review. The adaptation of word reading tasks is straightforward. Because the pilot study summary statistics were not included in the supplemental materials, I am unable to evaluate whether 30 items in the receptive vocabulary test, and 30 items in the expressive vocabulary test adaptations will be sufficient or not to avoid ceiling or floor effects. The authors indicate they will use sets 1 – 15 for the receptive vocabulary measurement which are the easier items in English. This could be enough items if children are not very familiar with English, but I would expect ceiling effects if children are very familiar with English. Similarly, 15 items may be too few for Meetei Mayek if children are very familiar with the language. More information on how vocabulary items were selected across languages is needed for the reader to evaluate test equivalence.

Response: We sought official permission to adapt the measures. The PRO-ED, Inc issued us a '*Student Translation Permission T4738 and T4739 for translating the TOWRE-2 and CTOPP-2 into Manipuri*' and we were explicitly told that "**PRO-ED does not allow posting of any actual test items**". So we are unable to share the items for review. Pearson (for WRMT) however did not respond to any of our communication. Nonetheless, we are compelled to be more cautious in dealing with copyrighted materials.

We have re-written the detail of vocabulary breadth measure (Page 18, main manuscript). There was a lack of clarity describing it, perhaps prompting the comment raised by Reviewer 1. Vocabulary breadth has 30 items: items 1-15 for receptive, 16-30 for expressive, in both languages. However, 1-15 receptive are not "easier items" in either languages. In fact, all items are arranged and revised (after the pilot result) in increasing order of difficulty. The details of items selection for vocabulary, and summary statistics and reliability estimates are included for all the measures wherever suitable are provided in *Supplementary File* (Page 7).

About the sufficiency of the number of items, we acknowledge that the higher the number of items, the higher will be the alpha estimate of reliability. However, there is no empirical evidence to quantity the optimum number of items for a 'good enough' vocabulary depth task (Zhang & Zhang, 2020). Thirty items per frequency level of 1000 words is arguably appropriate (Gyllstad et al., 2015). We have tried our best to ensure words chosen are of appropriate level for the target sample by referring to textbooks used in Grade 3 and below. As highlighted by Joen and Yamashita (2014), a major issue in vocabulary measures is the tradeoff between expressive words and more passive, selection of response.

For vocabulary depth, we chose the word comprehension subtest of the WRMT. It is described by the test authors as reading vocabulary test because participants have to read each of the item before giving their response. The selection of this test is informed by the research interest, which is vocabulary depth as indexed by multiple meaning and usage of words. Our emphasis is on print and reading experience. Also, this test has three domains- antonyms, synonyms, and analogies. It requires comprehension, usage and production of a discrete word in multiple forms. Hence, we believe that it is appropriate to qualify as a measure of vocabulary depth (for previous studies that have used a similar task to measure vocabulary depth, see Kieffer & Lesaux, 2012; Ouellette, 2006; Tannenbaum et al., 2006).

I also encourage the authors to more explicitly justify their selected effect sizes of interest. The sampling plan is described well based on the existing power analyses, but the authors may need to update the sampling plan pending a revised power analysis for research question 1…The authors propose ways to handle missing data via imputation. Please explicitly state which dataset (imputed or original with missing data) will be used in your analysis.

Response: We specified the smallest effect size of interest on Page 16-17 (Manuscript) and provided the summary of effect sizes on Table 1 and 2 (Page 2-4, *Supplementary File*). As previously stated, the sampling pan has been revised. If missing data is found to be missing not at random, we will run all analyses with the imputed data (for details, see Page 6 of the *Supplementary File*).

1D. Whether the clarity and degree of methodological detail is sufficient to closely replicate the proposed study procedures and analysis pipeline and to prevent undisclosed flexibility in the procedures and analyses.

