Vocabulary Learning and Instruction
Volume 11, Issue 2, December 2022
http://vli-journal.org

A Commentary on Four Papers in the Second Language Vocabulary Learning Strand for the JALT Vocabulary SIG

Junko Yamashita
Nagoya University

Keywords: L2 vocabulary acquisition, vocabulary learning app, discipline-specific word list/corpus, augmented reality, proper names

1 Introduction

Vocabulary supports virtually all aspects of language performance. Researchers' awareness of its importance has been expressed in multiple theories and instructional recommendations in different skills: speaking (e.g., Kormos, 2006; Levelt, 1989), listening (e.g., Field, 2008; Rost, 2011), reading (e.g., Perfetti, 2007; Perfetti & Stafura, 2014), and writing (e.g., Schoonen et al., 2011). In the domain of second language (L2) proficiency, Hulstijn (2015) proposed a core-periphery model, in which various components underlying L2 proficiency are divided into two groups (core and periphery). The core component includes linguistic knowledge and its processing efficiency (e.g., vocabulary and syntax), and the peripheral component consists of language-general abilities (e.g., interactional ability, strategic competencies, and metalinguistic knowledge). Hulstijn contends that the core component is likely to influence L2 proficiency more strongly than the peripheral component. A recent collection of meta-analyses that synthesized correlations between various components and four skills (reading, writing, listening, and speaking) generally supported this model (Jeon & In'nami, 2022). Looking more closely at vocabulary, the meta-analyses commonly endorsed the strong correlation of vocabulary knowledge with all these skills. As briefly introduced here, the more we study with increasingly rigorous methods, the more evidence we obtain for the importance of vocabulary in L2 performance. Then, we have every reason to facilitate L2 vocabulary acquisition through pedagogical endeavours. The Vocabulary Learning strand in the JALT Vocabulary SIG (2022) features four papers that share this aim, even if each one of them worked on distinct topics.

In this commentary, the following four papers have been reviewed and discussed: Vocabulary Learning Outside of the Classroom: Institutional Use of a Spaced Repetition Vocabulary Learning App (Mueller & Hays, 2022), Developing a Discipline-Specific Corpus and High-Frequency Word List for Science and Engineering Students in the Graduate School (Uehara et al., 2022), Rasch-based

Please cite this article as: Yamashita, J. (2022). A commentary on four papers in the second language vocabulary learning strand for the JALT Vocabulary SIG. *Vocabulary Learning and Instruction, 11*(2), 101–113. https://doi.org/10.7820/vli.v11.2.yamashita

Instrument Validations for an Augmented Reality Vocabulary Acquisition Experiment (Dabrowski, 2022), and Orthographic Processing of Proper Names: A Proposal to Investigate the Orthographic Cue for Second Language Readers (Klassen, 2022).

2 The Four Papers

2.1 Vocabulary Learning Outside of the Classroom: Institutional Use of a Spaced Repetition Vocabulary Learning App (Kevin Mueller and George Hays)

Language learning necessitates a long-term commitment of learners being exposed to a massive amount of language input. A serious disadvantage in English as a Foreign Language (EFL) contexts is the sheer lack of English input/exposure. Therefore, one of the challenges of educators is to find ways to increase English input and provide learners with learning opportunities as well as to use effective methods of teaching. Because in-class time is limited, the only reasonable option to increase input would be using out-of-class time. Many educational institutions have been implementing mandatory self-learning schemes in their EFL curricula outside of the classroom. The EFL curriculum of the Global Teaching Institute (GTI) at Tokyo International University (TIU) has four such schemes. In this presentation, Mueller and Hays introduced WordEngine, an app that the GTI has been using since 2013 as one of their schemes, reported its effect, and discussed its merits and challenges.

WordEngine is “a bilingual (Japanese and English) vocabulary learning application for smartphones and computers” (Mueller & Hays, 2022). Learners start with a vocabulary test (V-check) that diagnoses their starting level of vocabulary knowledge. Based on this diagnostic data, the WordEngine provides learners with a personally tailored list of high-frequency words that they do not know yet. The learning theory behind this app is spaced learning (intervals between repetitions of a target item), which is considered more effective than massed learning (no intervals between repetitions) (e.g., Nakata, 2015). The learning task is paired-associate learning in three modes (visual, aural, and contextual) using flashcards. According to two studies cited by Mueller and Hays, WordEngine was not only faster and more effective but also resulted in larger English vocabulary gains by Japanese students compared to intensive reading, extensive reading, word lists, and paper flashcards.

