Guoyuhui Huang

The University of Hong Kong, Hong Kong SAR, China

hgyuhui@connect.hku.hk

Repetition supports the effects of Involvement Load Hypothesis on improving students' productive vocabulary performance

Bio data

Guoyuhui Huang is a Ph.D candidate at the Faculty of Education, the University of Hong Kong. Her research interests include understanding how technology can be utilized to enhance students' language learning.

Abstract

Productive vocabulary refers to retrieving and applying the words in speaking and writing. It forms the basis for EFL (English as a Foreign Language) learners to express themselves accurately and fluently. Recent years have witnessed a growth of publications examining the effects of the Involvement Load Hypothesis (ILH) on productive vocabulary performance, although with somewhat mixed results. The present study explored whether 'repetition' could complement the ILH in improving EFL learning of productive vocabulary. Correspondingly, two WeChat Applets (Applet 1.0 and Applet 2.0) were designed to help Chinese EFL learners apply productive vocabularies in the IELTS (International English Language Testing System) examination. Applet 1.0 was chiefly developed based on the ILH. Applet 2.0 was developed based on ILH with the additional support of repetitive reading activities. Specifically, learners studied with Applet 1.0 merely encountered each target item once. By contrast, learners studied with Applet 2.0 would meet each required vocabulary eight times. Fifty Chinese college students, divided into a control group (CG, n = 26) and an experimental group (EG, n = 24), participated in the present study. Both groups were asked to write a composition every week while learning with different applets (CG learned with 1.0 and EG studied with 2.0). Three paragraph writing tests, namely pre-test, post-test and delayed-test were administered to assess their productive vocabulary proficiency. We found the EG significantly outperformed the CG in terms of the post-test and delayed-test. Therefore, it was concluded that repetition and ILH were indeed compatible, which could result in better productive vocabulary acquisition.

Conference paper

Introduction

One explicit and well-recognized taxonomy for vocabulary knowledge is receptive and productive knowledge. Receptive vocabulary is often defined as the ability to comprehend the words in listening and reading; while, productive vocabulary requires producing the relevant words in speaking and writing (Schmitt, 2010). Put another way, receptive vocabulary knowledge refers to the ability of a learner to recognize a word, while

productive vocabulary refers to the ability of a learner to apply the word correctly. Since the application of a word is far more challenging than recognition, previous studies (e.g., Malmström et al., 2018) have reported that learning productive vocabulary is more demanding for L2 learners. However, there has been little research on effective instructional strategies to improve productive vocabulary knowledge among L2 learners, Therefore, a critical question remains: how could we help students effectively acquire productive vocabulary?

Literature Review

Previous Studies on Involvement Load Hypothesis

Over the past two decades, a wealth of research (e.g., Bao, 2015; Keating, 2008; Zou, 2017) has empirically showed that productive vocabulary acquisition might hinge on the degree of involvement in processing new words, which Laufer and Hulstijn (2001) called the Involvement Load Hypothesis (ILH). The ILH consists of three essential constructs: need (N), search (S) and evaluation (E). Furthermore, Laufer and Hulstijn (2001) defined some degrees of prominence for each component. Need is a motivational dimension of the hypothesis, which is defined as the learners' intention to understand the words. It is hypothesized to be absent (N-, 0 points), moderate (N+, 1 point) or strong (N++, 2 points). For example, need is absent when a target word is not required to accomplish a task. It is induced to a moderate degree when learning the target words is externally imposed by teachers or other authorities. By contrast, need is strong when the intention to learn the words is self-imposed. Search is a cognitive element. It refers to the attempt to ascertain the L2 form of a word or its equivalent L1 meaning. According to Laufer and Hulstihn (2001), it is absent (S-, 0 points) when no such effort is required (i.e., a reading comprehension task with some marginal glosses), and it is present (S+, 1 point) when students must seek the L2 form or L1 meaning to complete an assignment. With respect to evaluation (a cognitive element), it entails the comparison between a target word and other related words, or knowledge of a word, with the context of utilization to decide to check if it fits. Evaluation appears to be absent (E-, 0 points), moderate (E+, 1 point) or strong (E++, 2 points). It is absent when learners do not need to determine which word or tense of the word to employ. Moderate evaluation requires one to distinguish between the multiple meanings or forms to find the most appropriate one when the context is given (i.e., a gap-filing activity). By contrast, evaluation is strong if learners need to combine the new word with others to create an original context (as opposed to a given). For example, students need to use the target words in writing sentences or compositions.

