

# *The Effect of Reading Purpose on Incidental Acquisition and Retention of Vocabulary from Reading*

Xiaoyan MA

*Faculty of Education, University of Macau, MCSAR*

Barry Lee REYNOLDS\*

*Faculty of Education and Centre for Cognitive and Brain Sciences,  
University of Macau, MCSAR*

*\*Corresponding author: BarryReynolds@um.edu.mo*

*The main aim of this study was to determine whether different purposes for reading can result in more or less incidental vocabulary acquisition from reading. Three intact classes of Grade 7 junior high school students from Mainland China were recruited and oriented to read for different purposes: interest (n = 42), exam (n = 45), and baseline (n = 45). After reading, the three groups completed an unexpected receptive form and productive meaning test, followed by an unexpected receptive meaning test as an immediate posttest. Two weeks later, the participants were given the same assessments as the delayed posttest. The posttest results for all types of vocabulary knowledge showed that the interest group outperformed the exam group, which outperformed the baseline group. However, on the delayed posttests, the same pattern of results as for the posttest was shown only for receptive form knowledge. The productive meaning and receptive meaning assessment outcomes showed that the exam and interest groups had similar performance and outperformed the baseline group. The implication is that teachers should clearly orient learners prior to giving them any task that intends to promote vocabulary learning, especially those that involve reading.*

## Introduction

The importance of vocabulary in first language acquisition studies has long been recognized, and the same is true for second language acquisition (SLA) studies. According to Meara (2002), vocabulary acquisition is critical for language learning, as it allows one to master various types of language abilities, such as listening, speaking, reading, and writing. Early on, Wilkins (1972, p. 258) claimed that “without grammar, very little can be conveyed; without vocabulary, nothing can be conveyed.”

Even if the importance of vocabulary learning is understood, teachers may not be able to devote enough time to teaching all the vocabulary that is necessary to function in a language. This is because learners must recall thousands of words to be considered proficient (Nation, 2013). It would be impossible in one’s lifetime for a teacher to teach and a learner to learn all of the necessary words (Nation, 2013). In spite of this, some learners are continuously found to have mastered and acquired extensive vocabularies (Huckin & Coady, 1999).

Learners who have been successful in vocabulary learning have done so because most of their vocabulary learning occurs while they are engaged in other tasks such as reading, listening, speaking, and writing (Joe, 1998; Laufer & Hulstijn, 2001). After Nagy and colleagues (1985) hypothesized that most of the words that native English speakers acquired occurred through reading, especially during the school years, many SLA scholars began investigating whether this was true for second language learners as well. Since then, there have been numerous studies showing how incidental vocabulary acquisition (IVA) is possible as a by-product of engagement in various tasks, most often reading (e.g., Huckin & Coady, 1999; Laufer & Hulstijn, 2001; Qian, 2019; Schmitt, 1998; Teng, 2018).

While most vocabulary acquisition researchers agree that vocabulary can be acquired incidentally from reading, the vast body of research on IVA from reading has shown inconsistencies in the success of vocabulary learning (e.g., Elgort & Warren, 2014; Huckin & Coady, 1999; Liang, 2018; Nagy et al., 1985; Pulido, 2003; Rahul & Ponniah, 2020). As a way to explain these differences in previous studies, Laufer and Hulstijn (2001) proposed the involvement load hypothesis (ILH), which posits that varying levels of involvement lead to different levels of IVA. That is, the deeper the learners involve themselves in the tasks, the greater the

possibility that they will incidentally acquire unknown words encountered in the task. The ILH has been widely used to frame studies aimed at examining learners' IVA from reading (e.g., Arseven, 2013; Eckerth & Tavakoli, 2012; Kaivanpanah et al., 2020; Lee & Pulido, 2017; Silva & Otwinowska, 2018; Xie et al., 2017; Yang & Cao, 2021). However, some recent studies (e.g., Bao, 2015; Hazrat & Read, 2021; Hu & Nassaji, 2016; Huang et al., 2012; Yanagisawa & Webb, 2021) have criticized the precision of the ILH and its ability to predict IVA from reading.

These researchers have claimed that one of the components of the ILH is lacking. *Need*, which represents motivation for learners to complete a given task, considers only motivation as occurring due to internal and external forces. However, some researchers (e.g., Hazrat & Read, 2021; Hulstijn & Laufer, 2001; Vallerand, 2007; Vallerand & Ratelle, 2002) have begun to question whether a more nuanced view of motivation is needed.