The GTI has been using WordEngine since 2013. The way in which this app was implemented in the curriculum has gone through several changes regarding, for example, the requirement level of completion, the ratio to include the performance on the app to the grade, and the range of students who use the app. For example, the requirement changed from 300 to 350 correct answers on the app per week in 2019; student app performance was counted as 10% of the final grade from 2019; students who use WordEngine expanded in 2021 to include majors in International Relations and Business Economics, as well as English majors. The effects of this app that the authors presented in terms of the increase in students’ vocabulary size and estimated Test of English for International Communication (TOEIC) scores are positive. In the 2021 academic year, the average students improved their vocabulary score by 40% (or 1,153 words). The average first-year students increased their TOEIC score by 30 points.

After demonstrating the effect of the app, the authors discussed the advantages of WordEngine and problems to solve in the future. The first advantage is the increase in vocabulary learning opportunities. Second, the app provides individually tailored words so that students can focus on the words that they need to learn most. Third, WordEngine can motivate students through its programme (Team Challenge) that provides class competition. Finally, students can use it anywhere and anytime. Despite these advantages, the authors addressed three challenges: making a closer connection between vocabulary learning in the app and classroom learning, expanding the range of activities, and finding a balance between the mandatory use of the WordEngine and instructors' freedom in using various elements of the curriculum.

This paper is very informative and interesting for practical reasons for me because my university also has an out-of-class mandatory learning scheme in the EFL curriculum. The fact that the GTI has been using WordEngine for nearly 10 years and considers its continued use in the future, along with the positive data shown in this paper, supports the usefulness of this app. Knowing about this app for the first time, several questions occurred to me. The first is whether the GTI ever considered different apps. The authors cited past studies that showed the effectiveness of this app compared with more traditional methods. However, the comparisons are not with web-based applications. I wonder whether there are any other similar vocabulary learning applications, and if so what is the advantage of the WordEngine. The second question is how we decide the requirement of out-of-class learning. The paper suggests that higher requirement (i.e., more correct answers on the app) leads to larger vocabulary gains. This is a likely tendency (i.e., more self-study/efforts, more gains). However, we cannot increase the demand infinitely as students have other subjects to study. The requirement should be an upper limit within the doable work. Third, relevant to the second question, I wonder how students reacted to the app. Though the paper suggests that the app has a motivating function, has the GTI ever had students' evaluations of this app? Students' opinions on the level of requirement may provide suggestions for setting the requirement. Fourth, it would be more convincing if the benefit of the WordEngine is tested not only using the algorithms embedded within the app but also with independent tests. A related interesting question is how the increase of vocabulary through the WordEngine relates to performance in teacher-taught courses. Fifth, I wanted to know more about WordEngine's specifications, including differences in learning conditions (visual, aural, contextual) and how spaced learning is implemented in the design of learning tasks. Sixth, the paper shows that the gains (vocabulary size and TOEIC score) were larger in students with higher initial vocabulary, which seem to correspond to the more active use of the app. Even if the app is generally beneficial, it would be useful to specify what affects the gains. Is it the initial level (V-check score/general proficiency) or engagement (the degree of active use of the app)? These two factors are often associated, i.e., students with larger vocabulary sizes show more active use of the app because it is less demanding for them. If so, we may want to consider how we can support students who start the app with smaller vocabulary sizes.

2.2 Developing a Discipline-Specific Corpus and High-Frequency Word List for Science and Engineering Students in the Graduate School (Suwako Uehara, Hibiya Haraki, and Stuart McLean)

Vocabulary size strongly supports reading comprehension. There is a consensus that a lexical coverage of 95 to 98% is necessary for non-frustrating reading comprehension. Word lists and lexical coverage analyses can guide vocabulary teaching and learning because they suggest which words to learn to reach the desired level of lexical coverage. There are broadly three types of vocabulary: general-purpose words, academic words (common across a range of disciplines), and technical words (academic words but discipline-specific). L2 learners first learn general-purpose words to acquire the foundational vocabulary; when they become university students or researchers they need to learn academic and technical words to read research articles in their fields of study. The purpose of Uehara et al. (2022) is to create a discipline-specific word list for graduate students in science and engineering fields.

