

EXPLORING CHANGES IN PRODUCTIVE VOCABULARY KNOWLEDGE AMONG LOWER- PROFICIENCY ENGLISH LEARNERS THROUGH A SELF-REFLECTION WRITING TASK (SRWT)

Hosam Elmetaher
(hosamel@gmail.com)

*Akita University, Akita, Japan
1-1 Tegatagakuenmachi, Akita, 010-8502, Japan*

Abstract: This study investigates changes in productive vocabulary knowledge among pre-intermediate L1 Japanese learners of English using the Self-Reflection Writing Task (SRWT). Unlike established tasks such as the Lexical Frequency Profile (LFP; Laufer & Nation, 1995), Productive Vocabulary Levels Test (Laufer & Nation, 1999; PVLTL), and Lex30 (Meara & Fitzpatrick, 2000), the SRWT requires learners to write a brief reflective text on a familiar topic related directly to their learning experiences. The study included 24 L1 Japanese learners of English enrolled in a compulsory English Communication Skills class. Participants completed the SRWT by reflecting on their final presentation performance, writing between 130-160 words. The task was repeated at the end of each study quarter (Q1, Q2, and Q3). Responses were analyzed using the new JACET8000 corpus (JACET, 2016), categorizing vocabulary as either frequent (1k) or infrequent (non-1k). The findings suggest a possible two-phase pattern in learners' productive vocabulary development. Initially, learners appear to rely primarily on frequent (1k) vocabulary, establishing a foundational lexical base. As their vocabulary knowledge develops, they gradually begin incorporating more infrequent (non-1k) lexical items in their writing. This shift may indicate a deeper stage of vocabulary development in which learners move beyond basic lexical resources to express ideas more precisely. These findings highlight the importance of strengthening frequent vocabulary knowledge in language instruction while gradually supporting learners' expansion toward less frequent vocabulary.

Keywords: productive vocabulary, self-reflection, vocabulary development, writing assessment

DOI: <https://doi.org/10.15639/teflinjournal.v37i1/65-78>

Lexical knowledge, encompassing both receptive and productive aspects, is arguably the central focus of all language teaching. Receptive knowledge, often referred to as “passive knowledge,” comprises the vocabulary used in listening and reading, while productive knowledge, also known as “active knowledge,” involves the vocabulary used in speaking and writing.

These two aspects of vocabulary knowledge (receptive and productive) are generally considered closely related. For example, at early stages of language development, learners' ability to recognize words often correlates with their ability to produce those words; if learners can recognize the word “research,” they are more likely to be able to use it in their speech or writing. However, the relationship between receptive and productive vocabulary knowledge is

not always straightforward (Elmetaher, 2021, 2022; Lee, 2025). This complexity arises primarily from differences in (i) the size of the vocabulary, (ii) the methods of assessment, and (iii) the construct of each type of knowledge.

Firstly, it is generally assumed that receptive vocabulary knowledge is typically larger than productive vocabulary knowledge (Laufer, 1998; Nation, 2001; Read, 2000; Webb, 2005, 2007). This may be because we do not need to actively use all the words we recognize, and recognizing words does not always translate into the ability to produce them (Elmetaher, 2022, 2024).

Secondly, assessing productive vocabulary knowledge is considered more challenging than assessing receptive vocabulary knowledge (Miralpeix, 2019). Nation and Webb (2011) highlight two main reasons for this difficulty: (i) tests of productive vocabulary knowledge face more challenges in awarding credit for partial word knowledge, a process that is simpler in receptive vocabulary tests, and (ii) these tests may not fully represent a learner's entire lexicon, especially when they rely on specific prompts that may not cover all lexical fields familiar to the learner.

Thirdly, unlike receptive vocabulary knowledge, productive vocabulary knowledge is widely regarded as a multifaceted construct (Fitzpatrick, 2007). In recent discussions, Schmitt (2019) and Gu (2019) suggest that productive vocabulary encompasses not only the ability to produce words but also the capacity to use them appropriately in different contexts. Schmitt (2019) argues that productive vocabulary knowledge involves knowing a lexical item well enough to use it effectively when encoding communicative content in speech or writing. Gu (2019) emphasizes the importance of contextual appropriateness, stating that true competency in vocabulary involves understanding when and how to use specific words appropriately, depending on the interlocutors and communicative context. These multiple dimensions make productive vocabulary knowledge inherently complex and therefore challenging to assess accurately. Given this complexity, it is important to examine how productive vocabulary knowledge has been assessed in previous research and whether existing tasks adequately capture learners' vocabulary use in context. The following section reviews existing productive vocabulary assessment tasks and their limitations in order to situate the rationale for using the newly developed Self-Reflection Writing Task (SRWT) employed in the present study.

