An Expanded Conceptual Model of the Involvement Load Hypothesis for Incidental Vocabulary Acquisition

Liu Zhen
School of Foreign Languages, Hanshan Normal University, China
549888524@qq.com

Latha Ravindran
Faculty of Education, Languages, Psychology and Music, SEGi University
latharavindran@segi.edu.my

Mansour Amini
Faculty of Languages, Literacies and Translation, Universiti Sains Malaysia
mansour@usm.my

ABSTRACT

Since the Involvement Load Hypothesis was developed two decades ago, a colossal amount of empirical studies have been conducted to verify its predictive or explanatory power through comparison of various kinds of word-focused tasks with different involvement load indices. Extant findings barely lend a full credence to the prediction of the hypothesis, and modifications to its model were proposed but spread in different papers based on their experimental results with a limited number of variables manipulated. Despite these corrections, recent studies still use the old model to explain task comparison results. By integrating refinements supported by tangible experimental evidence, this study proposed an expanded theoretical framework of the Involvement Load Hypothesis with a new component, frequency of word usage, and modification to the tenets of the existing three components.

Keywords: involvement load hypothesis, incidental vocabulary acquisition, vocabulary task

1. Introduction

Vocabulary knowledge is a crucial component in second language acquisition closely relating to both L2 subskills and overall proficiency (Webb & Nation, 2017). In English learning,

vocabulary inadequacy is a major obstacle faced by many EFL learners (Goh & Ang-Aw, 2018) because of its astronomical vocabulary size. For at least 9,000 word families is considered to be sufficient to read extensively (Vilkaitė-Lozdienė & Schmitt, 2020), but such an amount is unreachable for most EFL learners, whose vocabulary size can rarely reach over 4,000 words after finishing secondary school education (Miralpeix, 2020). Therefore, EFL learners need to pick up words by themselves as a concomitant in reading books or watching videos (Pellicer-Sánchez, 2020; Perez, 2020). This method of learning is called incidental vocabulary acquisition, which was proved to be a feasible way to accumulate a large number of words but at a very slow pace (Thomas, 2020). So word-focus tasks while reading or listening are advocated to improve learning efficiency (Laufer, 2020). Although it is traditionally held that words rehearsed frequently and processed more elaborately tend to have stronger traces in memory, such mental effort cannot be operationalized nor can the relative effect of different tasks be predicted until the introduction of the Involvement Load Hypothesis (ILH) (Newton, 2020).

The ILH has been extensively tested to verify its predictive or explanatory power, and the two papers by Hulstijn and Laufer have been cited and referred to over 2000 times (Yanagisawa & Webb, 2021). However, the results produced are mixed. Very few research yield results exactly as the prediction of the hypothesis (San Mateo-Valdehíta & Chacón-García, 2019; Silva & Otwinowska, 2018), while most studies only partially support the ILH (Huang et al., 2012; Xie et al., 2017). Therefore, there are proposals to refine the hypothesis to boost its explanatory power, but the refinements advocated are subtle and spread in different papers based on their own findings from experiments focusing on a limited number of variables (Lee & Pulido, 2017; Teng & Zhang, 2021; Yang & Cao, 2021; Zou, 2017). Also, despite these refinements, recent studies still adopt the original model of the ILH to explain experimental results of task comparison (Ansarin & Khabbazi, 2021; Ehsani & Karami, 2022). There is still no overall theoretical framework accounting for vocabulary acquisition (Schmitt, 2019) and there have been calls to expand the original model of the ILH (Nation & Webb, 2011; Rassaei, 2017; Taheri & Rezaie Golandouz, 2021). To address this gap, this paper aims to refine the ILH by assembling the sporadic modifications proposed in the existing literature to an expanded

theoretical framework, which will serve for future empirical research on task effectiveness on incidental vocabulary learning.

