TESL-EJ Teaching English as a Second or Foreign Language
The Electronic Journal for English as a Second Language

August 2026 – Volume 30, Number 2
https://doi.org/10.55593/ej.30118a4

The Effects of Prior Vocabulary Knowledge and Addressees in Sentence Writing Tasks: An Exploratory Study

Gaia Oikawa
Graduate School of International Cultural Studies, Tohoku University, Sendai, Japan
<oikawa.gaia.r5@dc.tohoku.ac.jp>

Abstract

This study examined the effects of prior vocabulary knowledge and the addressee of sentence writing on second language (L2) vocabulary learning among Japanese university EFL learners. Thirty-nine participants wrote example sentences for 19 target words under four conditions (general, self, family, and peer) in a within-participant design. Prior vocabulary knowledge was measured by a vocabulary size test (VST-NJ8). Vocabulary learning was assessed through meaning recall and recognition tests administered at a pretest and a posttest. The results of Bayesian generalized linear mixed-effects models showed that vocabulary size strongly predicted vocabulary learning outcomes for both recall and recognition tests. Moreover, while no clear condition effects were observed, writing for peers showed a modest positive tendency compared to writing for an unspecified addressee, especially in the meaning recall test. These findings highlight that sufficient prior vocabulary knowledge is essential for maximizing the benefits of sentence writing tasks. The findings also suggest that addressee prompts may influence the content of learners' writing, with a potential influence on vocabulary learning. This study contributes to our understanding of L2 vocabulary learning through sentence writing by highlighting the role of learner-related factors and task design, with pedagogical implications for more effective vocabulary teaching.

Keywords: sentence writing task, intentional vocabulary learning, L2 vocabulary learning, prior vocabulary knowledge, addressee of writing, addressee effect

Background

For successful vocabulary learning, not only the quantity of exposure (i.e., repetitions) but also the quality of attention to input is crucial (Webb & Nation, 2017). The quality of attention encompasses processes such as noticing, retrieval, varied use, repeated encounters, and elaboration. Among various vocabulary learning activities, sentence writing has been argued to be particularly effective in promoting varied use and semantic elaboration through a generative process with target words. By creating original sentences, learners can actively encode lexical items in meaningful contexts, which is expected to induce deeper processing and strengthen form–meaning mappings (Craik & Lockhart, 1972; Craik & Tulving, 1975). Despite this strong theoretical motivation, empirical findings regarding the effectiveness of sentence writing for vocabulary learning have been mixed. While some studies have reported positive effects of sentence writing (e.g., Pichette et al., 2012; Silva et al., 2024; Webb, 2005), others have found null or even negative effects (e.g., Barcroft, 2004). Such inconsistencies have

often been attributed to methodological differences across studies, such as learner- and task-internal factors that potentially shape how sentence writing functions as a vocabulary learning activity.

One potentially critical learner-related factor is prior vocabulary knowledge. Sentence writing requires learners not only to retrieve form–meaning mappings but also to draw on grammatical and lexical resources to construct meaningful sentences. Writing research has consistently shown that vocabulary knowledge plays a central role in writing performance (Johnson et al., 2016; Kilic, 2019). For example, Kilic (2019) reported that vocabulary knowledge accounted for approximately 26% of the variance in L2 writing performance, with productive vocabulary knowledge exerting a stronger influence than receptive knowledge. Given this close link between vocabulary knowledge and writing ability, it is likely that learners' prior vocabulary knowledge influences the writing process, thereby affecting the extent to which sentence writing facilitates vocabulary learning. Nevertheless, this possibility has rarely been examined in previous research.

Another underexplored task-related factor concerns the addressee of the sentence-writing task. Existing studies have primarily focused on the effectiveness of sentence writing per se or on comparisons with other learning activities (e.g., Folse, 2006), implicitly treating sentences as context-free products. However, writing is inherently audience-oriented, and the intended addressee may shape learners' cognitive engagement during sentence production. From a memory perspective, the self-reference effect suggests that information related to oneself is more easily remembered than information related to others (Symons & Johnson, 1997). In addition, writing for peers may also promote learning with deeper cognitive processing during writing with perspective-taking, richer episodic memory from peers, and peer interaction that can elicit discussion about target words (i.e., LREs). Despite these theoretical considerations, the role of addressee manipulation in sentence writing tasks has received little attention in research on sentence writing for vocabulary learning.

Building on these gaps, the present study investigates how prior vocabulary knowledge and the intended addressee influence the effectiveness of sentence writing tasks for L2 vocabulary learning. By examining these variables within sentence writing, this study aims to provide a more nuanced account of when and for whom sentence writing facilitates vocabulary learning.

Literature Review

Sentence Writing Tasks for Vocabulary Learning

Vocabulary learning is commonly classified into intentional and incidental learning (Webb & Nation, 2017). Intentional vocabulary learning involves deliberate efforts to learn lexical knowledge (e.g., meanings, spellings, and spoken forms), whereas incidental vocabulary learning occurs as a by-product of meaning-focused activities. Although learning efficiency differs across these types, both are considered essential for developing adequate vocabulary knowledge (Nation, 2022). Because learners in the present study were explicitly instructed to learn vocabulary through sentence writing, the learning context is best characterized as intentional vocabulary learning.

A sentence writing task requires learners to write sentences while using target words. The effectiveness of sentence writing for vocabulary learning can be theoretically grounded in two complementary frameworks: the generation effect and the levels of processing theory. From the generation-effect perspective (Slamecka & Graf, 1978; Wittrock, 1974), learning is enhanced when learners actively generate information rather than passively receive it. Since a sentence writing task requires learners to use target words in self-generated contexts, it is expected to strengthen memory representations relative to non-generative activities such as

reading example sentences. Webb and Nation (2017) also emphasized that varied and contextualized use of vocabulary enhances the quality of attention to input, thereby facilitating the establishment of robust form–meaning mappings. Empirical support for generative processing has been demonstrated by Joe (1998), who showed that generation through retelling after reading significantly facilitated incidental vocabulary learning compared to a non-retelling group.

Complementing the generation effect, levels of processing theory (Craik & Lockhart, 1972; Craik & Tulving, 1975) posits that information processed more deeply is more likely to be retained in long-term memory. Elaboration plays a central role in enhancing the depth of processing. When learners encode information, richer information about structure, phonology, and semantics is likely to strengthen the encoding (i.e., elaborated processing). Therefore, it is thought that sentence writing tasks promote semantic elaboration when learners write original sentences using target vocabulary. In addition, thinking about the situations or sentences repeatedly while using new words also strengthens memory (i.e., elaborative rehearsal). In other words, when learners think of many situations where a target word can be used, the elaborative rehearsal strengthens the memory of the target word, thereby facilitating learning. Although the present study concerns intentional vocabulary learning, research on incidental vocabulary learning, particularly studies informed by the Involvement Load Hypothesis (Laufer & Hulstijn, 2001), has also offered a useful framework for understanding the potential value of sentence-writing tasks relative to cognitively less involving tasks (e.g., reading with glosses; Kim, 2008; Zou, 2017; for a meta-analysis, see Yanagisawa & Webb, 2021).

Building on these theoretical accounts, a growing body of empirical research has examined the effectiveness of sentence-writing tasks for L2 vocabulary learning. Several studies have reported positive effects. For example, Pichette et al. (2012) found that French EFL learners who wrote sentences for target words showed significantly greater gains on a cued form recall test than those who read example sentences. Similarly, Silva et al. (2024) reported that sentence writing led to significantly greater gains than composition writing among Polish EFL learners at the B2 level or above, as measured by the Vocabulary Knowledge Scale (Wesche & Paribakht, 1996). Bao (2015) also found that sentence writing produced greater gains than a sentence-combining task, although its advantage was less clear relative to other comparison tasks. Kim (2008), moreover, showed that tasks with higher involvement loads yielded greater vocabulary gains (Experiment 1), while sentence writing produced gains comparable to those of composition writing (Experiment 2), regardless of proficiency level.