"The manuscript provides some detail of the methods to be used, but I would suggest more detail be included regarding the test development (probably best in the supplemental material), and the proposed analytic plan (in the manuscript)… . For example, will the number of variables be reduced before analysis.. Some reviewers may also point out that it seems data will be collected from multiple schools, so children are nested within schools. Will this level of clustering be accounted for, and does it need to be accounted for?

Response: We acknowledge that clustering of students in schools is a relevant factor. However, clustering at the level of school is not of research interest. We are not examining partition of variances at school clusters because the inferences are of no substantial relevance for the proposed study. At best, it can be an exploratory question which we do not intend to cover in this study. The minimum of groups required for clustering is still widely debated with no clear consensus (Bell et al., 2010; McNeish & Stapleton, 2016). The widely followed norms are rules of thumb. A simulation study by Mass and Hox (2006) shows that clusters less than 50 at level two is more likely to yield biased standard errors. If we include a random effect when there is a small number of clusters (below 10), it might also render the study to be underpowered (McNeish et al 2017).

We intend to collect data in six or seven schools. This means that there will be less than eight clusters at the level of school in our sample. In schools that have more than one classroom, known as sections, the same subject teacher is in charge of all the classrooms. For example, the English teacher in School X will be teaching both section A and B of Grade 3. Similarly, for Meetei Mayek and all other courses. We expect this to minimize the influence of classroom-level cluster and provide relative uniformity within the sample of each school.

I would suggest being more explicit about the sensitivity analyses in the main manuscript, for example, exactly state that you will run the analysis with and without outliers/missing data being addressed and compare the results, for example. Which results do you anticipate including in the main manuscript?

Response: We will run sensitivity analyses to check the differences in results with or without outliers. We will provide a quantitative estimate to rule out the influence of such factors and ensure confidence in the robustness of the results based on the sensitivity analysis outputs. We will provide the interpretation of robustness values in the manuscript.

Reviewer 2

Title:*"After reading the manuscript, I am not sure the title relates to the questions actually addressed that well. I believe something along the lines of "The relationship between reading and vocabulary knowledge in English-Meetei Mayek biliterates" would be more appropriate. Based on the title, I would expect that participants are literally reading two languages at the same time or that academic book language is more investigated specifically."*

Response: We thank you for this incisive feedback. Taking this into consideration, we revised the title as '*Reading and vocabulary knowledge in English-Meetei Mayek biliterates*'. We agree that the suggested title better conveys the study we intend to conduct.

Abstract

"You may want to add information..How many children will be recruited? What is their age (grade is not informative enough because ages per grade differ across countries)?

Response: We edited the abstract with explicit information of the target sample of children (N= 354) and the expected average age of 10 years.

Introduction

"Overall, I had sometimes trouble following the arguments that were made in the introduction (especially in the beginning). A lot of different hypotheses were introduced without really elaborating how they related to each other (when talking about the relationship between reading and vocabulary acquisition). .. I do suggest editing the introduction to be clearer"

Response: We have rewritten the introduction for readability and better flow of the theoretical rationale. We attempted to avoid sudden shift of the content and edited to improve clarity of the introduction.

"Research Question 2 – how can you exclude the possibility of a third factor…"

Response: To formalise our research questions, we represented the relationship between variables using Directed Acyclic Graph (DAGs) (Figure 1 and 2, Page 1, *Supplementary File*). This clearly identifies the covariates and provides scientific justification for statistical controls. To address the question of *a third factor* we specify the variables selection using DAGs as explained in the manuscript (Page X) and presented in the supplementary file. This allows us to accurately identify and justify the covariates controlled. Additionally, to quantify the role of any potential third factor, we will follow the omitted variable bias approach and run sensitivity analyses (Page 23, Manuscript) of the regression models.

Method:

- *What are the approximate ages in year 3 and 4? This feels like it could be very language and culture dependent.*

- *How will the written consent look like for the children?*
- *Can you say a bit more about the Meetei Mayek adaptations of the standardized tests? Were these developed for this project and/or have they been used before?*
- *Will vocabulary knowledge (breadth/depth) be tested in both languages? If so, how?*
- *Please be more specific about data exclusion criteria. Is the number 70% calculated across all data or for each task separately? Are there any other factors that may lead to a participants' data being excluded?*

Response:

We have decided to change the participants to include only Grade 3 (instead of 3 and 4). The expected average age based on the pilot and another study is 10 years.