Previous studies have indicated that young learners engage in reading for different purposes (e.g., Zhang & Duke, 2008). Sometimes young learners initiate their own engagement in reading, and sometimes this engagement is the result of teacher expectations at school. Even when teachers are responsible for this engagement, they can be more or less explicit about why learners should engage in reading. Thus, it is important for vocabulary researchers to understand these classroom contexts that might lead to different vocabulary acquisition outcomes (Reynolds, 2020). Therefore, in the current study, we considered different types of *need* and whether they could differentially affect IVA outcomes from reading. More specifically, this study investigated reading purposes: for interest (i.e., interest group readers), for passing an exam (i.e., exam group readers), and for no specific purpose (i.e., baseline group readers).

## **Literature Review**

### **Definitions of Incidental Vocabulary Acquisition**

There has been controversy regarding the role that IVA plays in second language learners' vocabulary growth (Cobb, 2007, 2008; Laufer, 2005; McQuillan & Krashen, 2008). These debates have attracted researchers' exploration of various factors affecting L2 learners' IVA from reading.

The usage of the term IVA is often associated with Laufer's (1989) work. Laufer (1989) defined IVA by contrasting it with purposeful language learning. IVA refers to a process in which learners pick up words when they are completing language learning tasks such as reading articles or listening to songs (Laufer, 1989). In contrast, purposeful language learning refers to language learners who memorize words with the purpose of learning vocabulary (Laufer, 1989). Similarly, Schmidt (1994, p. 7) characterized IVA as "learning without the intent to learn, or learning of one thing when the learner's primary objective is to do something else, such as communicating."

Recent studies have defined IVA by associating it with the completion of communicative activities. Nation (2013) defined IVA as a process in which learners' attention is drawn to the content of reading materials rather than learning words mechanically; that is, readers' attention is drawn to the information conveyed by the text. Joe (1998) also used the term to refer to the situation in which students' attention is drawn to reading comprehension rather than the learning of words. According to Huckin and Coady (1999, p. 183), IVA is the acquisition of new words due to engagement in "meaning-focused communicative activities such as reading or listening." Laufer and Hulstijn (2001) claimed that most words readers learn could be considered to have been acquired as a by-product of engagement in listening, speaking, reading, or writing activities. For the current study, we adopted Laufer and Hulstijn's (2001) definition, which indicates that vocabulary learning is a by-product of any activity completed when learners have not been informed that their vocabulary knowledge will be assessed in the future. The consequence of choosing this definition was that the three groups of learners were asked to engage in reading and had their acquisition of vocabulary assessed without being informed beforehand. Hence, IVA was seen as a by-product of the reading activity, and there was no expectation of any vocabulary assessment after the activity.

### **Empirical Studies on Incidental Vocabulary Acquisition from Reading**

The scope of research on incidental second language vocabulary acquisition is relatively broad. A certain degree of IVA will occur while engaging in various language activities, such as listening, speaking,

reading, and writing (Jin & Webb, 2020; Montero Perez, 2020; Van Zeeland & Schmitt, 2013; Zhao et al., 2016). However, according to the literature, research on IVA from reading has been explored the most (e.g., Liang, 2018; Qian, 2019; Rahul & Ponniah, 2020; Teng, 2018; Zhao et al., 2016). The research in this field involves many factors, including textual factors (e.g., Reynolds, 2016b; Reynolds & Ding, 2021) and learner factors (e.g., Elgort & Warren, 2014; Reynolds, 2016a; Reynolds & Bai, 2013; Tekmen & Daloglu, 2006).

One textual factor that significantly impacts learners' IVA is the frequency of exposure to the target words found in the texts given to research participants. Ellis (2002) proposed that the frequency of exposure to words could influence language learning because it helps to emphasize the target words and attract the reader's attention to those words. However, many researchers have been unable to settle on an exact number of encounters necessary for acquisition and have found that the role of repetition is often influenced by a variety of other factors (e.g., Pulido, 2004; Webb, 2008; Zahar et al., 2001).

Hence, learner factors also need to be considered when studying L2 learners' IVA. These factors include learners' vocabulary size (Tekmen & Daloglu, 2006) and learners' reading comprehension ability (Pulido, 2004), among others. Learners' vocabulary size is a significant factor that needs to be considered when testing IVA since participants' lexical coverage of the text influences text comprehension and unknown word guessing (Tekmen & Daloglu, 2006). According to some recent studies (Laufer, 2020; Laufer & Ravenhorst-Kalovski, 2010; Nation, 2006; Song & Reynolds, 2022), if learners intend to understand an article's content, they must know at least 95% of the vocabulary of the text read, while knowing 98% of the words is ideal.