In contrast, other studies have reported null or negative effects of sentence writing. Barcroft (2004) found that learners who learned words through sentence writing performed worse on written picture-naming tests (i.e., form recall) than those who learned words through viewing, both immediately and after a delay, even when repetitions and time-on-task were controlled. Similarly, Folse (2006) reported no significant differences between fill-in-the-blank tasks and original sentence writing. Webb (2005) further demonstrated that the effectiveness of sentence writing is contingent on time allocation: when time was strictly controlled, receptive learning activities (reading example sentences) were superior, whereas productive learning activities (sentence writing) yielded greater gains when learners were given sufficient time to complete the task.

Taken together, these findings suggest that sentence writing does not uniformly facilitate vocabulary learning. Rather, its effectiveness appears to depend on task conditions and learner characteristics that shape how cognitive resources are allocated during the activity. Importantly, however, most previous studies have focused on comparisons between sentence writing and other learning activities, while leaving several theoretically relevant factors within sentence

writing itself underexplored. Specifically, two factors that are of particular interest are prior vocabulary knowledge and the intended addressee of the writing.

Prior Vocabulary Knowledge

Although research has established the importance of vocabulary knowledge for all language skills including reading, listening, speaking, and writing (Milton, 2013; Stæhr, 2008), its role in sentence writing tasks has received limited empirical attention. This lack warrants examination because sentence writing requires learners to produce original sentences using target words; learners need to draw not only on form–meaning knowledge but also on sufficient grammatical and lexical knowledge to construct accurate and meaningful sentences (Nation, 2022). Also, for lower-proficiency learners, the cognitive demands of sentence writing tasks may be overwhelming, leaving insufficient attentional resources for processing words. Empirical evidence from writing research underscores the central role of productive vocabulary knowledge in writing performance. Kilic (2019), for example, showed that the depth of vocabulary knowledge (r = .39), receptive (r = .49), and productive vocabulary knowledge (r = .48) were significantly correlated with L2 writing performance, with the depth and productive vocabulary knowledge emerging as stronger predictors in regression analyses for predicting writing performance. Similarly, Johnson et al. (2016) reported a significant correlation between productive vocabulary knowledge and writing performance (r = .38). These findings suggest that productive vocabulary knowledge is particularly critical for completing tasks that require learners to actively use lexical items, such as sentence writing. Accordingly, the extent to which sentence writing facilitates vocabulary learning is likely to be affected by learners' vocabulary knowledge.

From a theoretical perspective, the cognitive load theory supports the importance of vocabulary knowledge (Sweller, 1988; Sweller et al., 2019). Learners with limited vocabulary resources are more likely to struggle to construct sentences, as they must allocate substantial cognitive resources to lower-level processes such as retrieving grammatical structures and spelling words accurately (Kormos, 2012). Such high cognitive demands leave fewer cognitive resources available for processing and encoding new lexical items. In contrast, learners with greater vocabulary knowledge can engage in the bottom-up process and are better positioned to allocate cognitive resources to target words, thereby increasing the likelihood that sentence writing will facilitate vocabulary learning.

Despite these theoretical considerations, most previous studies on sentence writing have not incorporated prior vocabulary knowledge in their analysis. One exception is Kim (2008), which examined the main effect of learner proficiency (undergraduates vs. students in an intensive English program). The study found no significant main effect of proficiency and no significant interaction between proficiency and task involvement. However, because the participants in that study were relatively proficient, further research including lower proficiency learners is needed. This omission needs addressing, as it obscures potential variability in the effectiveness of sentence writing across learners and limits our understanding of when and for whom sentence writing is most beneficial. Addressing this gap is therefore essential for a better understanding of sentence writing as a vocabulary learning activity.

Addressee of Sentence Writing

There are several theoretical reasons why the addressee of a sentence-writing task may influence vocabulary learning. Writing is inherently audience-oriented, and the intended addressee may shape how learners conceptualize content, select lexical items, and engage cognitively with target words. One relevant mechanism is the self-reference effect, which posits that information processed in relation to the self or close others (e.g., family members) is more

strongly encoded and more easily retrieved than information processed with reference to less personally relevant entities (Symons & Johnson, 1997). This advantage is typically attributed to elaborative encoding, whereby new information is integrated into existing knowledge structures and results in robust memory representation.

Empirical evidence from vocabulary learning research supports the self-reference effect. Pruss et al. (2025), for example, demonstrated that creating self-referential sentences resulted in greater vocabulary learning gains. Similarly, Qahl and Lambert (2025) showed that words encoded with reference to the self or intimate others (i.e., mother) were better retained than words associated with less personally relevant entities, such as public landmarks or synonyms.

In contrast, writing sentences for peers may also support vocabulary learning through different, yet complementary, cognitive processes. Peer-oriented writing requires learners to consider what contexts would be meaningful for specific others, which may involve perspective-taking and the integration of contextual information. Such processes are consistent with accounts of elaborative rehearsal and deeper semantic processing (Craik & Lockhart, 1972; Craik & Tulving, 1975). Moreover, when learners have limited information about others (e.g., peers), the task may place a greater burden on sentence generation by requiring perspective-taking and may prompt them to seek additional information. In this respect, sentence writing can become more similar to source-based writing, in that learners need to obtain, organize, and integrate information into their writing. Considering source-based writing has been shown to impose greater cognitive demands than writing based solely on one's own ideas (Plakans, 2008), peer-oriented writing potentially leads to more effortful processing of target words.

In addition to semantic elaboration, peer-oriented sentence writing may enrich episodic memory representations associated with target words. To produce meaningful sentences for specific peers, learners must integrate personal information about those individuals into the sentence context. Such additional, word-related information may result in richer episodic memory traces that draw on not only the learner's own experiences but also knowledge about others. These enriched representations may function as effective retrieval cues, consistent with the encoding specificity principle (Tulving & Thomson, 1973).

From a cognitive-interactionist perspective, peer-oriented sentence writing may also afford opportunities for interaction during task completion. Research on collaborative learning suggests that interaction can facilitate vocabulary learning by encouraging learners to negotiate meaning, exchange information, and attend more closely to linguistic form (Ellis et al., 2019). Collaborative contexts also allow for peer feedback, which has been shown to support learning (Dobao, 2014). Within this line of research, language-related episodes (LREs), defined as instances in which learners discuss or reflect on language use, have been identified as one mechanism through which interaction promotes learning (Swain & Lapkin, 2001). Empirical studies have consistently shown that collaborative learning conditions yield both greater vocabulary learning gains and a higher incidence of LREs than individual learning conditions (e.g., Kim, 2008; Nassaji & Tian, 2010; Rabie-Ahmed & Mohamed, 2022).

Despite these theoretical and empirical considerations, previous studies on sentence writing tasks have primarily focused on comparisons with other learning activities, rather than examining how variations within sentence writing itself—such as differences in the intended addressee—may influence vocabulary learning outcomes. Addressing the role of the addressee is therefore essential for a more complete understanding of the mechanisms through which sentence writing facilitates L2 vocabulary learning.

The Present Study

Building on the previous studies, it is reasonable to think that prior vocabulary knowledge and the intended addressee of sentence writing might impact the effectiveness of a sentence writing task for learning L2 vocabulary. However, these specific factors have not been covered thus far in the literature. Therefore, this study aimed to examine the influence of prior vocabulary knowledge and addressee in sentence creation tasks for L2 vocabulary learning by answering the following research questions:

RQ1: To what extent does prior vocabulary knowledge influence the learning gains through the sentence writing task?

RQ2: To what extent do the different addressees of sentence writing yield different vocabulary learning gains?

Based on the literature reviewed above, two tentative hypotheses were formulated.

H1: We hypothesized that larger prior vocabulary knowledge would facilitate vocabulary learning, as previous studies have reported (e.g., Kilic, 2019).

H2: We hypothesized that addressee type would affect vocabulary learning outcomes. Specifically, sentence writing addressed to the self and family would yield greater vocabulary gains than writing addressed to peers or a general audience, consistent with the self-reference effect (Qahl & Lambert, 2025). Alternatively, we also hypothesized that writing for peers could produce greater gains than the other conditions by promoting elaborative processing through perspective-taking, negotiation for meaning, and peer feedback (Ellis et al., 2019).