Productive Vocabulary Knowledge Tasks

Several tasks have been developed to assess productive vocabulary knowledge, each claiming to validly capture the construct. These tasks vary significantly, ranging from requiring participants to use word forms with both semantic appropriateness and grammatical accuracy within a given context, as in the Lexical Frequency Profile (LFP; Laufer & Nation, 1995), to completing words in a structured format, as seen in the Productive Vocabulary Levels Test (PVL; Laufer & Nation, 1999), or providing single-word associations, such as in the Lex30 task (Meara & Fitzpatrick, 2000).

The Lexical Frequency Profile (the LFP; Laufer & Nation, 1995)

The Lexical Frequency Profile (LFP) (Laufer & Nation, 1995) is a composition-based measure of lexical richness that has been utilized in several studies (e.g., Edmonds et al., 2022; Elmetaher, 2021; 2022; Laufer, 1994; Laufer & Nation, 1995). The LFP is designed to reflect

“the ability to use a word at one’s free will as free productive ability” (Laufer & Nation, 1999, p. 37). It requires participants to write two compositions, each approximately 300–350 words, within 60 minutes. The topics for these compositions are varied but generally cover debatable issues, such as “It is always what you do not have as a child that is important to you as an adult. Agree or disagree with this statement” (Laufer & Nation, 1995, p. 320). The compositions are then reviewed for misspellings and incorrectly used words, and analyzed using computer software to compare the vocabulary used against predefined vocabulary lists. The LFP score is determined by calculating the percentage of infrequent words (i.e., non-1k words) used by the participant in these compositions.

However, the LFP might possess at least three potential issues. Firstly, as Meara and Fitzpatrick (2000, p. 21) noted, free productive vocabulary tasks like the LFP may be “context-limited,” meaning that the specific topics assigned might not be equally familiar to all participants. Secondly, Meara and Fitzpatrick (2000) also pointed out that requiring non-native speakers to write two essays of 300 words each could be time-consuming, and potentially more challenging for lower-level learners. Thirdly, as Laufer and Nation (1999) argued, the LFP’s requirement for open-ended responses may not fully capture a learner’s productive vocabulary knowledge, as participants might tend to use more frequent words rather than less frequent ones. To address these issues, Laufer and Nation (1999) proposed a “constrained context” productive vocabulary level test designed to assess learners’ vocabulary knowledge across five different word-frequency levels, which will be reviewed next.

Productive Vocabulary Levels Test (PVL; Laufer & Nation, 1999)

The Productive Vocabulary Levels Test (PVL) (Laufer & Nation, 1999; Figure 1) is a “controlled productive” word completion task that has been widely cited in academic literature (e.g., Edmonds et al., 2022; Elmetaher, 2021; 2022; Fitzpatrick, 2007; Laufer & Nation, 1995, 1999; Meara & Alcoy, 2010; Read, 2012; Stæhr, 2009; Webb, 2009). In the PVL, the first few letters of each target word are provided (ranging from two to four letters) to direct participants towards a specific target word. The test assesses knowledge across five frequency levels: 2k, 3k, 5k, the University Word List (UWL), and 10k. Each level includes 18 test sentences, with the UWL representing 836 words and the other levels approximately 1,000 words each.

1. I am glad we had this opp_____ to talk.
2. There are a doz_____ eggs in the basket.
3. Every working person must pay income t_____.

Figure 1. The PVL (Laufer & Nation, 1999)

Note: Adapted from “A Vocabulary-size Test of Controlled Productive Ability,” by Laufer, B., & Nation, P., 1999, *Journal of Language Testing*, 16(1), p. 46. Copyright 1999 by SAGE Journals.

Laufer and Nation (1999) state that a participant’s score on the PVL is determined by the total number of correct responses. “Mastery” of a level is typically indicated by scoring around 15 or 16 correct answers out of 18, which equates to 85% to 90% accuracy (p. 41).

However, there are at least three potential issues with the PVLТ. First, the scoring responses for the three versions of the PVLТ were based on an outdated corpus (Nation, 1984). An analysis of the 90 scoring responses for each version using the BNC-COCA (Nation, 2017) corpora revealed discrepancies: the total number of scoring responses varied across the PVLТ versions, with PVLТ 1 scoring 85, PVLТ 2 scoring 79, and PVLТ 3 scoring 78 (Elmetaher, 2022).

Second, the PVLТ's scoring system may obscure the extent of the knowledge it can detect. The test only credits responses that match the target items, as noted by Walters (2012), who highlighted that a test-taker might choose a different word to complete a PVLТ sentence—potentially a less frequent word—which could indicate a broader productive vocabulary than the test reveals (p. 173). In such cases, no credit is given for knowledge of the less frequent word.