2. Research Background & Framework

The ILH was conceived by Laufer and Hulstijin in their monumental paper (Laufer & Hulstijn, 2001), which indicated that the involvement load (IL) index of a certain task can be calculated through three components, Need, Search and Evaluation with levels of involvement signified through signs of minus (-) and (+). **Need** is of motivational dimension and has 3 degrees of prominence, with being absent (-) when a certain word is not used, moderate (+) when a word is required by external parties, or strong (++) when the need to use a word is intrinsically triggered (Laufer & Hulstijn, 2001). The other two components are of cognitive dimension attending to form-meaning mappings (Laufer, 2020). **Search** is to find the form or meaning of a given word with only two states, either absent (-) when words and their meanings are already provided in marginal glosses or moderate (+) to look up a word in a dictionary (Laufer, 2003). **Evaluation** is about comparison with three levels, with being absent (-), or moderate (+) when choosing a particular word among others or a meaning for a polysemantic word that fits a given context, or strong (++) when words being combined to create a grammatically and semantically legitimate original sentence (Laufer & Hulstijn, 2001). The involvement load of a task is the sum of the prominence degree carried by the three components and under equivalent conditions, and it assumes that the task with higher involvement load (abbreviated as IL in the following section) indices is more effective than a task with lower indices (Laufer, 2020). It was first tested by the authors in another paper (Hulstijn & Laufer, 2001), which served as a norm to test tasks with different involvement loads in experimental settings.

3. Review of Literature & Discussion

According to the meta-analysis by Yanagisawa and Webb (2021), the ILH only accounts for 15% of the variance of the ILH testing results of initial acquisition, and a much lower percentage, only 5%, is observed in delayed tests after one to four weeks. For instance, Laufer (2003), Eckerth and Tavakoli (2012), and Yang et al. (2017) found that tasks with higher IL significantly

outscored tasks with lower IL in the immediate test but such superiority disappeared after one or two weeks. There are also studies indicating that the ILH even fails to explain the results of the immediate test (Bao, 2015; Kaivanpanah et al., 2020; Keating, 2008; San Mateo-Valdehíta & Chacón-García, 2019). Therefore, there are two major strands of research attempting to boost the explanatory power of the ILH. One research direction is to account for the variance unexplained by the ILH model in the vocabulary learning process by examining moderating factors peripheral to the task, such as learners' characteristics like metacognitive strategies (Teng & Zhang, 2021), language aptitude (Yang & Cao, 2021) or working memory capacity (Ansarin & Khabbazi, 2021), and their interactive effects with tasks on vocabulary learning outcomes, and take further steps to propose scenarios or learning conditions that the ILH can better be accommodated to. Another strand of research is concerned with modifying and expanding the core components to increase the proportion of variance in vocabulary tests that ILH can explain (Laufer, 2020; Papi, 2018; Silva & Otwinowska, 2018; Taheri & Rezaie Golandouz, 2021; Zou, 2017). The following is to introduce the sources and elaborate on the justification for modification to the original ILH model. And to briefly demonstrate the involvement load of a certain task, the initial letters of the three components will be used followed by their degree of prominence. For example, a task with moderate involvement in all three components will be listed as "task name" (N+S+E+).

The Need Component

Although the need is determined to be strong when the learner voluntarily employs the targeted words in a task, in experimental settings it is generally held at a moderate level (Kaivanpanah et al., 2020; Rassaei, 2017; San Mateo-Valdehíta & Chacón-García, 2019; Teng & Zhang, 2021; Yang & Cao, 2021) because the motivation to use the targeted words is always imposed by the researcher, an external party (Laufer & Hulstijn, 2001). The self-driven motivation of the learner against the designated words employed is a paradox situation, which is not very likely to be resolved in the current stage. Rassaei (2017) marked the task of self-generating and answering questions to be of strong Need component. Although the question was composed autonomously by the learner, the targeted words were deliberately required in the task direction

to be used in the answers and thus violate the tenets of the ILH. Also, this experimental result didn't support the superiority of the question-generated task, which was less conducive to overall vocabulary learning than another task with equal IL indices (Rassaei, 2017). This shows that the effect of strong Needs remains to be explored (Yanagisawa & Webb, 2021). The perceived necessity to engage with a certain word is difficult to observe and varies greatly among the students or on the different words, and it is difficult to manipulate in empirical research (Tang & Treffers-Daller, 2016). In addition, if a strong Need continues to be viewed as a hidden impulse contingent solely by the learners, then it is not appropriate as a task model for a candidate. Instead, a strong Need can be a phenomenon associated with the word-focus task, and then this notion is operative in the experiment.