Method

The materials and dataset for this study are available on the Open Science Framework (OSF): https://osf.io/fc3z5/overview.

Participants

A total of 47 Japanese first-year university students studying English as a foreign language (EFL) participated in the study. Eight participants were excluded from the final analysis due to incomplete data resulting from absence during the treatment or testing sessions (n = 7), and outliers defined as values exceeding ± 2.5 *SD* from the mean on the posttests of vocabulary outcome measures (n = 1) (Winter, 2019). Consequently, data from 39 participants were retained for statistical analysis. The sample size reflects the constraints of an intact classroom setting. Accordingly, the present study is intended to provide an initial exploratory investigation regarding addressee effects in sentence-writing tasks rather than definitive estimates with broad generalizability.

All participants had received an English education from elementary school in Japan. Participants' English proficiency ranged from A2 to B1 of the Common European Framework of Reference for Languages (CEFR), estimated based on the results of vocabulary size test scores. The participants had English classes twice a week as a mandatory subject in their curriculum. The primary learning objective of the course was to prepare students for the Test of English for International Communication (TOEIC®).

Sentence Writing Task

Participants were required to write a sentence for each target word under different conditions (i.e., for a general audience, themselves, family, and peers) (for an assignment procedure, see Procedure section). The selection of addressees in the present study was theoretically and

empirically motivated. The self and family conditions were included based on evidence that these addressees reliably elicit the self-reference effect, which has been shown to facilitate vocabulary learning. In particular, Qahl and Lambert (2025) demonstrated that word encoding while thinking about the self and intimate others (e.g., family) led to significantly better retention of target words, thereby providing an empirical rationale for including these addressees in the present sentence-writing paradigm. In contrast, the peer condition was motivated by theoretical frameworks emphasizing interaction, perspective-taking, and memory-based elaboration as reviewed above. Writing for peers may prompt learners to consider others' knowledge and experiences, thereby engaging cognitive and social processes distinct from self-referential encoding. In addition, peer-oriented tasks are common in classroom settings, enhancing the ecological validity of this condition. Finally, a general addressee condition was included as a baseline comparison. This condition involved writing sentences for a non-specific audience and closely resembled the sentence-writing tasks used in previous research. As such, it provided a baseline against which the effects of more personally or socially oriented addressees could be evaluated.

In the sentence writing task, participants were provided with a worksheet and completed the task via Google Docs (see Table 1). The worksheet included explicit instructions for each condition, along with a mini-dictionary of target words including written forms, L1 meanings, example phrases with L1 meanings, and part of speech. The sentence writing task was conducted individually. However, since classroom seating was not fixed, some participants were unfamiliar with their peers. To facilitate sentence creation in the peer condition, they were allowed to ask their peers for personal information when necessary (e.g., club activities, part-time jobs). Such interaction was limited to short exchanges during sentence writing for the Peer condition and did not occur continuously throughout the task. During these exchanges, learners sometimes referred to peers' personal information that could be useful for sentence creation and, when needed, confirmed the meanings of target words using the mini-dictionary provided in the worksheet. Representative sentence examples are presented in Table 2 (for additional anonymized examples sampled across proficiency levels, see Supplemental Material S2).

Google Docs' automated proofreading function provided immediate corrective feedback on spelling and grammatical errors, as well as suggestions for surface-level replacements. This function was intended to help participants focus on sentence creation (i.e., content generation) and to minimize potential confounding influence from misspelling or grammatical errors during learning. Unlike generative AI tools, the proofreading function did not generate sentences or content for learners; instead, feedback was provided only after learners had produced their own sentences. Opportunities to use the proofreading function were equivalent across participants and conditions, as all participants completed the sentence-writing task using the same Google Docs worksheet with identical settings in a within-participant design. Edit logs showed a mean of 23.13 revision episodes per participant (SD = 2.40, 95% CI [22.38, 23.88], range = 20–29).

To examine whether the addressee manipulation was reflected in learner output, a brief descriptive inspection of the sentences was conducted. Systematic differences in sentence characteristics were observed across conditions. In the Self condition, sentences frequently began with the first-person pronoun *I* and included possessive expressions such as *my*, indicating a high degree of personalization. In contrast, sentences in the Family and Peer conditions more often referred to the intended addressee through kinship terms (e.g., *my brother*) or third-person pronouns (*he or she*), resulting in more descriptive statements about the addressee. In addition, sentences in all conditions except the General condition tended to include proper nouns (e.g., school names) or other concrete situational information, whereas sentences in the General condition were comparatively less specific and more impersonal in

nature. Together, these patterns suggested that learners generally adapted their writing in ways consistent with the intended addressee (see Table 2 and Supplemental Material S2 for additional examples). These qualitative differences also suggest that the addressee prompts may have influenced not only surface-level wording but also the type of information learners drew on in sentence writing. This, in turn, suggests that the task conditions were meaningfully differentiated, providing a basis for possible addressee effects to emerge in learner output.

To further verify the manipulation, the author, an experienced EFL teacher, and another experienced EFL teacher independently evaluated whether each sentence aligned with the intended addressee. Inter-rater reliability was very high (Cohen's $\kappa = .92$), with a percent agreement of 99.3%, indicating almost perfect agreement. These results provide additional support that the addressee manipulation was implemented as intended.

Table 1. Conditions for Making Example Sentences with Target Words

Condition 1 *(General)*	Make an example sentence that might be useful and understandable for everyone that can be used for a published word book.
Condition 2 *(Self)*	Make an example sentence that might be useful for you.
Condition 3 *(Family)*	Make an example sentence that might be useful for your family (e.g., father, mother, grandparents).
Condition 4 *(Peer)*	Make an example sentence that might be useful for your peer.

Table 2. Representative Examples Illustrating the Addressee Manipulation Across Conditions

Condition 1 (General)	He and she have an intimate relationship.
Condition 2 (Self)	I forfeited my earphones.
Condition 3 (Family)	My brother is a Tohoku University undergraduate.
Condition 4 (Peer)	He knows cozy restaurants.

Vocabulary Size Test Based on the New JACET List of 8,000 Basic Words

Given that the present participants were Japanese learners of English as a foreign language (EFL), this study employed a vocabulary size test for Japanese learners using the New JACET List of 8,000 Basic Words (VST-NJ8: Hamada et al., 2021). VST-NJ8 was developed for measuring vocabulary size (i.e., how many words a learner knows) using a frequency-based sampling method from the New JACET List of 8,000 Basic Words (Committee of revising the JACET basic words, 2016). The New JACET List is a word list specifically designed for Japanese university-level EFL learners, reflecting contemporary language use in academic and daily communication contexts. The test contains 160 items, drawn from eight frequency levels, and was constructed and validated through item response theory modeling. Test takers need to select a Japanese translation with its correct English word from four options. The finalized version has shown to possess strong construct validity, including content, structural, generalizability, and external aspects. It also correlated with standardized English proficiency scores significantly ($r = .62$). The VST-NJ8 provides a reliable and validated estimate of learners' vocabulary breadth. For the present participants, $M = 118.02$, $SD = 17.69$, range = [43-144], indicating that, on average, participants had mastered the 5,000- to 6,000-word level of the JACET 8000.

Target Words

In this study, we used 19 English words as the target words, which consisted of seven verbs, six nouns, and six adjectives (see Supplemental Material S1). These words were drawn from the British National Corpus/Corpus of Contemporary American English (BNC/COCA) wordlists, specifically ranging from the most frequent 4,000 to 8,000 levels. This selection criterion ensured that the words were likely unfamiliar to participants while remaining analyzable in terms of morphological structure. Moreover, each target word appeared in

textbooks or word books for the TOEIC®, which ensured that the learning through this study was beneficial for participants. We excluded loanwords, highly transparent compounds, and words with readily identifiable Japanese cognates to minimize the potential influence of prior L1 knowledge. After selecting the candidate target words, we conducted a pilot test with two advanced-level Japanese English learners, one of whom was teaching English and the other had majored in English education, to confirm that the words were likely to be unknown to our participants.