Furthermore, Laufer and Nation assumed that providing the first letters of the target words would limit the use of other semantically appropriate words from different frequency bands. However, upon closer examination, some PVLТ sentences could plausibly accept multiple correct responses (Elmetaher, 2022). For instance, as Elmetaher (2022, p.28) highlighted, in the University Word List level of the PVLТ: “*There has been a recent tr_____ among prosperous families towards a smaller number of children.*” The intended response is “*trend*,” but words like “*tragedy*,” “*track*,” “*trouble*,” “*trauma*,” and “*transition*” could also be contextually appropriate. Another example: “*There are several misprints on each page of this te_____.*” The intended response is “*text*,” but “*textbook*,” “*test*,” and “*template*” are also possible.

Third, as Meara and Fitzpatrick (2000) argued, the PVLТ's limitation of only allowing one correct response per item may effectively identify what learners do not know but is less successful at identifying the full extent of what they do know. This approach may accurately assess knowledge at lower levels but might not adequately capture higher-level vocabulary knowledge, such as at the 10,000-word level, as it is unlikely that only 18 items can represent thousands of words.

In response to the PVLТ's limitations, Meara and Fitzpatrick (2000) proposed Lex30, a word-association productive vocabulary task. They argued that Lex30 could produce “rich vocabulary output” (p. 22), offering a better representation of learners' productive vocabulary knowledge. Lex30 will be reviewed in the next section.

Lex30 (Meara & Fitzpatrick, 2000)

Lex30 (Meara & Fitzpatrick, 2000; Figure 2) is basically a word association task. It has been widely utilized and cited in several recent research papers (e.g., Elmetaher, 2021; 2022; Fitzpatrick, 2007; Meara & Fitzpatrick, 2000; Walters, 2012). Lex30 requires participants to respond with up to four words for each of 30 stimulus words, allowing for a maximum of 120 responses. The stimulus words are carefully selected based on three criteria: (i) each word is highly frequent and originates from Nation's (1984) 1,000 frequency level words; (ii) the stimulus word is chosen to avoid eliciting a primary (predictable) response, such as “dog” or “black,” based on data from first-language (English) speakers (Edinburgh Associative Thesaurus; Kiss et al., 1973); and (iii) it does not elicit common (frequent) responses when compared with first-language (English) speaker data from the same thesaurus. This method aims

to assess a participant's productive vocabulary knowledge by prompting them to produce a variety of words that are not immediately obvious or overused.

1	attack				
2	board				
3	close				

Figure 2. Lex30 (Meara & Fitzpatrick, 2000)

Note: Adapted from “Lex30: An Improved Method of Assessing Productive Vocabulary in an L2,” by Meara, P., & Fitzpatrick, T., 2000, *System*, 28(1), p. 28. Copyright 2000 by Elsevier.

Completed Lex30 task papers are transcribed and lemmatized before being compared with online corpora to determine a Lex30 score. In the scoring process, all function words, proper nouns, numbers, and words within the first 1,000 frequency band are excluded. The Lex30 score comprises a simple count of all non-1,000 frequency words, reflecting the participants' use of infrequent words.

However, Lex30 presents two main potential issues. First, Walters (2012) argued that the ability of Lex30 to elicit different types of productive vocabulary knowledge might depend on the learner's level of L2 competence. She concluded that “Lex30 may be a valid test of productive vocabulary use for higher proficiency students, whereas it is more valid as a test of productive recall at the lower levels” (p. 183). This suggests that while Lex30 is effective for assessing advanced learners' ability to use diverse vocabulary, it may primarily measure the ability to recall words rather than active usage at lower proficiency levels.

Second, similar to the PVLT, the 30 stimulus words in Lex30 are drawn from an outdated corpus (Nation, 1984). An analysis using the BNC-COCA (Nation, 2017) revealed that only 18 out of the 30 words remain within the 1,000 most frequent words, contrary to the original assumptions of the task developers.

To conclude, the previously illustrated three productive vocabulary knowledge tasks—LFP, PVLT, and Lex30—though all claiming to measure productive vocabulary knowledge, differ significantly in their procedures, length, and scoring systems.

Introducing a New Method of Assessing Productive Vocabulary Knowledge

The earlier issues identified with the three well-cited productive vocabulary knowledge tasks (LFP, PVLT, and Lex30) highlight the need for a new assessment method. To address this gap, the present study uses the Self-Reflection Writing Task (SRWT) as a classroom-based approach for examining changes in productive vocabulary knowledge among lower-level learners. Unlike LFP (Laufer & Nation, 1995), PVLT (Laufer & Nation, 1999), and Lex30 (Meara & Fitzpatrick, 2000), the SRWT requires participants to produce a full written text based on a shared and familiar learning experience.

In the SRWT, participants initially complete a language task, such as a presentation or a language activity. Following this task, participants reflect on their performance, focusing on two main points: what they did well and what they could improve. This reflective writing takes place

in class under teacher supervision, ensuring that responses are personalized and preventing the use of external resources like dictionaries or online sites. Each student is required to type between 130-160 words, a length that is manageable for lower proficiency learners and fits within typical class time constraints. Typing responses also facilitates easier processing and analysis by the teacher.