The sum of the components with their degree is called the task involvement load (IL). Laufer and Hulstijn (2001) claimed a task with a higher IL was more effective and would yield better vocabulary acquisition and greater vocabulary retention. Recent years have observed a flurry of publications examining its predictive power on productive vocabulary acquisition, although with somewhat mixed success (e.g., Alavinia & Rahimi, 2019; Jafari et al., 2018; Tahmasbi & Farvardin, 2017; Zou, 2017). Admittedly, an emerging body of scholars (e.g., Pourakbari & Biria, 2015; Tahmasbi & Farvardin, 2017) still observed a significant decrease of productive vocabulary knowledge in the delayed-test, even when participants conducted the tasks with high IL (i.e., the sentence-writing task or the paragraph-writing task). Therefore, we posited there is room to improve the ILH. In other words, other factors might complement the ILH in improving students' delayed productive vocabulary performance.

Background Literature on the 'Repetition'

A recent meta-analysis on ILH (Yanagisawa & Webb, 2021) which scrutinized 42 empirical studies has aligned with our prediction that other factors might support the effects of the ILH. To begin with, they reported that the predictive ability of ILH was limited: it explained merely 15.0% and 5.1% of the variance in effect sizes on the

post-test and delayed-test of vocabulary knowledge, respectively. Moreover, they demonstrated that 'repetition' might positively contribute to immediate post-test learning gains, regardless of the amount of IL of the vocabulary learning tasks. Repetition refers to the number of times a learner encounters an unknown word in contexts. According to Folse (2006), repetition provided retrieval opportunities for students. Put differently, when a student encountered a word repeatedly while reading, he was likely to focus greater attention on that unfamiliar word in the first several encounters to infer its meaning and retrieve information learned about that word from the previous encounters (Pellicer-Sánchez et al., 2021).

However, the repetition effect was not found on delayed post-tests possibly due to the limited number of repetitions in the studies (Yanagisawa & Webb, 2021). The target words in a majority of studies were not repeated; when repetition was included, the mean frequency was quite low (M = 3.69, Mdn = 4) (Yanagisawa & Webb, 2021). If future studies were to improve students' delayed vocabulary performances, a higher number of repetitions might be required.

Contributions of the present study

Given the gaps in the previous literature, an attempt to expand and advance the theoretical and practical utility of ILH in productive vocabulary contexts is needed. An empirical study that overcomes the deficiencies in previous studies (e.g., inadequate repetitions) is also needed.

In the present study, we posited ILH with higher number of repetitions (more than four times, as indicated by the preceding literature review) might result in better productive vocabulary acquisition. To testify the hypothesis above, we designed two WeChat Applets: Applet 1.0 and Applet 2.0. Applet 1.0 was based on the ILH solely. Applet 2.0 was also informed by ILH supplemented with repetitive reading activities. Learners studied with Applet 1.0 merely encountered each target item once, while learners studied with Applet 2.0 would meet each vocabulary eight times. Furthermore, two groups of students: a control group (CG, n = 26) and an experimental group (EG, n = 24) studied with different Applets (CG studied with Applet 1.0 while EG learned with Applet 2.0) to reveal the discrepancies.

Method

Participants

A randomized control trial was embedded within a quantitative research design. A total of 50 EFL students (CG, n = 26; EG, n = 24), whose age range was between 18 and 20, participated in the present study. The Human Research Ethics Committee at the authors' university has endorsed an ethical approval.

Instruments and Materials

WeChat Applet (Applet) is a web application that can be accessed without downloading or installing. It could bring about a rapid transfer of digital data between instructors and learners, and thus has been proved useful for improving students' learning performances (e.g., Wu et al., 2018). Therefore, we used Applets to distribute learning materials in the present study.

Target words

Since the participants in the present study were going to take the IELTS (International English Language Testing System), we selected 40 words at IELTS level. Moreover, we selected ten verbs, ten adjectives, ten nouns and ten conjunctions to control the possible

confusion of part of speech with tasks. Students needed to learn ten target words per week. Accordingly, they finished learning by the end of the fourth week.