Procedure

In the first week, students were informed about the study procedures. The instructor explained the purpose of the study and the learning tasks, which formed part of regular classroom activities. It was made clear that, although all students completed these activities as part of the course, consent for the use of their anonymized data for research purposes was entirely voluntary and had no bearing on grades or academic standing. Students could decline or later withdraw consent for research use of their data without penalty. Following a question-and-answer session, written consent for the use of anonymized data was obtained. Under the institution's guidelines for research involving human participants, formal ethics review was not required because the project was based on regular coursework and used instructional content aligned with the course objectives (e.g., business English and TOEIC® preparation). All procedures complied with those guidelines.

After agreeing to participate, students completed the VST-NJ8 to assess their vocabulary size. After a 10-week interval, they took the pretest and completed the sentence writing task. This study employed a within-participant, item-based design (see Table 3). Nineteen target words were used, and each student produced sentences under all four conditions. Students were assigned to one of four groups, which determined how the target words were distributed across conditions. Within each group, the 19 target words were divided into four subsets (5, 5, 5, and 4 words), with each subset assigned to one of the four conditions. The assignment of target words to conditions was systematically rotated across the four groups such that each word appeared in each instructional condition across students. Although one condition within each student contained one fewer target word, this minor imbalance was counterbalanced across the four groups. As a result, the overall number of data points collected for each condition was comparable (see Table 3). Because the analyses used item-level mixed-effects models rather than condition-level averages, this imbalance was not expected to bias the estimates (for more details, see Data Analysis). This counterbalanced rotation of target words across conditions reduced the risk that condition effects would be confounded with item-specific difficulty. One week after completing the task, the students completed the posttests.

Table 3. Assignment of Target Words (TWs) Across Groups and Conditions

	Group 1	Group 2	Group 3	Group 4
General	TW 16 - 19	TW 11 - 15	TW 6 - 10	TW 1 - 5
Self	TW 1 - 5	TW 16 - 19	TW 11 - 15	TW 6 - 10
Family	TW 6 - 10	TW 1 - 5	TW 16 - 19	TW 11 - 15
Peer	TW 11 - 15	TW 6 - 10	TW 1 - 5	TW 16 - 19

Outcome Measures

This study employed meaning recall and meaning recognition tests. The meaning recall test was conducted followed by the meaning recognition test to prevent learners from seeing the meanings of the target items (Nation & Webb, 2011). These tests help capture the learning of form-meaning mapping with varying degrees of sensitivity from partial to fuller knowledge: the recall test measures more robust, productive vocabulary knowledge, whereas the recognition test assesses more partial receptive vocabulary knowledge. (Durrant et al., 2022;

Kremmel & Schmitt, 2016). Considering that each participant used each target word in different contexts with varying topics and forms (e.g., tense, aspect, number), this study employed decontextualized tests. It should be noted, however, that the decontextualized tests cannot measure the contextualized *use* aspect of vocabulary knowledge (see Nation, 2022, for more details).

The meaning recall test required participants to write the meanings of target items in Japanese (L1). In the multiple-choice meaning recognition test, participants selected the most appropriate Japanese translation from five options: one correct answer, three distractors, and an "I do not know" option. The inclusion of the "I do not know" option was intended to reduce random guessing in multiple-choice vocabulary tests (Zhang, 2013). Distractors were constructed according to five criteria: (a) the same part of speech as the target word, (b) similar word length or syllable counts, (c) the same semantic category (e.g., emotion-related terms or verbs of removal), (d) non-rare Japanese words for university-level learners, and (e) one distractor per item corresponded to another target word to detect inattentive responding. A pilot version was administered to the two advanced-level Japanese English learners to ensure the plausibility and clarity of distractors. After the pilot test, we revised some distractors based on the feedback to avoid potential confounding influences from distractors. Cronbach's α for the meaning recall tests was .65 (pretest) and .74 (posttest), and for the meaning recognition test, it was .59 (pretest) and .59 (posttest). Given that Cronbach's alpha for the meaning recognition measure was below the conventional benchmark, we conducted a follow-up reliability analysis using intraclass correlation coefficients (ICCs) derived from mixed-effects models. The results indicated the presence of systematic participant- and item-level variance, but the overall magnitude of these reliability estimates was modest (see Supplemental Material S4 for more details). Accordingly, the recognition measure may have had limited sensitivity to detect small condition differences, and weak or null effects should be interpreted with caution.

Scoring

Answers were marked dichotomously in both the meaning recall and recognition tests (0 = incorrect and 1 = correct). For the meaning recall test, answers were awarded a point when both the part of speech and translation in L1 were appropriate. For the meaning recognition test, a point was provided for the correct meaning. To ensure the reliability of scoring on the meaning recall test, we calculated inter-rater reliability with Cohen's Kappa. The author and a graduate student majoring in second language acquisition who was also working as a high school English teacher independently scored all responses. Interrater consistency was then calculated. The result confirmed consistently high reliability for both the pretest (κ = .96) and the posttest (κ = .97).

Data Analysis

Preliminary analyses were conducted to confirm that learning occurred through the treatment. Prior to the analyses, assumptions of homogeneity of variance (Levene's test) and normality (Shapiro-Wilk test) were examined. For the pretest scores, Shapiro-Wilk tests indicated violations of normality, consistent with a pronounced floor effect at pretest. Accordingly, we employed Welch's one-way ANOVA as a heteroscedasticity-robust parametric alternative to compare pretest scores across conditions. For the pre–post comparisons, Shapiro-Wilk tests on the difference scores suggested non-normality in some conditions; therefore, we used a non-parametric paired test (Wilcoxon signed-rank test). Details of the assumption checks and results of the tests are provided in Supplemental Materials S3 and S5.

Following the preliminary analyses, the effects of condition (General, Self, Family, Peer) on vocabulary learning were examined using Bayesian generalized linear mixed-effects models

(GLMMs) implemented in R (Version 4.5.2) with the *brms* package. Bayesian modeling was adopted for three reasons. First, Bayesian inference provides directly interpretable probability statements about parameters through credible intervals (CrIs), allowing the strength and direction of effects to be evaluated beyond dichotomous significance testing. Second, Bayesian models are well-suited to classroom-based research with relatively modest sample sizes because the use of weakly informative priors stabilizes estimation and regularizes extreme values. Third, Bayesian inference supports cumulative science by allowing prior knowledge to be explicitly incorporated into current analyses, facilitating transparent accumulation of evidence across studies (Wong, 2025). These features make Bayesian modeling particularly appropriate for the present exploratory investigation of addressee effects in sentence-writing tasks.

We fitted Bayesian generalized linear GLMMs separately for meaning recall and recognition tests. Both outcome variables were binary (incorrect = 0, correct = 1) and were modeled using a Bernoulli distribution with a logit link. Fixed effects included Condition, Vocabulary Size (VST-NJ8: standardized), and pretest performance (standardized). Condition was treated as a categorical predictor and dummy-coded, with the General condition specified as the reference level (General, Self, Family, Peer). Although both the VST-NJ8 and pretest scores reflect learners' prior vocabulary knowledge, they capture conceptually distinct aspects of that knowledge. The VST-NJ8 indexes learners' general vocabulary size, representing general lexical knowledge across a broad range of words, whereas the pretest scores reflect item-specific prior knowledge of the target words used in the experiment. Including both measures as covariates, therefore, allows us to disentangle general lexical proficiency from familiarity with the specific target items, thereby providing a more precise estimate of the effect of Condition (Maie et al., 2024). To address potential concerns about multicollinearity, we examined variance inflation factors (VIFs) for all fixed effects. Because conventional VIFs can be inflated in models including interaction terms, adjusted VIFs were also considered. The adjusted VIFs were below 3 in both models, indicating that multicollinearity was unlikely to be problematic. Random intercepts were specified for participants (ID) and items, and by-participant random slopes for Condition were included. This specification estimates participant-specific deviations for each condition while constraining correlations among random effects to zero, which improves model stability in designs with limited observations per participant.