Enormous amounts of work are necessary for constructing a corpus and making a word list: collecting purpose-appropriate materials, turning them into electric files, identifying and correcting noises/errors, annotating the corpus according to the purpose, and counting and listing words according to the definition of a word adopted in a project. Even with computer technologies, careful manual processing by human beings with specialist knowledge is inevitable. To increase the quality of human-involved processes, researchers have to make technology-based processes as automatic as possible so that they can direct their attention maximally to manual processes. This paper reports on innovative methods to use technologies and the challenges they encountered in their project.

Below is a summary of the processes in this project. The authors first collected base materials that create their corpus by asking professors in an Engineering and Science department to provide academic reading materials. Out of 1,182 texts collected, the authors used 330 texts (1.34 million tokens) from the Chemistry and Biotechnology programs in this study. They used flemma as a counting unit of a word (the word form not differentiating parts of speech [POS]) because POS difference had little effect in assessing English word knowledge of Japanese university students (McLean, 2018) and science-specific words are often not base words nor do they have multiple POS. They included flemmas with a range larger than 3 and frequency higher than 50, and deleted the first 2,000 flemmas in the New General Service List (NGSL) (i.e., general-purpose words). They used several data processing tools (MATLAB, PDFelements, and AntCont) in combination with manual inspection by human beings.

Researchers in corpus linguistics would find their use of MATLAB (a programing language/software) highly informative. They used this software for multiple purposes: to create text files from PDF, fix errors that may affect word count, add functions to report some useful information (summary report of analysis), solve problematic issues (ligature, line-end hyphens, British/American spellings, and word sequences without spaces), select words that meet their criteria (range > 3 and frequency > 50), remove general-purpose words and one to two letter strings. The authors made MATLAB scripts available, which will be an asset for interested researchers.

The authors reported that it took only 80 seconds to create text files from the 330 PDF documents by utilizing PDFelements and MATLAB. It is a remarkable speed. From this corpus, the authors obtained a list of 1,364 tokens that meet their criteria by utilizing AntConc and MATLAB. This list was submitted to a manual check by a science professor and was reduced to 1,220 tokens, which is the product of the work-in-progress study reported in this paper. The authors intend to compile a more balanced corpus by including more sub-disciplines of science and engineering fields.

This paper is a methodological paper that contributes to creating a discipline-specific corpus and a word list. Corpus linguists will fully understand the challenges described in the paper and appreciate the demonstrated way to use multiple data-processing tools, especially MATLAB, combined with manual processing. I support the inclusion of a technical specialist (a postgraduate engineering student in this case) in the project because this interdisciplinary collaboration, as well as using the data processing software, creates more time for applied linguists to engage in various issues that necessitate their attention and manual work. The methodology shared in this paper is applicable to the creation of corpora and word lists for various purposes.

My commentary includes several questions. The first is the question of how specific is discipline-specific, which relates to the question for whom the word list is useful. Uehara et al. (2022) cited a professor who stated "lists related to *one specific research lab* would be most beneficial for the students in their laboratory" (emphasis added). If we take the face value of this quote, the word list should be made for each research lab in a department. Is it practical from the cost-benefit balance (the more narrowly focused, the smaller number of users)? Similarly, we face ambiguity in the division within a discipline. To illustrate, a quick search on the Internet finds different sub-divisions in engineering (electrical/mechanical/chemical/civil or civil/mechanical/production/chemical/biomedical/aerospace). Which discipline, sub-discipline, or combination of them should be the target? Lastly, the characteristics of users are pertinent to this issue. If we compare the word list of this study with that of Ward (2009), both are engineering-specific, they are very different. The word list of this study (for graduate students) contains far more technical words that are infrequent in everyday life (e.g., *sperm*, *oocytes*, *oscillations*) than the one made by Ward (2009) who targeted undergraduate low-proficiency students (e.g., *system*, *shown*, *equation*). These word lists seem to represent two types of technical words: those that are accessible through general knowledge (e.g., *arm*, *lungs*, *penicillin*) and those that need specialized knowledge (e.g., *costal*, *xiphoid*, *hemoglobin*) (examples are from Nation, 2016, p. 145). This contrast suggests that the utility of a word list changes even in one discipline. Unlike general-purpose word lists, the definition of specificity would need to be made explicit to users so that they can make informed decisions on how to use the word lists.