Reading ability is another important factor that influences learners' IVA (Pulido, 2003). Pulido (2004) conducted an empirical study with 99 English-L1 adults learning Spanish-L2. Pulido (2004) found that the more proficient a learner's second language reading ability was, the more vocabulary was incidentally acquired and retained. Thus, it is necessary to control the reading ability of participants who are recruited for IVA studies, or there is the possibility that reading ability could confound the results.

In summary, a good IVA study requires strict control of both learner and textual factors. While many learner and textual factors have been

controlled in the previous studies reviewed above, the purpose of reading, an important motivating factor, has received much less attention. Thus, for the current study, reading purpose was manipulated as an independent variable, while learners' reading ability and existing vocabulary knowledge were controlled.

Previous researchers showed that orienting learners in some way prior to reading affected their processing of the texts and the learning that occurred from reading those texts (e.g., Narvaez & Van den Broek, 1999; Van den Broek et al., 2001). In general, researchers have suggested that readers are sensitive to their orientations before reading and that their IVA might be influenced by different reading purposes. Swanborn and de Glopper (2002) explored the impact of reading purpose on IVA for LI reading. The results showed that reading to learn more about a topic led to better IVA outcomes than reading for text comprehension or free reading. Swanborn and de Glopper (2002) explained these results by suggesting that learners who read for text comprehension had the impression that they could refer to the text afterward when taking the test. Hence, learners might regard it as unnecessary to pay much attention to learning new words when they are reading. Swanborn and de Glopper (2002) also advised that further studies should focus more on motivational factors for reading to better understand the relationship between learners' motivation and IVA.

Similarly, Joe (1998) examined the effects of text-based tasks on IVA. The results showed that reading and retelling a text promoted better IVA than reading a text without retelling. Joe (1998) suggested that the deeper learners engaged in an activity, the greater the vocabulary gains from reading. Therefore, Joe (1998) proposed that different text-based tasks or different reading purposes would lead to different retrievals of the previously unknown words incidentally acquired from reading those texts. More recent study results (Nguyen & Boers, 2019; Peters et al., 2009) have also indicated that different reading tasks (e.g., retelling and comprehension) can also differentially affect learners' vocabulary acquisition. Nguyen and Boers (2019) found better word learning when learners were asked to recount the contents of the language input through retelling than when learners were exposed to the language only through input. Similarly, Peters et al. (2009) found that learners' awareness of an upcoming test and post-reading vocabulary tasks both affected the vocabulary learning outcomes from language exposure.

Reynolds and Bai (2013) also discovered that freedom of reader choice has a positive effect on IVA. Reynolds and Bai (2013) conducted a study with 78 EFL learners in Taiwan who were separated into two groups: those who read texts based on their own choice of a topic and those who were assigned a topic to read by a computer system. The results showed a positive correlation between learners' interest in the topic of the reading materials and IVA results. This study indicated that when learners were given freedom of choice, their level of interest—motivation—in reading the articles increased, hence positively influencing IVA. However, since this study examined only two motivational conditions (i.e., amotivation and intrinsic motivation), further study is needed.

In summary, the limited research on reading purposes indicates that IVA for both L1 and L2 readers is affected by different motivational conditions. While some studies have investigated how reading purpose can affect reading, less is known about how reading purpose can motivate learners to process texts more or less deeply and how this level of processing might further affect the incidental gains in vocabulary knowledge from reading.

### **Involvement Load Hypothesis (ILH)**

The ILH was proposed based on the Depth of Processing Hypothesis ( Craik & Tulving, 1975). The Depth of Processing Hypothesis states that the depth of processed information determines how long learners remember the knowledge that was processed (Craik & Tulving, 1975). However, lacking an operational or practical definition, the Depth of Processing Hypothesis could not be directly applied to research (Laufer & Hulstijn, 2001). Thus, the ILH was developed to address the limitations of the Depth of Processing Hypothesis (Laufer & Hulstijn, 2001).

The ILH focuses on the acquisition of second language vocabulary in task-based settings. According to Laufer and Hulstijn (2001), involvement load consists of three components: *need* (N), *search* (S), and *evaluation* (E). *Need* is a non-cognitive and motivational characteristic of input with two levels: moderate and strong (Laufer & Hulstijn, 2001). *Need* is moderated when an external force imposes a need to complete a task. For example, when a teacher instructs learners to read, their *need* is moderate and extrinsically motivated. However, *need* will be strong when self-imposed, for example, when students choose to read because of their interest.



When learners have no apparent need to complete a task, either internally or externally imposed, *need* is missing.