Weakly informative priors were specified for all parameters to regularize estimation while avoiding overly strong assumptions. Models were estimated by using four chains, and standard convergence diagnostics indicated satisfactory convergence for all parameters. Posterior predictive checks suggested adequate model fit for both recall and recognition outcomes. Model robustness was examined through sensitivity analyses with alternative specifications and prior distributions. These analyses yielded comparable results, and leave-one-out cross-validation indicated no meaningful differences in model performance. For reasons of parsimony, only the baseline models are reported here; to ensure transparency and reproducibility, full model specifications, priors, diagnostics, and analysis codes are provided in the OSF repository.

Results are reported as posterior medians (β) with 95% credible intervals (CrIs). For ease of interpretation, odds ratios (*OR*s) and corresponding 95% intervals are also provided. In addition, we report the probability of direction (pd), which reflects the proportion of the posterior distribution supporting the estimated effect's direction, and the region of practical equivalence (ROPE), which quantifies the proportion of the posterior distribution that falls within a range considered practically negligible. Larger ROPE proportions indicate that an effect is likely to

be small or negligible, whereas smaller ROPE proportions indicate stronger evidence for a non-trivial effect.

Results

Descriptive Statistics and Preliminary Analysis

Preliminary analyses indicated no significant differences in pretest scores among the four conditions for either meaning recall or meaning recognition (Welch's one-way ANOVA, *p*s > .05; see Supplemental Material S5). Wilcoxon signed-rank tests further showed significant improvements from pretest to posttest across all conditions for both outcome measures (all *p*s < .05), with medium-to-large effect sizes for both the meaning recall and recognition tests (Plonsky & Oswald, 2014). These results suggest that learners did not differ in prior vocabulary knowledge of the target items before the learning, and developed their vocabulary knowledge through the intervention (for detailed statistics, see Supplemental Material S3 and S5).

Descriptive statistics showed that the Peer condition yielded the highest mean accuracy rate for both meaning recall and meaning recognition (see Table 4 and 5). For meaning recall, the Peer condition showed the highest mean accuracy, followed by the Family and Self conditions. The General condition showed the lowest accuracy.

Table 4. Descriptive Statistics of the Accuracy Rate for Meaning Recall Scores

	Pre			Post		
	Mean (*SD*)	**95% CI**	**Range**	**Mean (*SD*)**	**95% CI**	**Range**
General	.010 (.045)	[−.004, .025]	.00–.20	.059 (.115)	[.021, .097]	.00–.40
Self	.032 (.100)	[.000, .065]	.00–.40	.071 (.120)	[.032, .111]	.00–.40
Family	.022 (.066)	[.001, .043]	.00–.25	.091 (.130)	[.048, .134]	.00–.40
Peer	.021 (.077)	[−.004, .045]	.00–.40	.113 (.139)	[.067, .159]	.00–.40

Note. Each condition included observations from 39 participants. 0.10 = 10%.

For meaning recognition, the Peer condition again showed the highest accuracy, followed by the Self, General, and Family conditions. The difference in recognition accuracy between the Peer and General conditions was 6.5% (59.9% vs. 53.4%), whereas the difference between the Self and General conditions was 5.9% (59.3% vs. 53.4%).

Table 5. Descriptive Statistics of the Accuracy Rate for Meaning Recognition Scores

	Pre			Post		
	Mean (*SD*)	**95% CI**	**Range**	**Mean (*SD*)**	**95% CI**	**Range**
General	.265 (.289)	[.172, .359]	.00–1.0	.534 (.268)	[.446, .622]	.00–1.0
Self	.295 (.250)	[.214, .376]	.00–1.0	.593 (.246)	[.512, .674]	.00–1.0
Family	.260 (.215)	[.190, .330]	.00–.80	.533 (.277)	[.442, .624]	.00–1.0
Peer	.279 (.239)	[.202, .357]	.00–.80	.599 (.231)	[.523, .675]	.00–1.0

Note. Each condition included observations from 39 participants. 0.10 = 10%

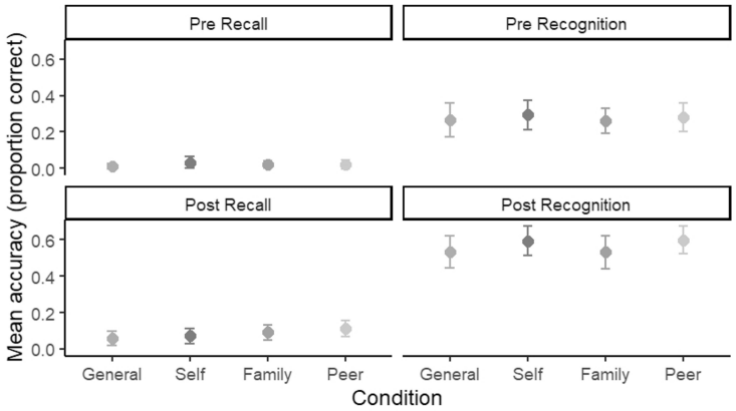

Figure 1. Mean and 95% CI for the Accuracy Rate by Test Timing and Condition

Bayesian Modeling

As shown in Table 6, vocabulary size (VST-NJ8) showed a robust positive effect on meaning recall accuracy, providing strong support that learners with larger existing lexical knowledge demonstrated higher recall accuracy. In contrast, the main effects of Condition (i.e., addressees) were modest (see Figure 2). None of the condition coefficients provided decisive evidence, as all 95% CrIs included zero and showed non-negligible overlap with the ROPE. Nevertheless, compared to the reference level (i.e., the General condition), the Peer condition exhibited a consistent positive tendency relative to the other conditions. The Family condition showed a weaker positive tendency, whereas the Self condition showed little evidence of a systematic effect. Post hoc pairwise comparisons based on *ORs* further indicated that none of the contrasts between conditions showed a clear difference, as all 95% credible intervals for *OR*s included one (see Supplemental Material S6).

Table 6. Posterior Estimates for Fixed Effects in the Bayesian GLMMs for the Meaning Recall Test

Predictor	β [95% CrI]	*OR* [95% CrI]	pd (%)	ROPE (%)
Intercept	−3.71 [−4.78, −2.66]	0.02 [0.01, 0.07]	100.00	0.00
Self	−0.09 [−1.04, 0.83]	0.91 [0.35, 2.29]	57.06	16.95
Family	0.29 [−0.68, 1.23]	1.34 [0.51, 3.42]	72.91	14.38
Peer	0.63 [−0.26, 1.52]	1.88 [0.77, 4.57]	92.01	7.12
VST-NJ8	0.72 [0.27, 1.21]	2.05 [1.31, 3.36]	100.00	0.00
Pretest	0.43 [0.21, 0.68]	1.54 [1.23, 1.97]	100.00	0.00

Note. β = posterior median (logit scale). OR = exp(β). CrI = 95% credible interval. pd = probability of direction. ROPE = proportion of posterior within the region of practical equivalence (i.e., logit ±0.1).

Figure 2. Mean and 95% Credible Intervals for the Predicted Probability of Correct Answer in the Meaning Recall Test

For the meaning recognition test (see Table 7), vocabulary size positively predicted accuracy, indicating that learners with greater vocabulary knowledge were more likely to correctly recognize target word meanings. With respect to Condition, the effects were again modest (see Figure 3). None of the condition coefficients reached a decisive level of evidence, as all 95% CrIs included zero and showed substantial overlap with the ROPE. However, the Peer condition showed the strongest positive tendency among the four conditions. The Self condition also showed a positive tendency, whereas the Family condition showed little evidence of a systematic effect. Estimated marginal means revealed a consistent ordering of recognition accuracy across conditions, with the Peer condition showing the highest predicted probability of correct recognition, followed by the Self, General, and Family conditions. Post hoc analysis indicated that none of the contrasts between conditions showed a clear difference, as all 95% credible intervals for *OR*s included one (see Supplemental Material S6).