Second, I wondered whether the 1,220 tokens are all equally important in the case of discipline-specific words. For general-purpose vocabulary, words to be prioritized for learning can be determined in relation to the text coverage (e.g., words that make the coverage closer to 98% should be given emphasis; the higher the frequency, the more contribution to the coverage). Does the same logic apply to the discipline-specific words or are they special types of words that need to be learned irrespective of relative frequency for discipline-specific understanding?

The third question is the definition of a word. *Studies in Second Language Acquisition* (2021, Vol. 43, Issue 5) featured critical commentaries on the lexical unit (e.g., word type, lemma, flemma, or word family). Although supporters of flemma are increasing, there are different views and opinions among L2 vocabulary researchers. Overall, the choice of units depends on the purpose of analysis, learner proficiency, and technical efficiency. The use of flemma for Japanese university students may be backed up by McLean's (2018) rigorous study. However, whether flemma is a good unit for L2 learners with different backgrounds is a question for future exploration. I wonder if it is feasible to apply different units and compare the resulting word lists and text coverage.

Finally, and in the long run, we should investigate how we can use the word list in vocabulary teaching and learning. Is the word list useful for self-study by graduate students, for professors to teach discipline-specific words and concepts, or for EFL instructors to teach technical words as a scheme of English for specific purposes (ESP)? Making a word list is the first step and future studies should explore how best to use it in pedagogy.

2.3 Rasch-Based Instrument Validations for an Augmented Reality Vocabulary Acquisition Experiment (Adam Dabrowski)

The overarching goal of Dabrowski's (2022) study is to test the effect of augmented reality (AR) technology on vocabulary learning. Toward this goal, his paper in this volume presents the validation result of three vocabulary tests to be used in his future main study (I call this paper a pilot study). Though this is my first time seeing the term AR, I can endorse the importance of researching educational technology in the modern world where we are facing the rapid advancement and availability of information technology and students' increasing familiarity with such technologies and digital devices. AR is theorized to enhance vocabulary acquisition by adding visuospatial information (visuospatial bootstrapping or VSB) in the process of learning form-meaning mapping of new vocabulary.

Dabrowski's main study compares the effectiveness of three vocabulary learning conditions. For this purpose, he developed a set of three vocabulary tests and obtained validity evidence in this pilot study. The participants were 39 adults (university faculty, undergraduate students, and language instructors) who were competent in English or Japanese; their first languages (L1s) are English, Japanese, Chinese, Dutch, and Portuguese. The three learning conditions are all paired-associate learning. Pseudowords are used as new vocabulary to be learned, which is one of the best methods to eliminate many potential contaminations (e.g., participants' prior knowledge of target words). The pseudowords are all concrete nouns, which would make the implementation of AR possible or easier compared to abstract nouns. Thirty pseudowords are created and divided into three sets of thematic categories (Desk, Kitchen, and Music) with 10 pseudowords in each. The unit of learning and testing was one set (i.e., participants learned and were tested in one thematic category).

One learning condition was a traditional paper-based word card (WC) where three words are laid horizontally in the following manner: pseudoword/Japanese word/English word. The other two conditions are tablet-based AR-VSBs: AR1 (context-independent) and AR2 (context-dependent). In AR1, participants saw a group of objects spread on a desk with corresponding labels (identical to the WC condition) imposed on each object. AR2 differs from AR1 only in that the objects were displayed in a context (Appendix B in Dabrowski, this volume). The sequence of the study is as follows: pre-test, pseudoword learning (10 minutes), immediate post-test, and 1-week-delayed post-test. This paper analyzed the scores of the delayed post-test.