However, unlike *need*, *search* and *evaluation* are two cognitive components. The term *search* refers to the process of looking up the meaning of unfamiliar words in reference materials or getting this information from a peer or teacher (Laufer & Hulstijn, 2001). In other words, the endeavor to identify a connection between the form and meaning of an unknown word is known as *search*. For example, when a learner tries to find the L2 translation of an L1 word by actively looking up the word in a dictionary, the *search* of this learner is strong; when the learner is asked to do so by external authorities such as the teacher, then the *search* is moderate. When learners are forbidden to search or do not actively look up an unfamiliar word, *search* is missing.

Finally, *evaluation* entails making a choice “based on a criterion of semantic and formal appropriateness of a word and its context” (Laufer & Hulstijn, 2001, p. 15). *Evaluation* entails making a choice on the meaning of a given word, comparing its meaning to that of other words, or determining its appropriate use in a certain context. For example, *evaluation* is moderate when learners are required to recognize the differences between words or the differences between several senses of a word; however, “*evaluation* is strong when the evaluation requires making a decision about additional words that will combine with the new word in an original sentence or text” (Laufer & Hulstijn, 2001, p. 15). When learners do not recognize the differences between words or evaluate the usage of a new word in a text, *evaluation* is missing.

The three components of the ILH (*need*, *search*, and *evaluation*) are measurable in terms of their prominence. If a component is missing, it receives a score of 0; if the component has a moderate presence, it is given 1; and the component receives 2 if there is a strong presence. The points for each of the components can be added together to give an involvement score for an IVA task; this allows the comparison of different tasks to be easy and straightforward.

In summary, the ILH supports the idea that the retention of previously unknown words encountered during the completion of tasks depends on the involvement load calculated for those tasks (Laufer & Hulstijn, 2001). In this study, different reading purposes were given corresponding scores to represent the degree of *need*. For the present study, both the exam group (*need* = 1; *search* = 0; *evaluation* = 0) and the baseline group (*need* = 1; *search* = 0; *evaluation* = 0) received 1 for



their involvement load because they were externally motivated, while the interest group received 2 (*need* = 2; *search* = 0; *evaluation* = 0) because this group was intrinsically motivated. The reason why *search* and *evaluation* were both 0 for all three groups was that the learners were forbidden to look up any unfamiliar words in a dictionary while reading, and they did not need to recognize different meanings of words or evaluate different usages of words while reading.

### **Research Questions**

Based on the ILH, the present study should find that the incidental vocabulary outcomes for the interest group are higher than those for both the exam and baseline groups. In addition, IVA outcomes for the exam and baseline groups should be the same. Thus, a quasi-experimental study was conducted to investigate these assumptions and was guided by two research questions:

1. Which reading-purpose condition is the most effective in enhancing vocabulary learning (immediate effects) through reading?
2. Which reading-purpose condition is the most effective in enhancing vocabulary retention (delayed effects) through reading?

## **Methodology**

### **Research Design**

A quasi-experimental between-subjects design was used in the current study. A quasi-experimental design “aims to find a cause-and-effect link between an independent and dependent variable” (Gibbons & Herman, 1997, p. 2). In circumstances where a real experiment is impossible due to ethical or practical reasons, it is useful to have a quasi-experimental design. In this study, using a between-subjects design, one intact class was assigned to each of the three reading conditions.

### **Participants and Context**

Altogether, 140 Mainland Chinese junior high school students studying English as a foreign language participated in this study. Among these participants, 72 were male and 68 were female. All of them were in

Grade 7, with an average age of 13.3 years ( $M = 13.326$ ,  $SD = 0.598$ ), and they had been studying English for more than five years at the time the study was conducted. None of the participants had lived or studied in an English-speaking country for more than six months. The three intact classes were randomly assigned to three conditions: those who read for passing an exam (exam group) ( $n = 46$ ), those who read for interest (interest group) ( $n = 48$ ), and those who read for no specific purpose (baseline group) ( $n = 46$ ). Because the aim of this study was to determine the effect of reading purpose on IVA, the researchers ensured that these three groups were homogeneous. All the participants

1. were native Chinese speakers who had studied English as a foreign language for four years in primary school and one year in junior high school;
2. used the same English textbooks and additional learning materials;
3. had nearly identical educational backgrounds in Guangdong province, China;
4. voluntarily participated in the study;
5. were made aware that participation would not influence their grades or school evaluations.