Table 7. Posterior Estimates for Fixed Effects in the Bayesian GLMMs for the Meaning Recognition Test

Predictor	β [95% CrI]	*OR* [95% CrI]	pd (%)	ROPE (%)
Intercept	0.27 [−0.30, 0.84]	1.30 [0.74, 2.32]	82.31	18.99
Self	0.17 [−0.38, 0.73]	1.19 [0.69, 2.07]	72.20	24.37
Family	−0.04 [−0.62, 0.54]	0.96 [0.54, 1.72]	56.59	27.14
Peer	0.25 [−0.28, 0.80]	1.28 [0.76, 2.22]	81.76	19.99
VST-NJ8	0.30 [0.04, 0.59]	1.36 [1.04, 1.80]	98.84	4.14
Pretest	0.59 [0.37, 0.82]	1.81 [1.45, 2.27]	100.00	0.00

Note. β = posterior median (logit scale). *OR* = exp(β). CrI = 95% credible interval. pd = probability of direction. ROPE = proportion of posterior within the region of practical equivalence (i.e., logit ±0.1).

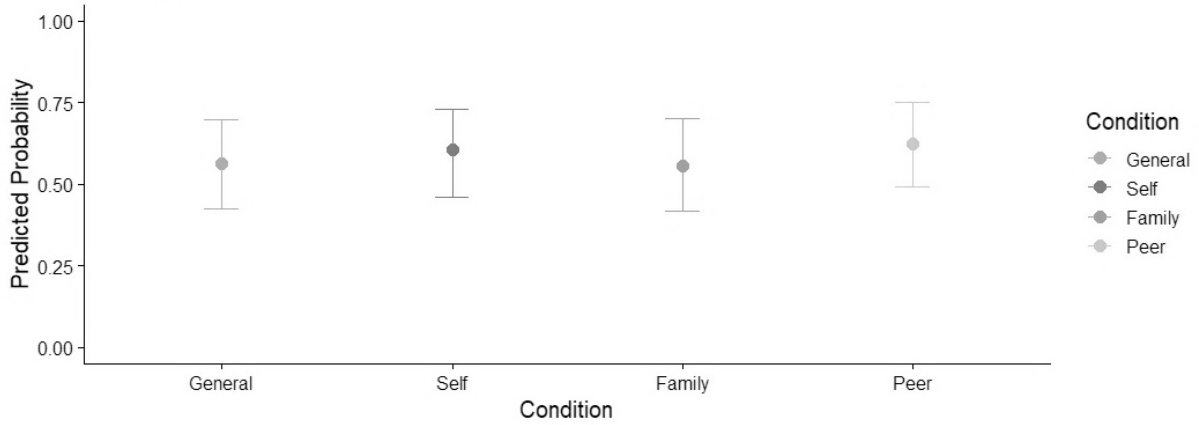

Figure 3. Mean and 95% Credible Intervals for the Predicted Probability of Correct Answer in the Meaning Recognition Test

To examine whether the data cleaning procedure influenced the results, we also fitted the Bayesian models with the same structure using the unscreened dataset that retained the data excluded in the primary analyses. The overall pattern of results remained substantively unchanged, and the main conclusions were identical to those reported in the primary analyses.

Discussion

The present study examined the effects of prior vocabulary knowledge (i.e., vocabulary size) as well as the addressee effects in the sentence writing task for L2 vocabulary learning. The results showed that vocabulary size was a strong predictor in both meaning recall and recognition tests, underscoring the crucial role of lexical knowledge in shaping the effectiveness of sentence writing tasks for vocabulary acquisition. Regarding the addressee

effects, there was no clear evidence of condition difference as the 95% CrIs for the β estimates included zero across all contrasts, although writing for peers showed a modest positive tendency, especially in the meaning recall test. However, this result should be interpreted cautiously, as any peer-related advantages remained limited and inconclusive. In addition, although learning occurred, the absolute gains, especially in meaning recall, were modest.

Regarding RQ1, the importance of prior vocabulary knowledge aligns with earlier findings that emphasize its role in productive language use (Milton, 2013; Stæhr, 2008). In sentence writing tasks, learners need to not only map form to meaning but also draw upon additional vocabulary and grammatical knowledge to construct sentences. Learners with limited vocabulary resources are more likely to struggle with bottom-up processes including leveraging grammatical knowledge and recalling form-meaning mappings of not only target items but also other words for sentence construction, which increases cognitive load and diverts attention away from consolidating target word meanings (Kormos, 2012; Sweller, 1988; Sweller et al., 2019). Indeed, although a mini-dictionary used in this study included spellings, meanings, and example phrases, and Google Docs provided automated corrective feedback that helped learners concentrate on the sentence writing task, the VST-NJ8 score was a strong predictor, suggesting that learners with lower lexical proficiency may have experienced difficulty in sentence composition. In other words, learners with greater vocabulary knowledge were better positioned to benefit from the task and demonstrated superior performance on both recall and recognition tests. This point is particularly important because previous research on sentence writing has rarely examined prior vocabulary knowledge directly. One partial exception is Kim (2008), which reported no main effect of proficiency and no interaction between proficiency and task involvement, whereas the present study showed that prior vocabulary knowledge was a strong predictor of learning in the sentence-writing task. One possible explanation is that the present study included relatively lower proficiency learners, making the role of prior vocabulary knowledge more pronounced. The present findings, therefore, contribute to the literature by suggesting that sentence-writing tasks may not be equally effective for all learners, but rather may depend in part on learners' pre-existing lexical resources.

Compared to relevant previous studies that investigated the pedagogical influence of the sentence writing task, the present study provides a key theoretical insight by explicitly demonstrating the impact of prior vocabulary knowledge on the efficacy of sentence writing. Previous studies on the sentence writing tasks reported mixed results: some reported positive effects (e.g., Pichette et al., 2012; Silva et al., 2024), but others did not (e.g., Barcroft, 2004; Folse, 2006). Based on the present results, one potential account for the inconsistency in the previous studies may be the lack of consideration of prior vocabulary knowledge. Studies that reported null or negative effects often included lower-proficiency learners or imposed strict time constraints, possibly increasing the cognitive demands and reducing task effectiveness (i.e., resource-dispersing variables, Robinson, 2021). Our findings suggest that sentence writing tasks may be beneficial, especially for learners with sufficient lexical foundations to support higher-level processing.

Regarding RQ2, although the effects of the condition were inconclusive, it is worth noting that writing for peers showed a modest positive tendency. One possible reason is that the condition might engage additional cognitive processes compared to writing for a general audience. From a levels-of-processing perspective (Craik & Lockhart, 1972; Craik & Tulving, 1975), writing for an unspecified audience (i.e., the General condition) might primarily involve considerations of clarity and intelligibility. By contrast, writing for peers may require learners to consider how target words could be meaningful for specific others, which could involve perspective-taking as well as the integration of relevant contextual information. Such elaborative processes could, in principle, increase cognitive engagement during task completion (Plakans, 2008).

From a memory-based perspective, the peer-related information used for sentence construction may have enriched learners' episodic memory. Because learners were required to create sentences by referring to information about their peers in the Peer condition, they may have obtained richer episodic memory associated with the target words, drawing not only on their own experiences but also on those of their peers. This episodic enrichment could have provided additional retrieval cues (Tulving & Thomson, 1973). Importantly, however, these explanations remain speculative, as the present study did not directly measure perspective-taking, interactional processes, or episodic memory. Therefore, any observed audience-related effects should be interpreted cautiously.

From a cognitive-interactionist perspective, another possible explanation for the observed effects in the Peer condition is that writing sentences for peers may have increased opportunities for interaction during task completion. In the present study, participants occasionally consulted their peers to obtain information needed to construct sentences using the target items. Such interactions may have prompted learners to revisit target words by recalling their meanings and considering how relevant information could be integrated into sentences, giving rise to lexical LREs. These interactional processes may have contributed to repeated access to form–meaning mappings and additional exposure to the target words (Kim, 2008; Nassaji & Tian, 2010; Rabie-Ahmed & Mohamed, 2022). This interactional account may also help explain the difference between the Peer and Family conditions. Although both involved writing for specific others, the Family condition likely afforded more readily available content, as learners already knew relevant information about their family members, thereby reducing the need for information-seeking interaction. By contrast, the Peer condition may have been more likely to create information gaps that encouraged information exchange and co-construction.