Three tests were identical in format but differ in items. The test had two sections: a Yes/No (YN) section and a meaning-recall (MR) section (Appendix C in Dabrowski, this volume). The instruction of the YN section was "*Have you learned any of the words below? Please choose 'yes' or 'no' for each word.*" Fifteen (15) pseudowords were listed: 10 learned pseudowords, for which "yes" is the correct answer and 5 additional ones, for which "no" is the correct answer. This section seems to measure the memory of seeing the orthographic form of the word, i.e., form recognition. The MR section asked the learner to type the meaning of each word. Participants were allowed to use some explanatory expressions if they did not remember the exact word that corresponds to the target word (e.g., both "dog" and "animal that barks" were accepted). This is a productive test testing the meaning recall of the 10 items in the learning section. The tests were administered with Google Forms on the Internet. Test items were analyzed by running separate Rasch analyses for each test. Results showed the overall quality of the items as being good according to the author. Therefore, with minor changes to some items, the tests are ready for the main study.

In this commentary, I focus on the design feature of this pilot study starting with five questions. The first question is about the choice of participants. It is justified if this is the type of people in the main study. However, from the items used in this pilot study, the target is more likely to be lower proficiency learners such as undergraduates or secondary school students. I am also puzzled by the variation of participants' L1. Although pseudowords make the target items unknown just like novel L2 words, using two languages (English and Japanese) as meaning labels may create different learning conditions for different participants even if differential item functioning (DIF) analysis did not find effects of L1. Again, the L1 background is justified if it represents the background of participants in the main study. Test quality can change according to test takers. I would appreciate clarification on the choice of participants in the pilot study.

The second question relates to the test format. The author seems to assume that both test formats test form-meaning mapping. However, the YN test seems to be testing the memory of seeing the target form. I wonder if the choice of these test formats is only to validate tests, that is, we can easily expect that the YN test is easier than the MR test, and the result supported it (i.e., part of validity evidence). I would appreciate some explanation on this point.

Third, the difference between AR1 and AR2 is not clear to me except for Set 2 (Kitchen). We can see some objects more clearly in Sets 1 and 3 (e.g., a laptop and a guitar) in AR2 than in AR1, but is this the context? There is no explanation of the context. I hope the author provides a definition and operationalisation of context. The fourth question is pertinent to the items. Target items (pseudowords) are constructed carefully, but I wonder about L1 words. Since a new form is linked to a known meaning through L1, some control of L1 words would be necessary (e.g., frequency, concreteness, familiarity, and typicality in each thematic category). For example, I am familiar with all Kitchen items, but "pick" in the Music category was not in my Japanese vocabulary (till I saw the picture), since I do not play the guitar. I also wondered why canned coffee is in the Music category. Lastly, I wonder whether the use of AR is equally effective for all learners. The main study could include learner features such as skills to use information technology or preference for paper-based versus tablet-based learning.

Some suggestions for future exploration are discussed. The first idea relates to the semantic category effect (Finkbeiner & Nicol, 2003; Tinkham, 1993; Waring, 1997): learning words in semantically related sets interferes with learning (the reason is arguably that presenting words in semantically related sets creates competition between words of related concepts). If this effect applies to AR vocabulary learning, the three thematic sets in this study may cause semantic interference. Although this is not problematic for the main study (because all conditions should be equally affected), it would be interesting to examine, in a future study, whether the semantic category effect is observed in the AR learning condition. If so, a slight adjustment of word sets may facilitate learning more. If AR overcomes the semantic category effect, it would increase pedagogical implications. Another interesting issue is whether the traditional WC condition facilitates learning to the same extent as the AR condition if visuospatial information (pictures) is added to the WC. We can add pictures to the paper-based WC as well. If the AR condition is still superior to the traditional WC-with-picture condition, the usefulness of AR would be strengthened. At the same time, more explanation would be needed for the AR effect than simply addressing VSB.