The literature revealed that vocabulary size and reading ability are two factors that affect the IVA of L2 learners. These two factors were carefully considered and well-controlled in the present study. To control the English reading proficiency of the participants, the international Cambridge Key English Test (KET) was used. The descriptive statistics of the KET results for the baseline group were  $M = 11$ ,  $SD = 4.457$ ; for the exam group,  $M = 10.222$ ,  $SD = 4.517$ ; and for the interest group,  $M = 10.119$ ,  $SD = 2.998$ . The test of homogeneity of variances was checked, and the result showed that the assumption of equal variance was fulfilled ( $p = .202$ ). A one-way ANOVA was run to determine whether the KET results for the three groups were significantly different. The one-way ANOVA did not reveal any statistically significant differences among the reading proficiency of the three classes,  $F(2, 129) = 0.619$ ,  $p = .540$ . Thus, it can be considered that the participants' reading proficiency among the three groups was the same.

The Updated Vocabulary Levels Test (UVLT) (Webb et al., 2017) was also used to compare the three groups' existing vocabulary knowledge. The overall UVLT descriptive statistics for the baseline group were  $M = 46.267$ ,  $SD = 9.766$ ; for the exam group,  $M = 46.4$ ,  $SD = 8.486$ ;

and for the interest group,  $M = 46.024$ ,  $SD = 6.888$ . The test of homogeneity of variances was checked, and the result showed that the assumption of equal variance was fulfilled ( $p = .239$ ). Furthermore, a one-way ANOVA was run to determine whether the overall UVLT results for the three groups were significantly different. The one-way ANOVA did not reveal any statistically significant differences among the overall vocabulary levels of the three groups,  $F(2, 129) = 0.022$ ,  $p = .979$ .

Meanwhile, according to the UVLT results, all the participants in the current study had mastered the first 1,000-word level. The descriptive statistics for the first 1,000-word level result in the baseline group were  $M = 29.333$ ,  $SD = 0.477$ ; for the exam group,  $M = 29.311$ ,  $SD = 0.468$ ; and for the interest group,  $M = 29.310$ ,  $SD = 0.468$ . The test of homogeneity of variances was checked, and the result showed that the assumption of equal variance was fulfilled ( $p = .611$ ). A one-way ANOVA did not reveal any statistically significant differences among the three groups,  $F(2, 129) = 0.135$ ,  $p = .874$ .

Thus, it can be considered that when analyzing the overall results or the first 1,000-word level results, there were no statistically significant differences among the three participant groups' existing vocabulary knowledge. In summary, the three groups of participants had similar English reading proficiency and vocabulary levels. Hence, the three groups of participants were suitable for the current study.

## **Materials and Instruments**

### *Reading Material*

The current study aimed to understand the effect of reading purpose on EFL learners' IVA from reading. Thus, one critical step was to choose appropriate reading material. To ensure that the complexity of the selected reading material was suitable for the participants in the current study, Laufer's (2020) recommendation that the rate of unknown words should be between 2% and 5% was followed. This means that learners should recognize at least 95% of the tokens in the reading material to allow for IVA to occur.

Although we acknowledge the possible limitation that the participants in the baseline group were potentially interested in the same topics as the interest group, the interest group's selection of the reading material topic dictated what all three groups read. This selection process was completed

by providing the interest group with a list that consisted of 50 different reading topics (see Appendix A for the list of 50 topics) extracted from their junior high school English textbook (Liu et al., 2012). This textbook considered students' life experiences and learning needs. The interest group was asked to mark all topics that they would like to read more about. According to the results, each participant in the interest group ( $n = 48$ ) chose approximately 16 topics ( $M = 16.104$ ,  $SD = 5.890$ ) out of the total 50 topics. Among these topics, *trips and vacation* was selected by 42 out of the 48 participants in the interest group.

As the selected topic was broad, we further provided the titles of three passages related to *trips and vacation* for the participants in the interest group to choose from. Finally, the participants in the interest group selected the most popular passage with the title "Something You Need to Know about the UK" (Liu, 2020) (see Appendix B). The passage introduces some different aspects of the United Kingdom, including its history, weather, sports, and some tips for tourists. After the process of target word selection, this text was simplified so that it met the criteria of Laufer's (1997) recommendation for 95% lexical coverage. Finally, after the adaptation, the passage contained 576 tokens, of which 96.7% occurred in the first 1,000-word families of English, as verified by *Range for Texts* v.5.1 (Cobb, 2015). The text was also read and verified by a native English speaker to ensure that, after simplification, there were no oddities or abnormalities caused by this process. Based on the participants' UVLT results, in addition to the target words, they should have had knowledge of the other tokens that appeared in the text. Last, the learners' classroom teacher also confirmed that there was no grammar structure that appeared in the text that had not been previously taught to the learners.