Motivation is unequivocally an important variable, whether intrinsic or extrinsic, which closely correlates with achievements of L2 vocabulary learning and permeates all stages of its learning process (Folse, 2011; Tanaka, 2017; Tseng & Schmitt, 2008). The magnitude of motivation has a tremendous impact on vocabulary learning outcomes (Lee & Pulido, 2017), but it is rather a complicated construct and is measured in different ways (Macintyre et al., 2009). However, Zheng (2012) indicated that motivation covaried with many other factors. So, it may provide a solution to distinguish the different magnitudes of this implicit entity through a more explicit phenomenon. Although no concrete scheme was proposed to refine the Need component, the following study may shed light on a way around to integrate the motivational construct into the ILH model. Instead of viewing motivation from a purely quantitative perspective, Papi (2018) applied the regulatory fit theory in the experiment to reveal a qualitative connection between the motivational and connective dimensions of second language learning and concluded that open tasks entailed creativity, which led to active participation and trigger eager behavior fitting a promotion focus. Whereas close task drives learners to adopt intent and considerate strategies to avoid mistakes and thus take a prevention stand according to the regulatory fit theory. Therefore, a task requiring creativity can initiate a stronger Need and more involvement in cognitive processing. However, since creativity can be more tangibly captured in the Evaluation component expatiated in the subsequent section, the Need in the new model will only be a binary variable with absent (-) or moderate (-) level.

The Search Component

Unlike Need and Evaluation, Search is a binary variable with only two levels at all attempts to locate the form or meaning of a word, such as consulting a dictionary or asking a teacher, and is considered to impose equal mental effort. In her recent paper, Laufer, who co-designed the ILH noted that the Search could be strong when the learner tried to locate a word form to express a familiar meaning (Laufer, 2020). However, no tenets or task instance was put forward in her paper to operationalize such a strong Search and the statement was vague with no clear boundary from moderate search. Because consulting an L1-L2 dictionary can fulfill the purpose of finding an L2 form for a familiar L1 meaning, but it is considered to be a moderate search in the original model. And fortunately, findings from Huang and Lin (2014) can be enlightening to propose a strong Search stipulation. With every targeted word occurring three times in the reading materials, they found that the group reading with gloss for the first occurrence, retrieving for the second occurrence, and with gloss again in the third, significantly and consistently outflanked another group reading with gloss in every occurrence all through the tests on form recall, meaning recall and meaning recognition conducted immediately after reading and even in another batch after two weeks. This finding suggests that opportunity to retrieve words from the mental lexicon after the initial encounter strongly escalates the form-meaning connections. Although there is no comparison of the gloss-retrieval task with dictionary consultation, which is supposed to induce moderate Search, previous research showed that looking up words in a dictionary cannot yield such an overwhelming and long-lasting effect compared to marginal gloss (Ghabanchi & Ayoubi, 2012; Un-udom, 2018). So, combining the new idea of Laufer (2020), a strong Search is to further retrieve word form internally from the mental lexicon after the initial construction of form-meaning connection.

There are also some other strong Search elements advocated, such as semantic mapping (Ong & Zhang, 2018) and guessing unknown words from context (Khoii & Sharififar, 2013). In addition, there is empirical evidence that lexical inferencing often leads to wrong answers (Nassaji, 2006) and has negative influence on encoding new words into long-term memory (Martínez-Fernández, 2008; Yoshii, 2013). Also, these two tasks entail more word comparison

to assess the suitability of a word in a sketch map or deduce its meaning through clues provided by other words around. That is how the word is used in context. Such cognitive efforts are Evaluation rather than Search behavior (Teng & Zhang, 2021), and therefore not suitable to integrate into the new model.

Evaluation Component

Evaluation, especially strong evaluation requiring vocabulary output, is considered to better facilitate vocabulary acquisition than input-based activities because it strengthens the form and meaning linkages of the words (Hu & Nassaji, 2016; Huang et al., 2012; Pichette et al., 2012). Laufer (2003) assigned tasks with the same overall IL indices, reading with a dictionary and filling in blanks (N+S+E+), and writing a composition with glossed words (N+S-E++) to her participants, but found that the composition task helped to retain the double amount of words two weeks later, suggesting that productive task left a deeper trace in memory. Wang et al. (2014) had a similar discovery that the group who read the passage and write sentences with glossed words (N+S-E++) remarkably outperformed another group conducting tasks with equivalent IL indices to read and fill blanks in the text (N+S+E+). Kim (2011) suggested that stronger evaluation imposed the greatest cognitive effort to process words and thus was the most significant factor in the ILH. Teng and Xu (2022) further demonstrated that the strong Evaluation component was so powerful to shadow the Search component and make it insignificant. Although strong Evaluation has been demonstrated to be more conducive to vocabulary learning, a further problem to be settled is that different output-oriented tasks with strong Evaluation produce significantly different effects.