2.4 Orthographic Processing of Proper Names: A Proposal to Investigate the Orthographic Cue for Second Language Readers (Kimberly Klassen)

It has been assumed that proper nouns/names (PN) are unproblematic for L2-English learners because they are easily understood as such because of the orthographic convention and contextual support. This assumption has long affected the way in which the required vocabulary size for fluent reading comprehension is estimated from the text coverage analysis. However, researchers have started questioning this assumption and called for empirical investigations to test it (e.g., Brown, 2010; Nation & Kobeleva, 2016). This new research agenda can significantly impact vocabulary instruction. To illustrate, the text coverage of PN in the British National Corpus is 2.57%; the required vocabulary size to reach the desired 98.0% coverage for independent comprehension is 4,000 to 5,000 words (97.94 to 98.98% coverage) if the assumption is correct; but it is 8,000 to 9,000 words (97.86 to 98.23% coverage) if all PNs are unknown (Nation, 2013, p. 21, Table 1.3). The estimation difference is not negligible.

Several studies are relevant to testing the assumption. Studies on the effects of cultural knowledge on reading comprehension offer some insights because changing PNs from L2 to L1 (e.g., *Seattle* to *Osaka* for L1-Japanese students) is a dominant method to manipulate the cultural familiarity of texts (Erten & Razi, 2009; Klassen, 2020; Sasaki, 2000). The results were mixed; some found effects of cultural knowledge (i.e., L1 PNs facilitated comprehension) (Erten & Razi, 2009; Sasaki, 2000), but others did not (Klassen, 2020). Klassen (2020) argued that the result may be caused by the lower vocabulary size of her participants, whereby foundational vocabulary size is necessary for the effect of cultural knowledge to appear. More directly relevant to the PN assumption is Klassen (2021), who examined the effect of context (whether learners can identify PN as such in a sentential context). The result showed a low success rate (28.4%). These past studies suggest that unfamiliar PNs may inhibit comprehension or the context may not be as helpful as assumed. Klassen's (2022) paper in this volume is along the line of this research agenda. She focuses on orthographic processing and proposes her research design to test the effect of the initial capital letter (ICL) on the identification of PN as names.

Klassen (2022) modeled her study after Peressotti et al. (2003), an L1-Italian word recognition study. Although the theoretical interest differs from Klassen's, it showed a faster recognition speed of PN written with ICL compared to the other three conditions: PN written with all small letters (ASL), common nouns (CN) with ICL, and CN with ASL. Based on this finding, Klassen asked to what extent the ICL facilitates the visual recognition of English PN by L1-Japanese university students. She hypothesized no facilitation effect of ICL because (1) learners do not benefit from L1 processing strategies because of the distinct writing systems between Japanese and English and (2) learners have less processing experience in English compared with L1 readers. Let us see how Klassen plans to test this hypothesis.

The participants are university students at the CEFR-B1 level. The items are 80 English PNs and 80 CNs that are matched for length and frequency (PC frequencies are obtained from a corpus of junior and senior high school English textbooks made in Japan and CN frequencies are from a list of the first 3,000 words in the COCA). To control for semantic category effects, item categories are restricted to People and Places. Items are presented in two conditions: with ICL or ASL (e.g., *Michael/michael*). The experimental task is a name decision task: the participant sees a stimulus and responds dichotomously (Yes or No) to a question *Is this a name?* Though there is no specification in the paper, this will be an online task, because both reaction times and accuracies are to be recorded for each item. Only PN data are analyzed with linear mixed effects modeling (LME) to test the effect of ICL.

This research proposal is certainly a welcome addition to the research topic of testing the long-held assumption about PN. To my knowledge, Klassen (2021) is the first published study that directly tested the assumption by examining the context effect. This study focuses on the form (the effect of ICL on the identification of PN). The design is simple (comparing two conditions), and the proposed task seems a fine choice for the narrowly focused research question. I just wondered whether the intention of the researcher is immediately clear to all participants. We expect a Yes response to PN and No to CN. However, some participants might choose Yes for CN (e.g., *Table/table*), thinking that "Yes, it is a *name* of a kind of furniture" (instead of, for instance, "No, it is a *label* of a kind of furniture"). A training session, which is already in the plan, would solve this problem. I would like to recommend including feedback on each training item, if it were not planned, to ensure students' understanding of the task in the way that the researcher intended.