## Measures

### *Receptive Form Test*

Different types of vocabulary knowledge were tested in the current study. Each target word appeared in the text with only one meaning sense. The receptive form test (see Appendix C) included multiple-choice items that asked the learners to choose the correct form among four options.<sup>1</sup> The items were scored dichotomously. Learners were given one

point when they correctly selected the key among the four options. The total score was six for the receptive form test. Cronbach's alpha for the receptive form test items on the posttest was .756, and the delayed posttest was .710, indicating fair internal consistency (Larson-Hall, 2010).

The format of the receptive form test is presented below. The learners were asked to choose which word they saw in the reading.

Sample Question 1

---

Which word did you see in the reading? If possible, write the Chinese translation of the word in the blank.

- A. consist
  - B. consist
  - C. konsist
  - D. komsist
- 

*Productive Meaning Test*

Productive meaning knowledge was tested at the same time as the receptive form test (see Appendix C). The productive meaning test included meaning translation questions that required the learners to write down the Chinese translations for the six target words. We created an answer bank for the productive meaning test that included the meanings only of the target words used in the text. This answer bank was referenced when marking the items to increase reliability. These items were also scored dichotomously. Learners were given one point when they correctly answered one of the correct translations according to the answer bank. The total score was six for the productive meaning test. The first author scored the productive meaning test for each participant twice with an interval of one week in between. Finally, we found 100% intra-grader reliability when comparing the results of the two scoring systems.

The format of the productive meaning test is presented below. The learners were asked to provide the Chinese translation after choosing which word they saw in the reading.

## Sample Question 2

Which word did you see in the reading? If possible, write the Chinese translation of the word in the blank.

- A. consist
- B. consist
- C. konsist
- D. komsist

Chinese Translation: \_\_\_\_\_

Four comprehension questions were also added to the receptive form and productive meaning test. These comprehension questions worked as distractors to ensure that the study's purpose was not revealed (Schmitt, 2010). At the same time, they were used to check whether the participants did indeed read the provided material during the experiment. The total score for the comprehension questions was 4, and the average score for the baseline group was 2.48; for the exam group, it was 3.57; and for the interest group, it was 3.32. Cronbach's alpha for the comprehension test items in the posttest was .651 and in the delayed posttest it was .862, indicating fair internal consistency—likely due to the small number of items on the test. Finally, it should also be noted that two versions of the receptive form and productive meaning test were created. One was given as the posttest, and one was given as the delayed posttest. The content was the same; however, the ordering was randomized to reduce the chance of practice effects (Nation & Webb, 2011; Schmitt, 2010).

### *Receptive Meaning Test*

The receptive meaning test (see Appendix D) included multiple-choice items that asked the learners to choose the correct meaning of a target word among four options. The items were scored dichotomously, with one point given to each item when a learner selected the correct key. Some nontarget words were also added to the receptive meaning test because it helped to make the target words less noticeable and reduced the chance of learners identifying the real target words (Schmitt, 2010). Cronbach's alpha for the receptive meaning test on the posttest was .790 and on the delayed posttest was .775, indicating fair internal consistency.

The format of the receptive meaning test is as follows. The learners were asked to choose the correct meaning of the target word they saw in the reading.

Sample Question 3

consist:

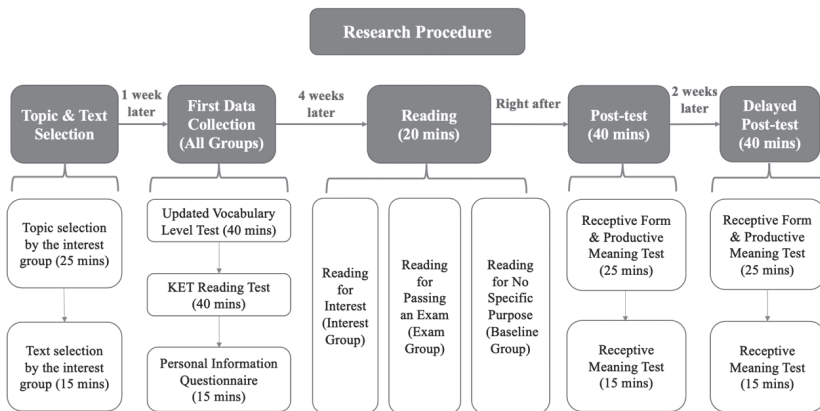
- A. 檢舉                      B. 剔除                      C. 組成                      D. 堅持

Finally, it should also be noted that two versions of the receptive meaning test were created. One was given as the posttest, and one was given as the delayed posttest. The content was the same; however, the ordering was randomized to reduce the chance of practice effects (Schmitt, 2010).