Only oral consent will be taken from the school authority (principal or head administrator), class teacher, and students. Regular parents-teachers' meeting is not a norm yet for most schools so we will not be able to reach out to the caregivers of each student participant. We will request the class teacher and/or school authority to take the consent from families on our behalf.

Vocabulary knowledge (both breadth and depth) and all reading measures will be tested in both languages. Seven English measures were adapted into Meetei Mayek in the pilot study. All Meetei Mayek measured were constructed specifically for this project. We consulted two linguists, one translator and three school teachers in Manipur. Each measure was revised four times before we did the pilot study.

The pilot study was conducted to test the appropriateness of the measures and the items. The sample comprises of 113 students of Grades 1 to 6. Out of the total sample, there were 22 Grade 3 students, that is, 19% of the pilot sample. The descriptive summary and reliability estimates of these tests are given in Table 4 (*Supplementary File*) as requested by Reviewer 1. The test of word reading efficiency (TOWRE) - sight word and non-word subtests are timed tasks so we do not have item-level estimates since no retest was done to give us a test-retest reliability. At best, we have the correlations between the two subtests. As requested by Reviewer 1, we have elaborated the process of selection of items in tests' construction in the *Supplementary file* (Page 10-14).

Data exclusion criteria has been elaborated. The criteria of missing data is specified and we added an item-level criteria (Page 20, Manuscript).

Results

- "*Outliers…should be decided in advance. Please be specific about which variables will be entered into the regression models for Research Question 2.*"

Response: Identification of outliers (3 *SD* above or below) are pre-specified and it will inform the decision for evaluating ceiling and floor effect.

- *"Please be specific about which variables will be entered into the regression models for Research Question 2."*

Response: We have three measures of word reading. We will average them to give us a composite word reading score, which will be used for all the regression analyses.

- *"What are the Bayesian analyses adding to Research Question 1?"*

Response: Bayesian analysis provides an additional option to quantify the evidence in support of the hypothesis, whereas a standard frequentist analysis seeks to falsify the hypothesis stated. In another words, it helps us interpret null result. Second, the Bayesian estimates such as correlations and their corresponding credible intervals are more computationally robust than the estimates generated by the frequentist approach (Lee & Wagenmakers, 2013; Stefan et al., 2019). For example, credible intervals, unlike confidence intervals, are independent of large-sample approximations and the statistical tests intended or selected (Kruschke, 2021). Posterior distributions of these estimates yield inferential indices such as the Highest Density Intervals (HDIs), region of practical equivalence (ROPE) provide additional information than what we can know from frequentist correlations such as Pearson (Harms & Lakens, 2018; Nuzzo, 2017). Relatedly, we explained this in our response to Reviewer 1 regarding null interpretation for Research Question 2.

-*"For Research Question 1, how do we know which ones are meaningful or how many differences (specific ones?) need to exist for them to be meaningful?"*

Response: As mentioned in the response to Reviewer 1, we will interpret this based on confidence intervals and equivalence testing for which have specified equivalence bounds (Page 19, manuscript)and conducted apriori power analysis (Page 2-3 of *Supplementary File*).