At the same time, the present findings should be interpreted in light of the study design. The Peer condition differed from the other conditions not only in addressee specification but also in the extent to which it afforded peer information exchange during task completion. Accordingly, any differences observed in the Peer condition should not be interpreted as a pure addressee effect; rather, they likely reflect a broader peer-oriented task configuration, such as interlocutor familiarity (Liu et al., 2025; Zabihi & Ghahramanzadeh, 2022). Future research should therefore examine more explicitly how interactional variables influence vocabulary learning.

With respect to the different results from previous studies on the self-reference effect in vocabulary learning (Pruss et al., 2025; Qahl & Lambert, 2025), several methodological differences may help explain the difference. First, compared to Qahl and Lambert (2025), who employed a receptive encoding task, the present study adopted a productive output task. Whereas Qahl and Lambert required participants to process self-, family-, or other-relevant information while viewing target words, the present task additionally required learners to generate sentences using the target words. Given that sentence writing requires not only knowledge of target words but also grammatical and discourse-level knowledge to compose well-formed sentences, these differences in task demands may have influenced how the self-reference effect contributed to vocabulary learning outcomes. Second, compared to Pruss et al. (2025), who employed an oral production task, the modality of output may also have played a role. Because the present study involved written sentence production, learners may have had more time to plan and reflect on their output. In addition, written output can function as an external memory resource, potentially reducing cognitive load and freeing cognitive resources for lexical processing. Such modality-specific affordances have been argued to support language learning in written production tasks (Ishikawa & Suzuki, 2016; Suzuki, 2012), and may partly account for differences between the present findings and those of prior studies.

Pedagogical Implications

The present findings suggested several pedagogical implications for implementing sentence-writing tasks in vocabulary instruction, particularly with respect to task set-up, scaffolding, audience prompts, and time-on-task. From a task set-up and scaffolding perspective, teachers may need to consider learners' existing lexical knowledge before introducing sentence-writing tasks. The present results showed that learners with larger vocabulary sizes benefited more from sentence writing, whereas learners with more limited lexical knowledge may struggle with sentence composition and allocate fewer attentional resources to target words. Accordingly, sentence-writing tasks may be most effective when learners have already established basic form–meaning mappings for the target items. Preparatory activities, such as brief pre-task vocabulary reviews or providing example sentences, may help scaffold learners' use of vocabulary and reduce the cognitive load associated with lexical access. By alleviating demands related to lower-level processes (e.g., retrieving word forms, meanings, or grammatical structures), such scaffolding can allow learners to devote greater attentional resources to sentence construction itself, thereby increasing the likelihood that they will benefit from deeper processing associated with the task.

With respect to addressee prompts, the findings suggest that specifying a concrete addressee may influence how learners engage with sentence-writing tasks and what they are likely to write. When teachers aim to highlight the kinds of contexts in which a word is appropriately used (e.g., register), specifying an addressee may be beneficial. In addition, because the potential benefits of specifying an addressee may involve memory activation and peer interaction during task completion, allocating sufficient time-on-task may further support lexical processing during sentence writing (Webb, 2005). In the present study, the sentence-writing task required approximately 40 minutes for 19 target words and, in some cases, involved learners seeking additional information about their peers. In regular classroom settings, however, teachers may adapt this procedure by distributing sentence writing across shorter segments, assigning it as homework, or providing brief, pre-selected prompts about peers to reduce time and information-seeking demands. Such adaptations may help teachers balance the potential benefits of addressee-oriented sentence writing with typical classroom constraints.

Conclusion

The present study examined how prior vocabulary knowledge and different addressees in sentence-writing tasks influence vocabulary learning. The findings suggested the possibility of addressee-related differences, with the Peer condition showing a modest positive tendency in the meaning recall test. In addition, the results demonstrated that prior vocabulary knowledge played a critical role in learning outcomes for both meaning recall and recognition, underscoring the importance of a solid lexical foundation when implementing productive vocabulary learning tasks. Taken together, the present study extends previous research by indicating that variability in task design (i.e., addressee) and learner-related variables may influence vocabulary learning outcomes in a sentence-writing task.

Despite these contributions, several limitations should be addressed in future research. First, participants in the present study produced only one sentence per target word. Given that repetition has been identified as a key factor in vocabulary learning (Uchihara, 2023), future studies should examine the effects of repeated sentence-writing practice to capture longer-term learning trajectories. Second, the sample size was relatively small (N = 39), which limits statistical power and the generalizability of the findings. Although this study employed Bayesian modeling to mitigate some of the challenges associated with small-sample estimation (e.g., relatively low power), future research should examine addressee effects in sentence-

writing tasks using larger and more diverse samples to improve external validity. Third, this study measured learning gains using meaning recall and recognition tests. Because vocabulary knowledge is multifaceted (Nation, 2022) and sentence writing is inherently a contextualized learning activity, future studies should include contextualized outcome measures to more holistically assess learning (Webb, 2005). Fourth, although the present study discussed potential mechanisms underlying the benefits of peer-oriented writing, interactional processes (e.g., language-related episodes) were not directly observed. Future research should therefore examine these processes more explicitly to clarify the mechanisms through which addressee effects may emerge. Finally, although peer pairs were assigned mechanically in the present study, differences in learners' familiarity or perceived closeness with their peers were not directly measured. Such differences may influence peer interaction processes (e.g., information exchange), which in turn could affect the content of the sentences produced and the richness of episodic memory associated with the target words (Liu et al., 2025; Zabihi & Ghahramanzadeh, 2022). Future research should therefore manipulate or assess peer familiarity to examine how interpersonal closeness moderates the effects of peer-oriented sentence writing on vocabulary learning.

Acknowledgments

The author is grateful to all participants for their cooperation. The author would also like to thank Dr. Ha Hoang and anonymous reviewers for their insightful and constructive comments on this study. Finally, the author would like to express sincere gratitude to Masato Sugawara for his assistance in preparing rater data for this study.

About the Author

Gaia Oikawa is a PhD student in the Graduate School of International Cultural Studies at Tohoku University, Japan. His research focuses on second language vocabulary learning and instruction, with a particular interest in learner engagement. His recent work has appeared in *TESOL Quarterly*, *International Journal of Applied Linguistics*, and *Studies in Second Language Acquisition*. ORCID ID: 0009-0002-0811-5921

To Cite this Article

Oikawa, G. (2026). The effects of prior vocabulary knowledge and addressees in sentence writing tasks: An exploratory study. *Teaching English as a Second Language Electronic (TESL-EJ), 30*(2). https://doi.org/10.55593/ej.30118a4

References

Bao, G. (2015). Task type effects on English as a foreign language learners' acquisition of receptive and productive vocabulary knowledge. *System*, *53*, 84–95. https://doi.org/10.1016/j.system.2015.07.006

Barcroft, J. (2004). Effects of sentence writing in second language lexical acquisition. *Second Language Research*, *20*(4), 303–334. https://doi.org/10.1191/0267658304sr233oa

Committee of revising the JACET basic words. (2016). *The new JACET list of 8,000 basic words*. Kirihara Shoten.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*(3), 268–294. https://doi.org/10.1037/0096-3445.104.3.268

Dobao, A. F. (2014). Vocabulary learning in collaborative tasks: A comparison of pair and small group work. *Language Teaching Research*, *18*(4), 497–520. https://doi.org/10.1177/1362168813519730

Durrant, P., Siyanova-Chanturia, A., Kremmel, B., & Sonbol, S. (2022). *Research methods in vocabulary studies* (1st ed.). John Benjamins Publishing Company.