**Research Procedure**

The study lasted for one and a half months (see Figure 1). Data were collected during regular class time. Before the study began, topic selection and text selection activities were conducted with the interest group to ensure that the participants in the interest group were intrinsically motivated to read the provided text. Then, on the first day of the study, the UVLT, the KET reading test, and the personal information questionnaire were used to collect data from all participants and were administered in the following order: UVLT (40 minutes), KET reading test (40 minutes), and personal information questionnaire (15 minutes).

**Figure 1. Research Procedure**



Each of the three classes was assigned to one reading condition. The interest group read the article that they had previously selected based on the topic and text selection; the exam group was given the text and was



told they would have a comprehension test on the content; and the baseline group was given the text and was just told to read. All the participants were unaware that their vocabulary knowledge would be tested since this orientation could lead to the learners' awareness of the research purpose and violate the study's incidental nature (Nation & Webb, 2011). Each group received 20 minutes to finish the reading. After reading, the articles were gathered, and the participants were first given the receptive form and productive meaning test and then the receptive meaning test. This order was applied in the current study because we wanted to reduce the potential practice effect caused by the completion of the tests. Two weeks later, an unexpected, delayed posttest was administered in the same order as the posttest. The participants were not allowed to keep any of the materials provided to them.

### **Data Analysis**

Among the 140 participants, two students did not meet the requirement of having mastered the first 1,000-word families in English. Meanwhile, six students in the interest group were not interested in the final topic chosen by the majority of the learners in the interest group. Thus, these eight participants' data were not analyzed. Finally, the vocabulary learning and retention results of 132 participants were analyzed in SPSS 26.0. Repeated measures ANOVAs were used to analyze the immediate posttest and the delayed posttest scores to determine whether the differences among the three groups were statistically significant. The results of these tests were also compared among the three groups by calculating their effect sizes. By analyzing these quantitative data, the effect that motivation had on learners' incidental acquisition and retention of vocabulary was illuminated.

## **Results**

### **Descriptive Statistics**

Table 1 presents the descriptive statistics for the immediate and delayed posttest scores for receptive knowledge of form, receptive knowledge of meaning, and productive knowledge of meaning organized by groups. As shown in Table 1, all three groups performed better on the immediate

posttest than on the delayed posttest regardless of vocabulary knowledge type. Additionally, all three groups showed higher scores on the receptive form test than on the receptive meaning test. However, all three groups scored lowest on the productive meaning test.

The same pattern in the descriptive statistics was found for the immediate and delayed receptive form, receptive meaning, and productive meaning posttests. Specifically, the interest group always outperformed the exam group and the baseline group; the exam group always outperformed the baseline group.

**Table 1. Descriptive Statistics**

Test	Group	Immediate Posttest		Delayed Posttest	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Receptive Form	Interest Group ( <i>n</i> = 42)	4.357	1.186	2.524	1.065
	Exam Group ( <i>n</i> = 45)	2.622	1.193	1.778	0.951
	Baseline Group ( <i>n</i> = 45)	1.911	1.221	1.022	0.783
Receptive Meaning	Interest Group ( <i>n</i> = 42)	4.234	0.983	2.095	0.983
	Exam Group ( <i>n</i> = 45)	2.156	0.928	1.889	0.775
	Baseline Group ( <i>n</i> = 45)	0.867	0.842	0.511	0.661
Productive Meaning	Interest Group ( <i>n</i> = 42)	2.167	1.146	1.024	0.924
	Exam Group ( <i>n</i> = 45)	1.178	0.984	0.911	0.733
	Baseline Group ( <i>n</i> = 45)	0.333	0.522	0.2	0.505

Note. The total score for the mean (*M*) scores is 6.

The descriptive statistics in Table 1 show a higher level of IVA from reading when learners were reading for interest compared to reading for passing an exam or reading for no specific purpose. This trend was present for both the immediate and delayed posttests for all three types of vocabulary knowledge assessed: receptive form, receptive meaning, and productive meaning.

### Inferential Statistics

To answer the research questions, three repeated-measures ANOVAs were conducted to compare the three groups' incidental acquisition and retention of three types of vocabulary knowledge: receptive form, receptive meaning, and productive meaning. Post hoc LSD tests were also conducted to further understand the differences in the performance of the three groups. In addition, effect sizes (*d*) were calculated among the three groups (interest group, exam group, and baseline group) using

Plonsky and Oswald's standards (Plonsky & Oswald, 2014). The effect sizes can visually display the effect that reading purpose has on learners' incidental acquisition and retention of the targeted vocabulary (Wei et al., 2019).