Keating (2008) distinguished the effect of different strong Evaluation tasks by pointing out that generating cohesive discourse in composition writing imposed greater elaboration processing than writing separated sentences. Gohar et al. (2018) found that scaffolded by L1 glosses and example sentences, the composition writing group significantly outperformed the sentence writing group in a meaning recall test implemented. This test was implemented a week after the experiment and hierarchical multiple regression analysis revealed that the ILH only explain 48% of the variance of the scores, suggesting that the ILH failed to distinguish the different effects

of a connected discourse and isolated sentences. And Zou (2017) compared three evaluations task of output orientation accompanied by glosses, cloze-exercises (N+S-E+), sentence-writing (N+S-E++), and composition-writing(N+S-E++). And through a Vocabulary Knowledge Scale test on both receptive and productive knowledge of the words deployed immediately after the task and one week later, the author found that composition-writing > sentence-writing > cloze-exercises with a significance level smaller than 0.01 between each pair of post hoc comparison. And based on this quantitative result and findings from think-aloud protocols, and also drawing psychological concepts of chunking, hierarchical organization and pre-planning strategies, Zou (2017) put forward an augmented evaluation framework, in which evaluation (+) is moderate when the targeted words are being compared with others at phrasal level, or strong (+) when individual sentences are generated using the targeted words because chunking and pre-planning are involved, or very strong (++) to generate an original discourse with all targeted words chunking cohesively with a well-planned hierarchical structure.

Simply speaking, in Zou's model, the prominence of Evaluation depends on the scope of context in which the targeted words are processed - the more macro context the words are used, the higher involvement the task will be. However, the new model still fails to account for the differences among discourse writing tasks with the same strong Evaluation and overall IL indices revealed by recent studies. With meaning recognition tests, Rassaei (2017) found that compared to the tasks with the same load, summarizing the text (N+S-E++) or the task with higher IL indices, generating and answering questions (N++S-E++), the task of predicting and writing down the missing events in the story (N+S-E++) was more conducive (p<0.001) for vocabulary acquisition in both form recall and meaning recognition tests. The author emphasized that the creativity required in prediction triggered a greater involvement load than the reconstructive technique used to make a summary or generate questions, and called for a need to expand the Evaluation component. Taheri and Rezaie Golandouz (2021) also adopted a between-group design to compare three reading plus tasks with identical involvement load (N+S-E++), making sentences, summarizing the story read, or predicting and writing down the end of the story with L1 glosses given. The meaning recall was tested on 15 targeted words for each group and the result showed that prediction writing had a more pronounced effect (p<0.05)

on immediate vocabulary acquisition than the other two tasks. Besides attributing the success of the prediction task to its creative nature as Rassaei (2017), Taheri and Rezaie Golandouz (2021) also noted that participants in this group might enjoy more opportunities to employ the initially-learned words and other existing knowledge in a new linguistic context and thus enhance learning effect. Zhou and Wang (2021) investigated three tasks, continuation writing after reading a story, writing a summary for the text, and sentence paraphrasing in groups of Chinese senior English Majors. Vocabulary gains were collected through a modified Vocabulary Knowledge Scale (VKS) and statistical analysis showed that the continuation task group immediately acquired and later retained significantly more words than the other groups ($p < .05$). The authors explained the results by drawing on a psycholinguistic construct 'interactive alignment', which suggested that the reading then continuing writing process drove learners to converge their linguistic output to the text read previously and thus triggered the most intense interaction with the targeted words and input materials than the other two tasks. Since these studies consistently adduced empirical evidence that continuation or prediction writing tasks significantly help to acquire more unfamiliar words than other discourse generation tasks, and their findings were laid on the theoretical background, this type of task will be eligible to be manifested in the new model of the ILH.

In addition, when referring back to the *Need Section*, it can be seen that task triggering strong Need according to Papi (2018) has the features of being open, creative, and actively engaging, all of which are manifested through predictive or continuative writing task demonstrated in the aforementioned three studies. And in this way, the invisible strong Need can be externalized as an extra strong Evaluation and thus manipulative in the new model.