Although Klassen's (2022) reasoning behind her hypothesis sounds reasonable, there may be some other possibilities, which may be useful in the event that the hypothesis is not supported. First, regarding the L1 influence, L1-logographic readers of L2-English tend to rely on orthographic processing more than phonological processing (Koda, 2005), which is regarded as a logographic reading process transferred from L1. If so, Japanese students might be sensitive to the visual, orthographic form of a word. Second, regarding the limited L2 processing experience, we do not know the proficiency level at which students start using the ICL information in the recognition of PN. It may or may not be the CEFR-B1 level. Relevantly, Klassen correctly mentioned that there is a general shift from phonological processing to orthographic processing in the development of English reading, which seems the reason to fix the proficiency level in this proposal (so that students are at the same/similar developmental stage). However, to reiterate, we do not know the level of proficiency where the shift happens in L2 readers. Even if the shift has happened, it is a matter of degree; competent readers use both orthography and phonology according to the task or purpose of reading. It may be more fruitful to include L2 proficiency as a variable and analyze its effect in the LME. To do that, including a range of proficiency levels may be desirable.

The proposal represents a compact and interesting study. There are several ways to expand it for future studies. First, native-speaker baseline data would make a study more rigorous. Second, including more factors in the LME would be useful such as L2 proficiency, as mentioned above, and characteristics of words (e.g., familiarity, frequency, and length) even if some of these features are controlled. Although direct comparison with PN is not recommended due to the response type difference, analyzing CN data is interesting. Will this study find no cross-condition difference in CN like Peressotti et al. (2013)? In addition to expansions based on this proposal, future studies on the topic of PN processing should expand both dependent and independent variables. As for the independent variable, some other factors may affect PN processing. An example is phonological processing. Students often have pronunciation problems for long and unfamiliar (often non-English) names, which appear even in lower levels of graded readers (e.g., *Abanazar*, *Badr-al-Budur* from Grade 1 in the Oxford Bookworms series). Phonological difficulties impede reading fluency and memory integration, which may negatively affect comprehension. On the dependent variable side, different aspects of reading that may be affected by PN can be explored, such as word recognition, text reading fluency, and text comprehension. In addition, future studies can shift from a single-factor study to a multi-factor study by examining multiple variables simultaneously and observing interactions among them.

The assumption that PNs are unproblematic in comprehension is certainly too simplistic. Counter-evidence will appear in the near future with the growing awareness of the need to examine this issue. However, refuting the assumption does not mean questioning computing the text coverage of PN. Corpus analysis reveals the fact about the language. What should be modified is our way of interpreting the figures. Furthermore, even if the assumption that PN is not problematic is refuted, we should not go to the opposite extreme (PNs are always problematic). Many factors affect PN processing. Future studies should adopt more fine-tuned research designs to understand which PN is problematic in what context and for whom.

3 Conclusion

After reading the four papers in the Vocabulary Learning strand this year, I strongly felt the impact of information technology on L2 vocabulary research and instruction. All of the papers demonstrate the researchers' competence in implementing ever-advancing technologies in their research and education: an online app embedded in the curriculum (Mueller & Hays, 2022), multiple technical tools for corpus analysis (Uehara et al., 2022), application of AR and statistical analysis (Dabrowski, 2022), and measuring reaction times (Klassen, 2022). I was also impressed by the enthusiasm of researchers who share their ideas at the early stages of their studies. Three papers represent a research proposal, a pilot study, and a work-in-progress. Mueller and Hays's paper is based on the long-term running of their curriculum. However, considering their continued use of WordEngine in the future, this paper is considered a good start for testing and confirming the app's effect and finding to-be-solved questions. When the ideas shared in these papers fully bloom, they will make contributions to L2 vocabulary research with significant implications for vocabulary learning.

References

Dabrowski, A. (2022). Rasch-based instrument validations for an augmented reality vocabulary acquisition experiment. *Vocabulary Learning and Instruction, 11*(2), 69–93. https://doi.org/10.7820/vli.v11.2.dabrowski

Brown, D. (2010). An important assumption? The treatment of proper nouns in text coverage counts. *Reading in a Foreign Language, 22*(2), 355–361.

Erten, I. H., & Razi, S. (2009). The effects of cultural familiarity on reading comprehension. *Reading in a Foreign Language, 21*(1), 60–77.

Field, J. (2008). *Listening in the language classroom*. Cambridge University Press.