Participants' responses are then collected and analyzed using a relevant word frequency corpus, such as the New JACET8000 (JACET, 2016). This corpus is the most updated vocabulary list provided by the Japan Association for University English Education (JACET), specifically designed to align with Japanese English teaching materials and examination content, ensuring relevance and accuracy in assessment.

Unlike the previously discussed productive vocabulary knowledge tasks, the SRWT scores are based on both the percentage of infrequent words (non-1k) and frequent words (1k). This dual focus on both infrequent and frequent words marks a significant departure from traditional methods that emphasize only infrequent words, as seen in the LFP, PVL, and Lex30. The SRWT, in its design, procedures, and scoring system, addresses the limitations identified in the three previously mentioned productive vocabulary knowledge tasks: the Lexical Frequency Profile (LFP), the Productive Vocabulary Levels Test (PVL), and Lex30, as presented in Table 1.

Table 1. How SRWT Addresses the Potential Issues of LFP, PVL, and Lex30

Task	Description	Potential Issues	How SRWT Addresses Issues
Lexical Frequency Profile (LFP)	Measures lexical richness by requiring two compositions of 300-350 words on general topics.	Context-limited due to specific topics (Meara & Fitzpatrick, 2000). Time-consuming (Meara & Fitzpatrick, 2000), especially for lower-level learners. Might not fully estimate learners' productive vocabulary knowledge (Laufer & Nation, 1999)	Allows participants to write about familiar topics, reducing context limitations. Less time-consuming by using shorter, focused prompts. Considers both frequent and infrequent words, providing a comprehensive vocabulary assessment.
Productive Vocabulary Levels Test (PVL)	Controlled productive word completion task with prompts for five frequency levels.	Scoring responses are based on outdated corpora of Nation (1984). Only credits specific target items, potentially missing broader vocabulary knowledge (Walters, 2012).	Uses an updated corpus for analysis (New JACET8000; JACET, 2016) Credits broader vocabulary use, not limited to specific target items.

Task	Description	Potential Issues	How SRWT Addresses Issues
		Some sentences may accept multiple correct responses, complicating accurate assessment (Elmetaher, 2022)	Doesn't have restrictions on specific responses and doesn't target specific words.
		May not detect higher-level vocabulary knowledge accurately (Meara & Fitzpatrick, 2000).	Provides a more comprehensive view of vocabulary knowledge across different levels, including higher-level vocabulary.
Lex30	Word association task requiring responses to 30 stimulus words, focusing on infrequent words.	May elicit different types of vocabulary knowledge (use for higher/ recall for lower) based on learner's proficiency level (Walters, 2012). Scoring responses are based on outdated corpora of Nation (1984).	Might validate both recall and use of productive vocabulary. Uses updated corpus for analysis (New JACET8000; JACET, 2016)

To summarize, the SRWT addresses the key issues found in the LFP, PVL, and Lex30 by offering a more updated, relevant, valid, comprehensive, and practical assessment method. The next step is to implement this method in language classrooms to further validate its effectiveness.

METHOD

This study aims to explore changes in productive vocabulary knowledge among pre-intermediate L1 Japanese learners of English using SRWT. The study included 24 participants, all native Japanese speakers and first-year undergraduate students at a university in central Japan. Based on their placement and course level, participants were estimated to be at the A2 to B1 level of the Common European Framework of Reference for Languages (CEFR). The study spanned three quarters, with each quarter consisting of 14 sessions. Classes were held twice per week, with each session lasting 100 minutes.

Procedures

Participants were asked to respond to the SRWT by reflecting in writing on their final oral presentation performance, highlighting both strengths and areas for improvement. No specific format was provided, allowing participants to express themselves freely. The SRWT was administered three times, at the end of each study quarter (Q1, Q2, and Q3), under the

supervision of the classroom teacher. Responses were collected on an online platform, with a word limit set between 130-160 words. After collection, responses were checked for typos and analyzed using the updated JACET8000 word corpus (JACET, 2016) to examine word frequency levels. The analysis focused on the proportion of frequent (1k) and infrequent (non-1k) words in each response. To determine whether changes in vocabulary use were statistically significant across the three testing times, paired-sample t-tests were conducted for each consecutive pair of testing times (Time 1 vs. Time 2, Time 2 vs. Time 3, and Time 1 vs. Time 3). Effect sizes (Cohen's *d*) and 95% confidence intervals were also calculated to provide a clearer indication of the magnitude and precision of the observed changes.