Frequency of word needed in the task

Frequency is an important variable exerting a great impact on the learning process and outcomes of all aspects of vocabulary (Schmitt, 2010). It is common sense that the more frequently a word is encountered, the higher chance it will be acquired. However, the frequency of encounter is not easy to be operated in a theoretical model because first, for tangible learning to take place, the number of times needed to meet the words varied greatly from 10 to over 20

encounters according to different research findings (Brown et al., 2008; Pellicer-Sánchez & Schmitt, 2010), and secondly, learners may read a section back and forth and artificially increase word occurrence. And although it has been proven that word-focus tasks can considerably reduce the number of encounters required in extensive reading to achieve similar acquisition rates (Laufer & Rozovski-Roitblat, 2015), there is still a lack of evidence the marginal times of meeting them to trigger meaningful learning. Although the exact times a word is encountered is tricky to operate, the frequency it is needed in a task can be grossly calculated. And when this variable is equalized across different tasks, the relative effects of different tasks tend to corroborate with the prediction of the ILH as it is shown by the experiment conducted by Teng and Xu (2022). However, the ILH failed to work in the situation in which the targeted words are rehearsed more times in low indices tasks than high indices ones. Folse (2006) compared two tasks, filling-blanks (N+S-E+) and writing sentences (N+S-E++), but with a group doing the first task with each word once, and another one doing the same task with the words three times. Vocabulary test results showed that the triple blank-filling group picked up significantly more words immediately than the writing sentences group, and the effect size of their difference was very large (*d*=.91). Lu (2013) also showed that words elaborated three times in the blank-filling task were much better acquired than those used for once in summary writing, which actually induced much higher involvement load. Ansarin and Bayazidi (2016) adopted a with-in subject design by assigning three multiple-choice tasks (N+S+E+), blank-filling (N+S+E+), and sentence-writing (N+S+E++) to 72 college EFLs, who were allowed to look up words in the dictionary. 18 targeted words were encountered three times in the former two tasks but only once in the sentence writing, and a meaning-recall test showed that both multiple-choice and blank-filling tasks significantly outperformed the sentence writing, which had the highest IL indices among the three, suggesting that frequency of word usage in task is a more important factor than an elaboration on promoting vocabulary acquisition.

Although the frequency of word usage in a task was repetitively shown to be a crucial factor mediating the effect of the task, no scheme has been proposed on how it could be merged into the ILH model. The following study may shed light on a solution. In Silva and Otwinowska (2018), although the frequency of word usage was not listed as a variable, a widely-employed

task taken as a whole unit in other ILH research was disintegrated into sub-tasks and assigned IL index to the corresponding components each time when targeted words were used. For instance, reading with gloss and making sentences traditionally will be labeled as (N+S-E++) with moderate Need, but in this study, Need was labeled twice as strong (++) because it was used twice for reading comprehension first and then writing the sentences respectively. With such a more dedicated scheme of IL indices calculation, their research finding was in line with the prediction of the ILH. Therefore, instead of being listed as a component parallel with the other three ones, the frequency of encountering the targeted words can be combined with other components in the model to boost its predictive power.

4. An Expanded conceptual framework of the ILH

Based on the literature review in section 3, here is an expanded theoretical framework of the ILH consisting of a motivational component (Need), two cognitive components (Search, Evaluation), and a physical dimension (frequency of word usage). And the following is the new framework with refined or expanded tenets in underscored form and its source addressed. In addition, since the original model was not accompanied by a visualized framework in Laufer and Hulstijn (2001), figure 1 was adapted from the chart provided by Teng and Zhang (2021). As it was stated in the original model (Laufer & Hulstijn, 2001), the IL indices of a specific task are the sum of the prominence degree of the following four components, and it is hypothesized that task with higher IL indices will be more effective on facilitating vocabulary learning than the one with lower indices.

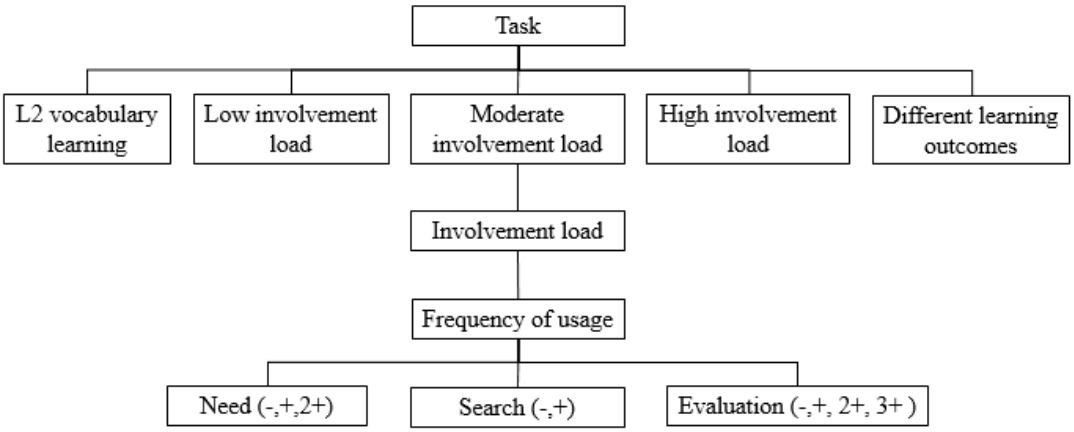

Figure 1: A Visualized Theoretical Framework of the Expanded ILH Model Adapted from Teng and Zhang (2021)