Paragraph-writing task

Previous studies have posited that paragraph-writing task was superior to other vocabulary tasks since it involved deeper processing (e.g., Kim, 2011; Zou, 2017). According to Zou's (2017) analysis, when students were writing paragraphs, semantic contexts of target words were generated, which was conducive to word learning. Therefore, every week, we asked students to finish a semantically acceptable and grammatically correct paragraph with ten target words they learned that week. In the end, they needed to submit four compositions in total.

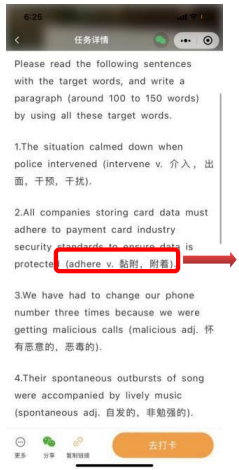
Learning materials in the Applet 1.0

Applet 1.0 distributed the reading tasks for the CG. Since students in the CG encountered each target word once, forty sentences were prepared for them. These sentences were adapted or selected from Oxford Advanced Learner's English-Chinese Dictionary (Hornby, 2009), Merriam-Webster Dictionary (online version, see at https://www.merriam-webster.com) and Cambridge Dictionary (online version, see at https://dictionary.cambridge.org), in which target words were embedded. The teacher-researchers and other five EFL teachers checked the appropriateness of vocabulary and syntax of the sentences for the participants. Figure 1 illustrated the learning interface in Applet 1.0. Each sentence appeared with one target word embedded. After each sentence, the corresponding glosses (parts of speeches and Chinese translations) for that target word were provided in the brackets. We take the word *adhere* as an example. In this example (see Fig.1), the target word is *adhere* and its gloss provides its part of speech (a verb), and Chinese translation (粘附，附着).

Learning materials in the Applet 2.0

Applet 2.0 delivered the reading materials for EG. As noted earlier, Applet 2.0 was supplemented with repetitive reading activities. According to Uchihara et al. (2019), although repetition might contribute to incidental vocabulary acquisition, there were diminishing learning gains as the number of repetition increased beyond a certain point (around 20 encounters). This has aligned with other studies claiming that more was better might not always be true (Elgort et al., 2018; Pelliver-Sánchez, 2016). Elgort et al. (2018)'s eye-tracking research has suggested repetition remained a plateau effect on the processing of target words after a certain number of encounters (eight to ten times). Considering all these useful findings, we let students in the EG meet eight times of each target word. Correspondingly, we designed the learning process as follows: each word appeared twice from every Monday to Thursday (2 times * 4 days = 8 times); every Friday was left for students to finish a paragraph writing task.

Therefore, altogether 320 sentences (40 target words * 8 repetitions) were selected from the dictionaries we mentioned above (Oxford Advanced Learner's English-Chinese Dictionary, Merriam-Webster Dictionary and Cambridge Dictionary). They were divided into four sets, with each presented per week. Each set included 80 sentences as well as ten target words. Every day from Monday to Thursday, students met 20 sentences per day, with each target word appearing twice. After each sentence, their corresponding glosses (parts of speeches and Chinese translations) for the target words were provided (see Figure 2).



adhere (target word) v. (part of speech) 黏附，附着 (Chinese translation: affix, stick)

Figure 1. Learning interface of the Applet 0.0

Figure 2. Learning interface of the Applet 2.0

Measures

We conducted three paragraph writing tests (pre-, post- and delayed-tests) with two demands on students to gauge their productive vocabulary proficiency. First, they were required to write down the Chinese translations of ten pre-selected words. Second, they should write a paragraph using these words as many as possible. One point would be given to a sentence precisely applied to a target word with an accurate Chinese translation. One point would also be assigned if other parts of sentences were wrong (Tahmasbi & Farvardin, 2017). For example, if the target word is *adhere*, students could be given one point if they wrote 'We should adhere to the regulations, no matter what is happened'. This is because the clause which contains the target word is both semantically acceptable and grammatically correct. The maximum score for each pre-, post-, and delayed-test was ten.