Ellis, R., Skehan, P., Li, S., Shintani, N., & Lambert, C. (2019). *Task-based language teaching: Theory and practice*. Cambridge University Press. https://doi.org/10.1017/9781108643689

Folse, K. S. (2006). The effect of type of written exercise on L2 vocabulary retention. *TESOL Quarterly*, *40*(2), 273–293. https://doi.org/10.2307/40264523

Hamada, A., Iso, T., Kojima, M., Aizawa, K., Hoshino, Y., Sato, K., Sato, R., Chujo, J., & Yamauchi, Y. (2021). Development of a vocabulary size test for Japanese EFL learners using the new JACET list of 8,000 basic words. *JACET Journal*, *65*, 23–45. https://doi.org/10.32234/jacetjournal.65.0_23

Ishikawa, M., & Suzuki, W. (2016). The effect of written languaging on learning the hypothetical conditional in English. *System*, *58*, 97–111. https://doi.org/10.1016/j.system.2016.02.008

Joe, A. (1998). What effects do text-based tasks promoting generation have on incidental vocabulary acquisition? *Applied Linguistics*, *19*(3), 357–377. https://doi.org/10.1093/applin/19.3.357

Johnson, M. D., Acevedo, A., & Mercado, L. (2016). Vocabulary knowledge and vocabulary use in second language writing. *TESOL Journal*, *7*(3), 700–715. https://doi.org/10.1002/tesj.238

Kilic, M. (2019). Vocabulary knowledge as a predictor of performance in writing and speaking: A case of Turkish EFL learners. *PASAA*, *57*(1), 133–164. https://doi.org/10.58837/CHULA.PASAA.57.1.6

Kim, Y. (2008). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. *Language Learning*, *58*(2), 285–325. https://doi.org/10.1111/j.1467-9922.2008.00442.x

Kormos, J. (2012). The role of individual differences in L2 writing. *Journal of Second Language Writing*, *21*(4), 390–403. https://doi.org/10.1016/j.jslw.2012.09.003

Kremmel, B., & Schmitt, N. (2016). Interpreting vocabulary test scores: What do various item formats tell us about learners' ability to employ words? *Language Assessment Quarterly*, *13*(4), 377–392. https://doi.org/10.1080/15434303.2016.1237516

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, *22*(1), 1–26. https://doi.org/10.1093/applin/22.1.1

Liu, J., Ding, F., & Niu, R. (2025). Effects of peer familiarity on Chinese English learners' L2 learning in collaborative writing. *International Journal of Applied Linguistics*, *35*(1), 57–72. https://doi.org/10.1111/ijal.12600

Maie, R., Eguchi, M., & Uchihara, T. (2024). Arbitrary choices, arbitrary results: Three cases of multiverse analysis in L2 research. *Research Methods in Applied Linguistics*, *3*(2), 100124. https://doi.org/10.1016/j.rmal.2024.100124

Milton, J. (2013). Measuring the contribution of vocabulary knowledge to proficiency in the four skills. In C. Bardel, C. Lindqvist, & B. Laufer (Eds.), *L2 vocabulary acquisition, knowledge and use: New perspectives on assessment and corpus analysis* (pp. 57–78). EUROSLA Monographs, 2.

Nassaji, H., & Tian, J. (2010). Collaborative and individual output tasks and their effects on learning English phrasal verbs. *Language Teaching Research*, *14*(4), 397–419. https://doi.org/10.1177/1362168810375364

Nation, I. S. P. (2022). *Learning vocabulary in another language* (3rd ed.). Cambridge University Press.

Nation, I. S. P., & Webb, S. (2011). *Researching and analyzing vocabulary*. Cengage Learning.

Pichette, F., de Serres, L., & Lafontaine, M. (2012). Sentence reading and writing for second language vocabulary acquisition. *Applied Linguistics*, *33*(1), 66–82. https://doi.org/10.1093/applin/amr037

Plakans, L. (2008). Comparing composing processes in writing-only and reading-to-write test tasks. *Assessing Writing*, *13*(2), 111–129. https://doi.org/10.1016/j.asw.2008.07.001

Plonsky, L., & Oswald, F. L. (2014). How big is "Big"? Interpreting effect sizes in L2 research. *Language Learning*, *64*(4), 878–912. https://doi.org/10.1111/lang.12079

Pruss, S., Karni, A., & Prior, A. (2025). Self-reference promotes vocabulary learning in a foreign language. *Psychonomic Bulletin & Review*, *32*, 2104–2113. https://doi.org/10.3758/s13423-025-02674-w

Qahl, T., & Lambert, C. (2025). Personal investment in language processing: The role of self-reference in initial lexical acquisition. *TESOL Quarterly*, *59*, S60–S87. https://doi.org/10.1002/tesq.3402

Rabie-Ahmed, A., & Mohamed, A. (2022). Collaborative and individual vocabulary learning in the Arabic classroom: The role of engagement and task demands. *Foreign Language Annals*, *55*(4), 1006–1024. https://doi.org/10.1111/flan.12636

Robinson, P. (2021). The cognition hypothesis, the triadic componential framework and the SSARC model: An instructional design theory of pedagogic task sequencing. In M. H. Long & M. J. Ahmadian (Eds.), *The Cambridge handbook of task-based language teaching* (pp. 205–225). Cambridge University Press. https://doi.org/10.1017/9781108868327.013

Silva, B. B., Kutyłowska, K., & Otwinowska, A. (2024). Learning academic words through writing sentences and compositions: Any signs of an increase in cognitive load? *Language Teaching Research*, *28*(3), 1143–1175. https://doi.org/10.1177/13621688211020421

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, *4*(6), 592–604. https://doi.org/10.1037/0278-7393.4.6.592

Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *The Language Learning Journal*, *36*(2), 139–152. https://doi.org/10.1080/09571730802389975

Suzuki, W. (2012). Written languaging, direct correction, and second language writing revision. *Language Learning*, *62*(4), 1110–1133. https://doi.org/10.1111/j.1467-9922.2012.00720.x

Swain, M., & Lapkin, S. (2001). Focus on form through collaborative dialogue: Exploring task effects. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 99–118). Pearson Education.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, *31*(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5

Symons, C. S., & Johnson, B. T. (1997). The self-reference effect in memory: A meta-analysis. *Psychological Bulletin*, *121*(3), 371–394. https://doi.org/10.1037/0033-2909.121.3.371

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*(5), 352–373. https://doi.org/10.1037/h0020071

Uchihara, T. (2023). How does the test modality of weekly quizzes influence learning the spoken forms of second language vocabulary? *TESOL Quarterly*, *57*(2), 595–617. https://doi.org/10.1002/tesq.3176

Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition*, *27*(1), 33–52. https://doi.org/10.1017/S0272263105050023

Webb, S., & Nation, I. S. P. (2017). *How vocabulary is learned* (1st ed.). Oxford University Press.

Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *The Canadian Modern Language Review*, *53*(1), 13–40. https://doi.org/10.3138/cmlr.53.1.13

Winter, B. (2019). *Statistics for linguists: An introduction using R* (1st ed.). Routledge. https://doi.org/10.4324/9781315165547

Wittrock, M. C. (1974). Learning as a generative process. *Educational Psychologist*, *11*(2), 87–95. https://doi.org/10.1080/00461527409529129

Wong, M. H. I. (2025). A Bayesian approach to small samples: Mixed-effects modeling in L2 interventional research. *Research Methods in Applied Linguistics*, *4*(3), 100231. https://doi.org/10.1016/j.rmal.2025.100231

Yanagisawa, A., & Webb, S. (2021). To what extent does the involvement load hypothesis predict incidental L2 vocabulary learning? A meta-analysis. *Language Learning*, *71*(2), 487–536. https://doi.org/10.1111/lang.12444

Zabihi, R., & Ghahramanzadeh, B. (2022). Proficiency pairing and engagement with language in L2 peer interaction. *System*, *105*, 102725. https://doi.org/10.1016/j.system.2022.102725

Zhang, X. (2013). The I don't know option in the vocabulary size test. *TESOL Quarterly*, *47*(4), 790–811. https://doi.org/10.1002/tesq.98

Zou, D. (2017). Vocabulary acquisition through cloze exercises, sentence-writing and composition-writing: Extending the evaluation component of the involvement load hypothesis. *Language Teaching Research*, *21*(1), 54–75. https://doi.org/10.1177/1362168816652418