1. Need
    1) Absent (-) no need to use the targeted words in the task.
    2) Moderate (+) the targeted words being required to use in the task.
2. Search
    1) Absent (-) no search behavior, e.g. marginal gloss already provided for the targeted word;
    2) Moderate (+) to find the form or meaning of a new word, e.g. consult a dictionary;
    3) Strong **(++)** to retrieve newly-learned words from the mental lexicon to express a familiar meaning (Huang & Lin, 2014; Laufer, 2020).
3. Evaluation
    1) Absent (-) no comparison of words or decision on a meaning for a polyseme;
    2) Moderate (+) to select a suitable word among others or choose an appropriate meaning of a polyseme for a given context;
    3) Strong (++) to chunk the targeted word with others at the sentence level, e.g. generate original but disconnected sentences with every targeted word (Zou, 2017);
    4) Very Strong (+++) to chunk all the targeted words together coherently into a well-organized discourse, e.g. writing a composition or summary (Zou, 2017);
    5) Extra Strong (++++) to chunk all the targeted words together coherently into a well-organized discourse, which merges into or continues to elaborate the text read previously, e.g. predict the ending or write a continuation for a story (Papi, 2018; Rassaei, 2017; Taheri & Rezaie Golandouz, 2021; Zhou & Wang, 2021).
4. Frequency of word usage

    When words are used in a task multiple times, the motivational or cognitive components involved each time should aggregate a degree of prominence accordingly, e.g. need can be counted twice and labeled as being strong (++) in the task of reading with answering comprehension questions embedded with the words, which will then be used again in the subsequent cloze exercises (Silva & Otwinowska, 2018).

5. Conclusion

This study synthesized scattered proposals of modifications to the ILH model in the extant literature and constructed a more comprehensive theoretical framework to predict and explain the relative effects of different tasks facilitating vocabulary learning. Some of the new tenets, such as extra strong Evaluation, have been proven to be legitimate in prediction or continuation writing in recent literature. However, empirical research is in demand to test the effect of the new component, the frequency of word usage, and the feasibility of the way it is combined with the other three components. Additionally, a comparison of different tasks is to be done to verify the explanatory power of the new framework. Moreover, future studies are needed to explore the measurement of motivational construct and its linkage to the cognitive components.

Funding

This study is funded by Guangdong Philosophy and Social Science Foundation 13th Five-Year Plan Project 'On the impacts of multimedia input and task output on incidental vocabulary acquisition' (GD17XWW14).

References

Ansarin, A. A., & Bayazidi, A. (2016). Task type and incidental L2 vocabulary learning: Repetition versus task involvement load. *Southern African Linguistics and Applied Language Studies*, *34*(2), 135-146.

Ansarin, A. A., & Khabbazi, S. K. (2021). Task-induced involvement load and working memory: Effects on active and passive vocabulary knowledge of EFL learners in a multimedia learning environment. *Eurasian Journal of Applied Linguistics*, *7*(1), 277-302.

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

Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading while listening, and listening to stories. *Reading in a Foreign Language*, *20*(2), 136-163.

Eckerth, J., & Tavakoli, P. (2012). The effects of word exposure frequency and elaboration of word processing on incidental L2 vocabulary acquisition through reading. *Language Teaching Research*, *16*(2), 227-252.

Ehsani, M., & Karami, H. (2022). Comparing the predictive power of involvement load hypothesis and technique feature analysis. *International Journal of Language Studies*, *16*(2), 165-188.

Folse, K. (2011). Applying L2 lexical research findings in ESL teaching. *TESOL Quarterly*, *45*(2), 362-369.

Folse, K. (2006). The effect of type of written exercise on L2 vocabulary retention. *TESOL*

Quarterly, *40*(2), 273-293.

Ghabanchi, Z., & Ayoubi, E. S. (2012). Incidental vocabulary learning and recall by intermediate foreign language students: The influence of marginal glosses, dictionary use, and summary writing. *Journal of International Education Research*, *8*(2), 85-96.

Goh, C. C. M., & Ang-Aw, H. T. (2018). Teacher-Examiners' Explicit and Enacted Beliefs About Proficiency Indicators in National Oral Assessments. In D. Xerri & P. Vella Briffa (Eds.), *Teacher Involvement in High-Stakes Language Testing* (pp. 197-216). Springer International Publishing.