FINDINGS AND DISCUSSION

This section presents the results of the analysis of learners' productive vocabulary knowledge across the three testing times. Table 2 presents the descriptive statistics for the proportions of frequent (1k) and infrequent (non-1k) words produced at each testing time, along with the observed differences and their statistical significance. Table 3 provides the detailed results of the paired-sample t-tests, including effect sizes and confidence intervals, to further examine the magnitude and significance of these changes.

Table 2. Changes in Productive Vocabulary Knowledge Percentages Across Three Testing Times

N= 24	Frequent words (1k) percentage	Infrequent words (non-1k) percentage
Time 1	85.80 (3.99)	14.20 (3.99)
Time 2	86.33 (3.96)	13.67 (3.96)
Time 3	83.95 (3.36)	16.05 (3.36)
Changes between time points		
Time 2- Time 1	0.53	-0.53
Time 3- Time 2	-2.38 **	2.38**
Time 3- Time 1	-1.85	1.85

Note: ** $p < .01$. Numbers in parentheses are standard deviations.

As illustrated in Table 2, the average percentage of frequent words (1k) produced in Time 2 (86.33%) was slightly higher than in Time 1 (85.80%), but this percentage decreased in Time 3 (83.95%). In contrast, the percentage of infrequent words (non-1k) produced in Time 2 (13.67%) was slightly lower than in Time 1 (14.20%) but increased in Time 3 (16.05%). These descriptive results suggest a shift in learners' vocabulary use across the three testing times, with a gradual increase in the use of infrequent vocabulary in the later stage. To further examine whether these differences were statistically significant, paired-sample t-tests were conducted (Table 3).

As indicated in Table 3, no statistically significant differences were observed between Time 1 and Time 2 for either frequent vocabulary ($t(23) = -0.42$, $p = .677$, $d = -0.09$, 95% CI $[-0.49$,

0.32]) or infrequent vocabulary ($t(23) = 0.42$, $p = .677$, $d = 0.09$, 95% CI [-0.32, 0.49]). However, a significant change was observed between Time 2 and Time 3. The proportion of frequent vocabulary decreased significantly ($t(23) = 2.92$, $p = .008$, $d = 0.60$, 95% CI [0.16, 1.03]), while the proportion of infrequent vocabulary increased significantly ($t(23) = -2.92$, $p = .008$, $d = -0.60$, 95% CI [-1.03, -0.16]). The effect sizes for these differences were moderate, suggesting that the observed changes represent meaningful shifts in learners' vocabulary use over time. No statistically significant differences were observed between Time 1 and Time 3 for either frequent vocabulary ($t(23) = 1.40$, $p = .176$, $d = 0.29$, 95% CI [-0.13, 0.69]) or infrequent vocabulary ($t(23) = -1.40$, $p = .176$, $d = -0.29$, 95% CI [-0.69, 0.13]).

Table 3. Paired-Sample Statistical Comparisons Across the Three Testing Times

Comparison	Frequent Words Change	<i>p</i>	Cohen's <i>d</i>	Infrequent Words Change	<i>p</i>	Cohen's <i>d</i>
T2-T1	0.53	.677	-0.09	-0.53	.677	0.09
T3-T2	-2.38	.008	0.60	2.38	.008	-0.60
T3-T1	-1.85	.176	0.29	1.85	.176	-0.29

The statistical pattern observed across the three testing times suggests that productive vocabulary development among lower-proficiency learners may not follow a continuous or immediately observable trajectory. Although no statistically significant differences were found between Time 1 and Time 2 or between Time 1 and Time 3, the significant and moderately sized shift between Time 2 and Time 3 indicates that measurable changes in productive vocabulary use may emerge only after learners have had sufficient opportunities to repeatedly encounter, consolidate, and actively use previously learned vocabulary. This interpretation is consistent with Nation's (2013) view that learning a word is a cumulative process requiring multiple encounters before different aspects of word knowledge become established. It also relates to Schmitt's (2019) observation that understanding how vocabulary knowledge develops from receptive knowledge to productive mastery remains an important area of vocabulary research.

In addition to the quantitative results, a qualitative examination of learners' written responses was conducted to observe how vocabulary was used across the three testing times. A small number of representative responses were reviewed to identify general patterns in lexical choice and expression. In the earlier reflections, learners relied mainly on frequent and general-purpose vocabulary to describe their performance, often using simple and repetitive expressions. For example, a learner wrote in the first reflection: "My presentation was good but I was very nervous and my English was not good." By the third testing time, however, learners showed a greater tendency to use more specific and evaluative vocabulary to describe their strengths and areas for improvement. For instance, one learner wrote: "I think my presentation was clearer this time because I practiced more and used better explanation for my topic." These examples illustrate how learners gradually began to use more specific vocabulary and clearer expressions to describe their performance. This suggests that learners were not only producing a higher proportion of infrequent words but were also becoming better able to select vocabulary that more precisely expressed their intended meaning in reflective writing.