Finkbeiner, M., & Nicol, J. (2003). Semantic category effects in second language word learning. *Applied Psycholinguistics, 24*, 369–383. https://doi.org/10.1017/S0142716403000195

Hulstijn, J. H. (2015). *Language proficiency in native and non-native speakers: Theory and research*. John Benjamins.

Jeon, E. H., & In'nami, Y. (2022). *Understanding L2 proficiency: Theoretical and meta-analytic investigations*. John Benjamins.

Klassen, K. (2020). The effect of cultural familiarity on reading comprehension: An approximate replication of Erten and Razi (2009). *Reading Psychology, 41*(3), 205–228. https://doi.org/10.1080/02702711.2020.1768976

Klassen, K. (2021). Second language readers' use of context to identify proper names. *The Reading Matrix, 21*(2), 85–105.

Klassen, K. (2022). Orthographic processing of proper names: A proposal to investigate the orthographic cue for second language readers. *Vocabulary Learning and Instruction, 11*(2), 94–100. https://doi.org/10.7820/vli.v11.2.klassen

Koda, K. (2005). *Insights into second language reading*. Cambridge University Press.

Kormos, J. (2006). *Speech production and second language acquisition*. Lawrence Erlbaum Associates.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. MIT Press.

McLean, S. (2018). Evidence for the adoption of the flemma as an appropriate word counting unit. *Applied Linguistics, 39*(6), 823–845. https://doi.org/10.1093/applin/amw050

Mueller, K., & Hays, G. (2022). Vocabulary Learning Outside of the Classroom: Institutional Use of a Spaced Repetition Vocabulary Learning App. *Vocabulary Learning and Instruction, 11*(2), 48–56. https://doi.org/10.7820/vli.v11.2.mueller

Nakata, T. (2015). Effects of expanding and equal spacing on second language vocabulary learning. *Studies in Second Language Acquisition, 37*(4), 677–711. https://doi.org/10.1017/S0272263114000825

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge University Press.

Nation, I. S. P. (2016). *Making and using word lists for language learning and testing*. John Benjamins. (Technical terms glossary in Introduction is useful)

Nation, I. S. P., & Kobeleva, P. (2016). Proper nouns. In I. S. P. Nation (Ed.) *Making and using word lists for language learning and testing* (pp. 55–64). John Benjamins.

Peressotti, F., Cubelli, R., & Job, R. (2003). On recognizing proper names: The orthographic cue hypothesis. *Cognitive Psychology, 47*, 87–116. https://doi.org/10.1016/S0010-0285(03)00004-5

Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11*, 357–383. https://doi.org/10.1080/10888430701530730

Perfetti, C. A., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading, 18*, 22–37. https://doi.org/10.1080/10888438.2013.827687

Rost, M. (2011). *Teaching and researching listening* (2nd ed.). Pearson Education.

Sasaki, M. (2000). Effects of cultural schemata on students' test-taking processes for cloze tests: A multiple data source approach. *Language Testing, 17*(1), 85–114. https://doi.org/10.1177/026553220001700104

Schoonen, R., Van Gelderen, A., Stoel, R. D., Hulstijn, J. H., & De Glopper, K. (2011). Modeling the development of L1 and EFL writing proficiency of secondary school students. *Language Learning, 61*(1), 31–79. https://doi.org/10.1111/j.1467-9922.2010.00590.x

Tinkham, T. (1993). The effect of semantic clustering on the learning of second language vocabulary. *System, 3*, 371–380. https://doi.org/10.1016/0346-251X(93)90027-E

Uehara, S., Haraki, H., & McLean, S. (2022). Developing a Discipline-Specific Corpus and High-Frequency Word List for Science and Engineering Students in Graduate School. *Vocabulary Learning and Instruction, 11*(2), 57–68. https://doi.org/10.7820/vli.v11.2.uehara

Ward, J. (2009). A basic engineering English word list for less proficient foundation engineering undergraduates. *English for Specific Purposes, 28*(3), 170–182. https://doi.org/10.1016/j.esp.2009.04.001

Waring, R. (1997). The negative effects of learning words in semantic sets: A replication. *System, 2*, 261–274. https://doi.org/10.1016/S0346-251X(97)00013-4