Gohar, M. J., Rahmanian, M., & Soleimani, H. (2018). Technique feature analysis or involvement load hypothesis: Estimating their predictive power in vocabulary learning. *Journal of Psycholinguistic Research*, *47*(4), 859-869.

Hu, H.-c. M., & Nassaji, H. (2016). Effective vocabulary learning tasks: Involvement Load Hypothesis versus Technique Feature Analysis. *The system*, *56*, 28-39.

Huang, L., & Lin, C. (2014). Three approaches to glossing and their effects on vocabulary learning. *System*, *44*, 127-136.

Huang, S., Eslami, Z., & Willson, V. (2012). The effects of task involvement load on L2 incidental vocabulary learning: A meta-analytic study. *The Modern Language Journal*, *96*(4), 544-557.

Hulstijn, J. H., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning*, *51*(3), 539-558.

Kaivanpanah, S., Alavi, S. M., & Ravandpour, A. (2020). The effect of input-based and output-based tasks with different and identical involvement loads on Iranian EFL learners' incidental vocabulary learning. *Cogent Psychology*, *7*(1), 1731223.

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.

Khoii, R., & Sharififar, S. (2013). Memorization versus semantic mapping in L2 vocabulary acquisition. *ELT Journal*, *67*(2), 199-209.

Kim, Y. (2011). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. *Language Learning*, *61*(s1), 100-140.

Laufer, B. (2003). Vocabulary Acquisition in a Second Language: Do Learners Really Acquire Most Vocabulary by Reading? Some Empirical Evidence. *The Canadian Modern Language Review*, *59*(4), 567-587.

Laufer, B. (2020). Evaluating Exercises for Learning Vocabulary. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (1st ed., pp. 351-368). Routledge.

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, *22*(1), 1-26.

Laufer, B., & Rozovski-Roitblat, B. (2015). Retention of new words: Quantity of encounters, quality of task, and degree of knowledge. *Language Teaching Research*, *19*(6), 687-711.

Lee, S., & Pulido, D. (2017). The impact of topic interest, L2 proficiency, and gender on EFL incidental vocabulary acquisition through reading. *Language Teaching Research, 21*(1), 118-135.

Lu, M. (2013). Effects of four vocabulary exercises on facilitating learning vocabulary meaning, form, and use. *TESOL Quarterly*, *47*(1), 167-176.

Macintyre, P. D., Mackinnon, S. P., & Clément, R. (2009). Toward the development of a scale

to assess possible selves as a source of language learning motivation. In D. Zoltán & U. Ema (Eds.), *Motivation, Language Identity and the L2 Self* (pp. 193-214). Multilingual Matters.

Martínez-Fernández, A. (2008). Revisiting the Involvement Load Hypothesis: Awareness, Type of Task and Type of Item. *Selected Proceedings of the 2007 Second Language Research Forum*, University of Illinois at Urbana-Champaign.

Miralpeix, I. (2020). L1 and L2 vocabulary size and growth. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (1st Edition ed., Vol. 1st, pp. 189-206). Routledge.

Nassaji, H. (2006). The relationship between depth of vocabulary knowledge and L2 learners' lexical inferencing strategy use and success. *The Modern Language Journal*, *90*(3), 387-401.

Nation, I. S. P., & Webb, S. (2011). *Researching and Analyzing Vocabulary*. Boston: MA, Heinle.

Newton, J. (2020). Approaches to learning vocabulary inside the classroom. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (1st ed., pp. 255-270). Routledge.

Ong, K. K. W., & Zhang, L. J. (2018). The effects of code-switched reading tasks on late-bilingual EFL learners' vocabulary recall, retention and retrieval. *System*, *72*, 13-22.

Papi, M. (2018). Motivation as quality: Regulatory Fit Effects on incidental vocabulary learning. *Studies in Second Language Acquisition*, *40*(4), 707-730.

Pellicer-Sánchez, A. (2020). Expanding english vocabulary knowledge through reading: Insights from eye-tracking studies. *RELC Journal*, *51*(1), 134-146.

Pellicer-Sánchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do Things Fall Apart? *Reading in a Foreign Language*, *22*(1), 31-55.

Perez, M. M. (2020). Incidental vocabulary learning through viewing video: The role of vocabulary knowledge and working memory. *Studies in Second Language Acquisition*, *42*(4), 749-773.