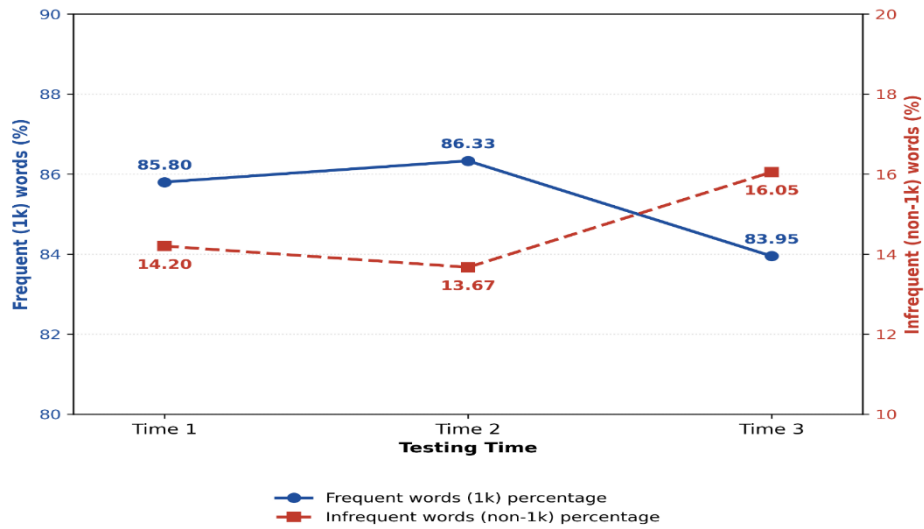
Although the examples above are illustrative rather than a formal qualitative analysis, the shift from simple, general vocabulary in the earlier reflections to more specific and evaluative vocabulary by the third testing time can be situated within Jiang's (2000, 2002, 2004) view of lexical development, which treats increasingly context-appropriate word use as a more advanced stage of productive knowledge than the mere accumulation of word forms. This gradual emergence is also consistent with recent longitudinal evidence from Min and Li (2025), who found that productive vocabulary knowledge became a significantly stronger predictor of L2 writing performance only as time progressed, even though receptive vocabulary was the dominant predictor early on.

Taken together, and as illustrated in Figure 4, these quantitative and qualitative findings may tentatively suggest a two-phase pattern in productive vocabulary development across the three testing times. In the initial stage, learners may focus on consolidating frequent vocabulary, becoming familiar with a wider range of words within the most frequent 1,000-word band before extending their productive use to less frequent items. This progression is consistent with Laufer and Nation's (1999) argument that words are generally learned in rough order of frequency, with high-frequency words offering the greatest communicative return and therefore serving as the natural starting point for productive use. This phase may be characterized by a focus on acquiring and using frequent vocabulary that enables learners to communicate basic ideas effectively and understand a wider range of contexts.

As learners become more comfortable with these frequent words, they might then extend their productive vocabulary to include a greater proportion of infrequent items. This general shift away from reliance on the most frequent vocabulary band is echoed in recent longitudinal evidence from Chinese EFL writing. Lou et al. (2026) tracked learners' compositions across four semesters using the Lexical Frequency Profile method. They found that the proportion of the most frequent 1,000-word band decreased overall, while use of the second and third 1,000-word bands increased over time — indicating a broader movement from reliance on frequent vocabulary toward greater use of less frequent words. A similar movement is evident in the present data: the moderate effect size observed between Time 2 and Time 3 suggests that this shift was not only statistically significant but also practically meaningful in terms of learners' ability to move toward more specific lexical choices in classroom-based reflective writing.

This possible two-phase pattern of productive vocabulary development, encompassing changes in both frequent and infrequent words, may offer an alternative interpretation to earlier research (e.g., Daller et al., 2007; Horst & Collins, 2006; Elmetaher, 2021, 2022), which often characterized productive vocabulary knowledge as inconsistent. Those studies primarily examined changes in infrequent vocabulary and interpreted fluctuations in the use of less frequent lexical items as evidence of inconsistency in productive vocabulary development. One possible explanation for these findings is methodological rather than developmental. Because lower-proficiency learners naturally rely heavily on frequent vocabulary, focusing exclusively on infrequent words may overlook meaningful improvements occurring within learners' core productive vocabulary. This reliance on high-frequency vocabulary has also been documented in recent corpus-based research beyond English. Abu-Rabiah (2025) found that Arabic-speaking learners of Hebrew as a second language relied predominantly on the most frequent 1,000-word

band in their productive written output upon entering higher education, with usage decreasing systematically across successively less frequent bands.



Note. Frequent (1k) and infrequent (non-1k) word percentages are plotted on separate axes (left: 80–90%; right: 10–20%) due to their different ranges. Values correspond to Table 2.