Procedure

To begin with, a pre-test was administered before the treatment to see if there was a significant difference of prior knowledge between groups. Subsequently, a four-week treatment was implemented. During the treatment, CG and EG students used Applet 1.0 and Applet 2.0 to study 40 target words, respectively. Moreover, both groups were required to submit a paragraph-writing assignment every week. After four-week learning, students were given a post-test. Two weeks later, a delayed test was administered.

Results

As demonstrated in Table 1, the prior knowledge before the treatment was similar (EG: M = 0.17, SD = 0.82; CG = 0.19, SD = 0.49), p = .89. After the treatment (see Table 2), both groups enhanced significantly in terms of productive vocabulary in the post-test (EG = 7.79, SD = 1.82; CG = 4.69, SD = 2.57) and the delayed-test (EG = 7.79, SD = 1.61; CG = 4.19, SD = 1.70), F (2, 96) = 273.63, p = .00, η^2 = .85). Furthermore, there was a significant main effects of group, F (1,48) = 47.10, p = .00, η^2 = .50. Table 2 also illustrated a significant interaction between time and group, F (2, 96) = 22.44, p = .00, η^2 = .32, indicating the EG significantly outperformed the CG in terms of improvement in productive vocabulary proficiency. The results also suggested an almost large effect size for time and an almost medium effect size for group. Furthermore, the results of pairwise

comparison of time and group (see Table 3) suggested a significant improvement of both groups from the pre-test to the post-test, and a sustained knowledge retention of two groups from the post-test to the delayed-test.

Table 1. Means and SDs for productive vocabulary scores for the EG and CG in the pre-test, post-test and delayed-test

	Pre-test		Post-test		Delayed-test	
Group	M	SD	M	SD	M	SD
EG (n = 24)	0.17	0.82	7.79*	1.82	7.79*	1.61
CG (n = 26)	0.19	0.49	4.69	2.57	4.19	1.70

Note. M = mean; SD = standard deviation
Maximum score = 10
*p < .001

Table 2. Results of mixed ANOVA on productive vocabulary tests

Source	Type III SS	df	MS	F	Significanc	Partial η^2
		Between-subject				
Group	185.25	1	185.25	47.10	.00	.50
Error	188.78	48	3.93			
		Within-subject				
Time	1174.81	2	587.41	273.63	.00	.85
Time × Group	96.33	2	48.17	22.44	.00	.32
Error (time)	206.08	96	2.15			

Note. SS = sum of square; MS = mean square

Table 3. Results of pairwise comparison of time and group

Group	(I)time	(J)time	MD	SE	Significance
EG	Pre-test	Post-test	-7.63	.45	.00
		Delayed-test	-7.63	.34	.00
	Post-test	Pre-test	7.63	.45	.00
		Delayed-test	- .00	.46	1.00
	Delayed-test	Pre-test	7.63	.34	.00
		Post-test	.00	.46	1.00
CG	Pre-test	Post-test	-4.50	.43	.00
		Delayed-test	-4.00	.33	.00
	Post-test	Pre-test	4.50	.43	.00
		Delayed-test	.50	.45	.27
	Delayed-test	Pre-test	4.00	.33	.00
		Post-test	- .50	.45	.27

Note. MD = mean difference; SE = standard error

Discussion

The results witnessed a significant performance of EG, since the average score of EG was approximately twice as large as CG in the post-test and delayed-test. The better performance of EG is largely due to the higher number of repetition. Studies (e.g., Webb, 2007) which examined the effectiveness of repetition on productive vocabulary acquisition illustrated that a higher number of repetitions might help students acquire productive knowledge of orthography, grammatical function, syntax and association, which was crucial for accurate language use. Take grammatical function as an example.

Students need to know the word *beauty* is a noun while the word *beautiful* is an adjective. Therefore, they could avoid ungrammatical utterances, such as 'She is a beauty lady.' Moreover, our quantitative results are aligned with the rationale of ILH, since both groups improved significantly better after the treatment. Of particular note, we thought participants in CG might forget how to apply the target words in the delayed-test. Perhaps most surprisingly, their productive vocabulary knowledge is still retained later. It seemed that tasks with high IL could also contribute to long-term retention. In the present study, we were not sure which (repetition or IL) carried more relative weight in terms of productive vocabulary's long-term retention (since EG students also performed well in the delayed-test). Therefore, further research comparing the effects of repetition and IL on vocabulary retention within a single study is required.