References

Bell, B. A., Morgan, G. B., Kromrey, J. D., & Ferron, J. M. (2010). The impact of small cluster size on multilevel models: A Monte Carlo examination of two-level models with binary and continuous predictors. *JSM Proceedings, Survey Research Methods Section*, *1*(1), 4057-4067. http://www.asasrms.org/Proceedings/y2010/Files/308112_60089.pdf

Deacon, S. H., Wade-Woolley, L., & Kirby, J. R. (2009). Flexibility in young second-language learners: examining the language specificity of orthographic processing. *Journal of Research in Reading, 32*(2), 215-229. **https://doi.org/10.1111/j.1467-9817.2009.01392.x**

Gyllstad, H., Vilkaitė, L., & Schmitt, N. (2015). Assessing vocabulary size through multiple-choice formats: Issues with guessing and sampling rates. *ITL-International Journal of Applied Linguistics, 166*(2), 278-306. Doi: :10.1075/itl.166.2.04gyl

Harms, C., & Lakens, D. (2018). Making 'null effects' informative: statistical techniques and inferential frameworks. *Journal of clinical and translational research, 3*(Suppl 2), 382–393.

Jasińska, K.K., Wolf, S., Jukes, M.C.H, & Dubeck, M. M. (2019). Literacy acquisition in multilingual educational contexts: Evidence from Coastal Kenya. *Developmental Science*, 22:e12828. https://doi.org/10.1111/desc.12828

Jeon, E.H., & Yamashita, J. (2014). L2 Reading Comprehension and Its Correlates: A Meta-Analysis. *Language Learning, 64*, 160-212. https://doi.org/10.1111/lang.12034

Kieffer, M.J., & Lesaux, N.K. (2012)Knowledge of words, knowledge about words: Dimensions of vocabulary in first and second language learners in sixth grade. *Read Writ* **25**, 347–373. https://doi.org/10.1007/s11145-010-9272-9

Kim, Y. S. G., Petscher, Y., & Vorstius, C. (2021). The relations of online reading processes (eye movements) with working memory, emergent literacy skills, and reading proficiency. *Scientific Studies of Reading*, *25*(4), 351-369. https://doi.org/10.1080/10888438.2020.1791129

Kruschke, J.K. (2021) Bayesian Analysis Reporting Guidelines. *Nature Human Behaviour*, *5*, 1282–1291 . https://doi.org/10.1038/s41562-021-01177-7

Lervåg, A., Hulme, C., & Melby-Lervåg, M. (2018). Unpicking the developmental relationship between oral language skills and reading comprehension: It's simple, but complex. *Child development*, *89*(5), 1821-1838. **https://doi.org/10.1111/cdev.12861**

Maas, C. J., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, *1*(3), 86-92. https://doi.org/10.1027/1614-2241.1.3.86

McNeish, D., & Stapleton, L. M. (2016). Modeling Clustered Data with Very Few Clusters. *Multivariate behavioral research*, *51*(4), 495–518. https://doi.org/10.1080/00273171.2016.1167008

McNeish, D., Stapleton, L. M., & Silverman, R. D. (2017). On the unnecessary ubiquity of hierarchical linear modeling. *Psychological methods*, *22*(1), 114–140. https://doi.org/10.1037/met0000078

Nuzzo, R.L. (2017). An Introduction to Bayesian Data Analysis for Correlations. *PM &R, 9* (12), 1278-1282. https://doi.org/10.1016/j.pmrj.2017.11.003

Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology, 98*(3), 554–566. https://doi.org/10.1037/0022-0663.98.3.554

Peng, P., Barnes, M., Wang, C., Wang, W., Li, S., Swanson, H. L., Dardick, W., & Tao, S. (2018). A meta-analysis on the relation between reading and working memory. *Psychological Bulletin*, *144*(1), 48–76. https://doi.org/10.1037/bul0000124

Pratham (2022) The Annual Status of Education Report.

Tannenbaum, K. R., Torgesen, J. K., & Wagner, R. K. (2006). Relationships between word knowledge and reading comprehension in third-grade children. *Scientific studies of reading*, *10*(4), 381-398. https://doi.org/10.1207/s1532799xssr1004_3

Zhang, S., & Zhang, X. (2022). The relationship between vocabulary knowledge and L2 reading/listening comprehension: A meta-analysis. *Language Teaching Research*, *26*(4), 696-725. https://doi.org/10.1177/1362168820913998