Figure 4. A Tentative Two-Phase Pattern of Productive Vocabulary Development Among Lower-Proficiency Learners Across the Three Testing Times

CONCLUSION

The findings of this study have several important implications for both language testing and classroom practice. In language testing, some high-stakes language assessments (e.g., IELTS) include scoring criteria that encourage the use of less common lexical items in writing and speaking tasks. While such criteria may reflect advanced lexical ability, the present findings suggest that productive vocabulary development among lower-proficiency learners may also involve meaningful changes in the use of frequent (common) vocabulary. Consequently, learners may make substantial gains in their productive vocabulary knowledge through increased control and more effective use of frequent lexical items, yet such development may remain unnoticed if assessment primarily rewards the use of less common vocabulary. Recognizing incremental changes in both frequent and infrequent vocabulary may therefore provide a more realistic and comprehensive picture of learners' productive vocabulary changes.

Second, the findings also have important implications for classroom practice. When designing teaching materials—whether based on published textbooks or developed by teachers—it is important to consider the types of vocabulary that are appropriate for learners' proficiency levels. For lower-proficiency learners, such as the participants in this study, strengthening control over high-frequency vocabulary may play a particularly important role in supporting productive language development. As learners develop stronger control of this core

vocabulary, they may be better prepared to expand their lexical resources by gradually incorporating less frequent vocabulary into their writing and speaking.

While this research might be the first of its kind to use the SRWT to track changes in both frequent and infrequent productive vocabulary knowledge among lower-level learners over three quarters, there are some limitations to consider.

As the study was conducted with a relatively small sample size of 24 participants, all of whom were L1 Japanese speakers with lower-level English proficiency. To further validate the findings and broaden their applicability, future studies should involve larger sample sizes, include participants with different first languages, and encompass a range of English proficiency levels.

Further, the study set a word limit for each Self-Reflection Writing Task to be between 130-160 words. While 130 words might be a reasonable expected response from lower-proficiency learners, having a maximum limit of 160 words might have restricted participants from producing more extensive responses. Future studies might consider removing the maximum word limit to encourage participants to produce as much as they can, while still setting a minimum acceptable word count.

Finally, while the results suggest changes in productive vocabulary knowledge over time, alternative explanations should also be considered. Because the SRWT involved repeated reflective writing tasks across the three quarters, some of the observed changes may partly reflect increased familiarity with the reflective genre or repeated-task effects rather than vocabulary development alone. In addition, improvements in general writing ability over the academic year may have influenced learners' ability to express their ideas more clearly and select more appropriate lexical items. Although the SRWT was designed to capture productive vocabulary knowledge in a classroom context, future studies could further clarify these possibilities by incorporating comparison tasks, control groups, or additional measures of writing development.

ACKNOWLEDGEMENT

I would like to express my deepest gratitude to everyone who supported and contributed to this research. Parts of the discussion of earlier productive vocabulary assessment tasks in this article are based on and adapted from my unpublished doctoral dissertation completed at Hiroshima University.

DECLARATION OF AI USE

During the preparation of this manuscript, the author used ChatGPT for proofreading and improving the clarity and readability of the language. After using this tool, the author reviewed and edited the content as necessary and takes full responsibility for the content of the manuscript.

REFERENCES

- Abu-Rabiah, E. (2025). The productive vocabulary size of second language learners upon entry into higher education. *Theory and Practice of Second Language Acquisition*, 11(1), 1–19. <https://doi.org/10.31261/tapsla.16594>