Limitations and Implications for Future Studies

First, the present study focused solely on improving students' learning performance. Extant studies have found learning productive vocabulary often frustrated students (e.g., Qian & Sun, 2019). And this may assuage negative feelings or reluctance to learn productive vocabulary in the curriculum. However, the present study did not consider 'motivation' or 'attitude' as variables. More nuance and clarity in how different types of motivational elements advance the effects of ILH and therefore result in better productive vocabulary performance are needed. Second, in a study that extensively relied on visual input, learning gains usually occur through repetitive visual input. Eye-tracking data could shed light on the role of repetition in the productive vocabulary learning process, yet this was not the purpose of the present study. In this respect, future studies should consider analyzing these electronic data to generate more robust and reliable results.

Conclusions

The present study is an attempt to investigate whether a higher number of repetition might bolster the effects of ILH and result in students' better productive vocabulary performance. As can be seen from the results, students appeared to benefit more from the combination of repetition and ILH. This study therefore provides additional evidence that other factors—as opposed to need, search, and evaluation—should be equal (Laufer & Hulstijn, 2001). More research with other proficiency groups is warranted to further refine our understanding of the potential factors contributing to the effects of ILH.

References

Alavinia, P., & Rahimi, H. (2019). Task types effects and task involvement load on vocabulary learning of EFL learners. *International Journal of Instruction*, *12*(1), 1501-1516.

Bao, G. (2015). Task type effects on English as a foreign language learners' acquisition of receptive and productive vocabulary knowledge. *System*, *53*, 84-95.

Elgort, I., Brysbaert, M., Stevens, M., & Van Assche, E. (2018). Contextual word learning during reading in a second language: An eye-movement study. *Studies in Second Language Acquisition*, *40*(2), 341-366.

Folse, S. (2006). The effect of type of written exercise on L2 vocabulary retention. *TESOL Quarterly*, *40*(2), 273-293.

Hornby, A. S. (2019). *Oxford advanced learner's English-Chinese dictionary* (7th ed.). Oxford: Oxford University Press.

Jafari, S., Izadpanah, S., & Rahmani, R. (2018). The effect of task-induced Involvement Load on unfamiliar L2 vocabulary learning: Sentence writing, summary writing, imaginary story writing and creative sentence writing. *Applied Research on English Language*, *7*(1), 67-88.

Keating, G. D. (2008). Task effectiveness and word learning in a second language: The involvement load hypothesis on trial. *Language Teaching Research*, *12*(3), 365-386.
Kim, Y. (2011). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. *Language Learning*, *61*, 100-140.
Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, *22*(1), 1-26.
Malmström, H., Pecorari, D., & Shaw, P. (2018). Words for what? Contrasting university students' receptive and productive academic vocabulary needs. *English for Specific Purposes (New York, N.Y.)*, *50*, 28-39.
Pellicer-Sánchez, A. (2016). Incidental L2 vocabulary acquisition from and while reading: An eye-tracking study. *Studies in Second Language Acquisition*, *38*(1), 97-130.
Pellicer-Sánchez, A., Conklin, K., & Vilkaitė-Lozdienė, L. (2021). The effect of pre-reading instruction on vocabulary learning: An investigation of L1 and L2 readers' eye movements. *Language Learning*, *71*(1), 162-203.
Pourakbari, A., & Biria, R. (2015). Efficacy of task-induced involvement in incidental lexical development of Iranian senior EFL students. *English Language Teaching*, *8*(5), 122-131.
Qian, Y., & Sun, Y. (2019). Autonomous learning of productive vocabulary in the EFL context: An action research approach. *Digital Scholarship in the Humanities*, *34*(1), 159-173.
Schmitt, N.(2010). *Researching vocabulary: A vocabulary research manual*. New York: Palgrave Macmillan.
Tahmasbi, M., & Farvardin, M. (2017). Probing the effects of task types on EFL learners' receptive and productive vocabulary knowledge: The case of Involvement Load Hypothesis. *SAGE Open*, *7*(3), 1-10.
Uchihara, T., Webb, S., & Yanagisawa, A. (2019). The effects of repetition on incidental vocabulary learning: A meta-analysis of correlational studies. *Language learning*, *69*(3), 559-599.
Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, *28*(1), 46-65.
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.
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.