Pichette, F., de Serres, L., & Lafontaine, M. (2012). Sentence reading and writing for second language vocabulary acquisition. *Applied Linguistics*, *33*(1), 66-82.

Rassaei, E. (2017). Effects of three forms of reading-based output activity on L2 vocabulary learning. *TESL Canada Journal*, *21*(1), 76-95.

San Mateo-Valdehíta, A., & Chacón-García, C. (2019). Learning word class in a second language through vocabulary learning activities: definition-choosing, gap-filling, and sentence-writing. *Journal of Spanish Language Teaching*, *6*(1), 49-63.

Schmidt, R. W. (1990). The Role of Consciousness in Second Language Learning. *Applied Linguistics*, *11*(2), 129-158.

Schmitt, N. (2010). *Researching Vocabulary: A Vocabulary Research Manual* (1st ed.). Palgrave Macmillan.

Schmitt, N. (2019). Understanding vocabulary acquisition, instruction, and assessment: A research agenda. *Language Teaching*, *52*(2), 261-274.

Silva, B., & Otwinowska, A. (2018). Vocabulary acquisition and young learners: Different tasks, similar involvement loads. *International Review of Applied Linguistics in Language Teaching*, *56*(2), 205-229.

Taheri, S., & Rezaie Golandouz, G. (2021). The effect of task type on EFL learners' acquisition and retention of vocabulary: an evaluation of the involvement load hypothesis. *Cogent*

Education, *8*(1), 1915226.

Tanaka, M. (2017). Examining EFL vocabulary learning motivation in a demotivating learning environment. *System*, *65*, 130-138.

Tang, C., & Treffers-Daller, J. (2016). Assessing incidental vocabulary learning by Chinese EFL learners: A test of the Involvement Load Hypothesis. In G. Yu & Y. Jin (Eds.), *Assessing Chinese Learners of English* (pp. 121-149). Palgrave Macmillan.

Teng, M., & Xu, J. (2022). Pushing vocabulary knowledge from receptive to productive mastery: Effects of task type and repetition frequency. *Language Teaching Research*, 13621688221077028.

Teng, M. F., & Zhang, D. (2021). Task-induced involvement load, vocabulary learning in a foreign language, and their association with metacognition. *Language Teaching Research*, 13621688211008798.

Thomas, N. (2020). Incidental L2 vocabulary learning: Recent developments and implications for future research. *Reading in a Foreign Language*, *32*(1), 49-60.

Tseng, W.-T., & Schmitt, N. (2008). Toward a model of motivated vocabulary learning: A structural equation modeling approach. *Language Learning*, *58*(2), 357-400.

Un-udom, N. (2018). The Effects of Task-Induced Involvement on EFL Vocabulary Learning. *Chophayom Journal, 29*(3), 181–196.

Vilkaitė-Lozdienė, L., & Schmitt, N. (2020). Frequency as a guide for vocabulary usefulness high-, mid-, and low-frequency words In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (1st Edition ed., Vol. 1st, pp. 81-96). Routledge.

Wang, C., Xu, K., & Zuo, Y. (2014). The effect of evaluation factor on the incidental vocabulary acquisition through reading. *International Journal of English Linguistics*, *4*, 59–66.

Webb, S., & Nation, P. (2017). *How Vocabulary is Learned*. Oxford University Press.

Xie, H., Zou, D., Wang, F. L., & Wong, T.-L. (2017). A review on recent development of the involvement load hypothesis. In S. K. S. Cheung, L.-f. Kwok, W. W. K. Ma, L.-K. Lee, & H. Yang, *Blended Learning. New Challenges and Innovative Practices* Cham.

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.

Yang, Y., & Cao, X. (2021). Effects of task involvement load on L2 vocabulary acquisition and their association with language aptitude. *The Asia-Pacific Education Researcher*, *30*(5), 421-430.

Yang, Y., Shintani, N., Li, S., & Zhang, Y. (2017). The effectiveness of post-reading word-focused activities and their associations with working memory. *System*, *70*, 38-49.

Yoshii, M. (2013). Effects of gloss types on vocabulary learning through reading: Comparison of single and multiple gloss types. *CALICO Journal*, *30*, 203–229.

Zheng, Y. (2012). Exploring long-term productive vocabulary development in an EFL context: The role of motivation. *System*, *40*(1), 104-119.

Zhou, X., & Wang, C. (2021). Effects of interactive alignment on L2 vocabulary learning by Chinese EFL learners. *Language Teaching Research*, 13621688211004629.

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.