- Daller, H., Milton, J., & Treffers-Daller, J. (2007). Modelling and assessing vocabulary knowledge. Cambridge University Press. <https://doi.org/10.1017/CBO9780511667268>
- Edmonds, A., Clenton, J., & Elmetaher, H. (2022). Exploring the construct validity of tests used to assess L2 productive vocabulary knowledge. *System*, 108, Article 102855. <https://doi.org/10.1016/j.system.2022.102855>
- Elmetaher, H. (2021). Investigating productive vocabulary knowledge development: A task-based approach. *Studies in European and American Cultures*, 28, 1-26. <http://doi.org/10.15027/52374>
- Elmetaher, H. (2022). Exploring the extent to which productive vocabulary knowledge tasks detect changes [Unpublished doctoral dissertation, Hiroshima University, Japan]. https://hiroshima.repo.nii.ac.jp/record/2033341/files/StuIntegrArtSci_3_65.pdf
- Elmetaher, H. (2024). Investigating the MRVAT sensitivity in detecting changes in the first 2000 most frequent words in the new JACET8000. *Korea TESOL Journal*, 20(1), 85-106.
- Fitzpatrick, T. (2007). Productive vocabulary tests and the search for concurrent validity. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Modelling and assessing vocabulary knowledge* (pp. 116-132). Cambridge University Press.
- Gu, P. (2019). Strategies for learning vocabulary. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 271-287). Routledge.
- Horst, M., & Collins, L. (2006). From faible to strong: How does their vocabulary grow? *The Canadian Modern Language Review/La Revue canadienne des langues vivantes*, 63(1), 83-106. <https://doi.org/10.3138/cmlr.63.1.83>
- JACET (2016). *Basic vocabulary for Japanese university students New JACET8000*. <https://language.sakura.ne.jp/s/voc.html>
- Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics*, 21(1), 47-77. <https://doi.org/10.1093/applin/21.1.47>
- Jiang, N. (2002). Form-meaning mapping in vocabulary acquisition in a second language. *Studies in Second Language Acquisition*, 24(4), 617-637. <https://doi.org/10.1017/S0272263102004047>
- Jiang, N. (2004). Semantic transfer and its implications for vocabulary teaching in a second language. *The Modern Language Journal*, 4, 416-432. <https://doi.org/10.1111/j.0026-7902.2004.00238.x>
- Kiss, G., Armstrong, C., Milroy, R., & Piper, J. (1973). An associative thesaurus of English and its computer analysis. In A. Aitken, R. Bailey, & N. Hamilton-Smith (Eds.), *The computer and literary studies* (pp. 153-165). Edinburgh University Press.
- Laufer, B. (1994). The lexical profile of second language writing: Does it change over time? *RELIC Journal*, 25(2), 21-33. <https://doi.org/10.1177/003368829402500202>
- Laufer, B. (1998). The development of passive and active vocabulary in a second language: Same or different? *Applied Linguistics*, 19(2), 255-271. <https://doi.org/10.1093/applin/19.2.255>
- Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16(3), 307-322. <https://doi.org/10.1093/applin/16.3.307>
- Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, 16(1), 33-51. <https://doi.org/10.1177/026553229901600103>

- Lee, S. (2025). The relationship between receptive and productive knowledge of L2 English collocations. *International Journal of Applied Linguistics*, 35(1), 109-133. <https://doi.org/10.1111/ijal.12605>
- Lou, X., Wu, C., & Ma, G. (2026). Longitudinal profiling of Chinese EFL learners' lexical frequency in writing. *Arab World English Journal*, 17(2), 290-309. <https://dx.doi.org/10.24093/awej/vol17no2.19>
- Meara, P., & Alcoy, J. C. (2010). Words as species: An alternative approach to estimating productive vocabulary size. *Reading in a Foreign Language*, 22(1), 222-236. <https://nflrc.hawaii.edu/rfl/item/217>
- Meara, P., & Fitzpatrick, T. (2000). Lex30: An improved method of assessing productive vocabulary in an L2. *System*, 28(1), 19-30. [https://doi.org/10.1016/S0346-251X\(99\)00058-5](https://doi.org/10.1016/S0346-251X(99)00058-5)
- Min, C., & Li, D. (2025). From words to writing: Vocabulary growth and its impact on L2 composition over time. *SAGE Open*, 15(4). <https://doi.org/10.1177/21582440251412208>
- Miralpeix, I. (2019). *L1 and L2 vocabulary size and growth*. Routledge.
- Nation, P. (1984). *Vocabulary lists: Words, affixes, and stems*. Victoria University of Wellington.
- Nation, P. (2001). *Learning vocabulary in another language*. Cambridge University.
- Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge University Press.
- Nation, P. (2017). *The BNC/COCA Level 6 word family lists (Version 1.0.0) [Data file]*. <http://www.victoria.ac.nz/lals/staff/paul-nation.aspx>
- Nation, P., & Webb, S. (2011). *Researching and analyzing vocabulary*. Heinle.
- Read, J. (2000). *Assessing vocabulary knowledge and use*. Cambridge University Press.
- Read, J. (2012). Piloting vocabulary tests. In G. Fulcher, & F. Davidson (Eds.), *The Routledge handbook of language testing* (pp. 307-320). Routledge.
- Schmitt, N. (2019). Understanding vocabulary acquisition, instruction, and assessment: A research agenda. *Language Teaching*, 52(2), 261-274. <https://doi.org/10.1017/S0261444819000053>
- Stæhr, L. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition*, 31(4), 577-607. <https://doi.org/10.1017/s0272263109990039>
- Walters, J. (2012). Aspects of validity of a test of productive vocabulary: Lex30. *Language Assessment Quarterly*, 9(2), 172-185. <https://doi.org/10.1080/15434303.2011.625579>
- 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. (2007). The effects of synonymy on second-language vocabulary learning. *Reading in a Foreign Language*, 19(2), 120-136. <http://www2.hawaii.edu/~readfl/rfl/October2007/webb/webb.pdf>
- Webb, S. (2009). The effects of receptive and productive learning of word pairs on vocabulary knowledge. *RELC Journal*, 40(3), 360-376. <https://doi.org/10.1177/0033688209343854>