Before running the ANOVAs, it was important to check for assumptions (Larson-Hall, 2010). Mauchly's test of sphericity was checked, showing that the assumptions of equal variances were fulfilled. The results of the ANOVAs run on the data found a significant interaction effect. Hence, pairwise comparisons were conducted, and it was found that reading purpose had a statistically significant effect on the IVA of receptive form knowledge,  $F(2, 129) = 53.634$ ,  $p < .001$ , receptive meaning knowledge,  $F(2, 129) = 127.120$ ,  $p < .001$ , and productive meaning knowledge,  $F(2, 129) = 38.699$ ,  $p < .001$ . The LSD post hoc tests run on the delayed posttest results found both similar and different patterns in the results as those shown for the posttest results. Specifically, the LSD post hoc tests found the same pattern in the delayed posttest results as the posttest results for the acquisition outcomes for receptive form knowledge: all three groups performed significantly differently from one another, with the interest group outperforming the exam group and the exam group outperforming the baseline group. However, for receptive meaning and productive meaning knowledge, the pattern in the results in the LSD post hoc tests was different for the posttests and the delayed posttests. The LSD post hoc test results from the analysis of the delayed posttest data showed that the interest group and the exam group were significantly different than the baseline group, but significant differences were not found between the interest group and the exam group. This pattern in data was the same for both the receptive meaning and productive meaning test scores.

To further understand the effect of the differences between the groups, effect sizes for the immediate posttest were calculated. For the receptive form test, when compared with the baseline group, both reading for interest ( $d = 2.44$ ) and reading for passing an exam ( $d = 1.08$ ) had a large effect on the acquisition of receptive form knowledge. When comparing the two experimental groups, a medium ( $d = 0.74$ ) effect was shown. For the receptive meaning test, when compared with the baseline group, both reading for interest ( $d = 3.69$ ) and reading for passing an exam ( $d = 1.46$ ) had a large effect on the acquisition of receptive meaning knowledge. When comparing the two experimental groups, a large effect ( $d = 2.17$ ) was shown. Similarly, for the productive meaning

test, when compared with the baseline group, both reading for interest ( $d = 2.06$ ) and reading for passing an exam ( $d = 1.07$ ) had a large effect on the acquisition of receptive meaning knowledge. Finally, when comparing the two experimental groups, a large ( $d = 1.03$ ) effect was shown.

The effect sizes for the delayed posttest were also calculated. For the receptive form test, when compared with the baseline group, both reading for interest ( $d = 1.48$ ) and reading for passing an exam ( $d = 1.05$ ) had a large effect on the acquisition of receptive form knowledge. When comparing the two experimental groups, a large effect ( $d = 1.46$ ) was shown. For the receptive meaning test, when compared with the baseline group, both reading for interest ( $d = 1.89$ ) and reading for passing an exam ( $d = 1.92$ ) had a large effect on the acquisition of receptive meaning knowledge. When comparing the two experimental groups, a small effect ( $d = 0.23$ ) was shown. Similarly, for the productive meaning test, when compared with the baseline group, both reading for interest ( $d = 1.11$ ) and reading for passing an exam ( $d = 1.14$ ) had a large effect on the acquisition of receptive meaning knowledge. When comparing the two experimental groups, a negligible effect ( $d = 0.14$ ) was shown. In other words, both the interest group and the exam group have significantly different results on learners' IVA delayed posttest scores compared to learners who read without a specific reading purpose for all types of vocabulary knowledge assessed.

## Discussion

In summary, learners who were intrinsically motivated to read had better IVA outcomes than learners who were extrinsically motivated to read. The results partially supported the ILH, indicating that intrinsic motivation can trigger more incidental learning than extrinsic motivation. However, in the delayed receptive meaning and the delayed productive meaning tests, we failed to find a significant difference between the interest group and the exam group. Additionally, contrary to the results predicted by the ILH, the exam group and the baseline group showed statistically significant differences even though they were accorded the same involvement load.

The current study examined the effect of reading purpose on learners' incidental learning and retention of vocabulary. In general, all three groups of learners performed better on the immediate posttest than

on the delayed posttest. This phenomenon naturally occurs because recently learned vocabulary knowledge is bound to decay over time (McLaughlin et al., 1983). This is because without any further priming that occurs through practicing and reviewing, it is natural for vocabulary gains from incidental learning to decay because this learning is fragile (Nation, 2013).

When analyzing the results of the different vocabulary assessments for both the immediate and delayed posttests, the findings revealed that different types of vocabulary knowledge were acquired under different motivated reading conditions. The learners were found to have incidentally acquired more receptive form knowledge than receptive meaning knowledge, followed by productive meaning knowledge. These results are aligned with previous studies (e.g., Schmitt, 1998; Webb, 2008) that found learners attained form knowledge first before meaning knowledge. The results of the current study also supported the results of previous studies in that the learners were found to have acquired and retained more receptive knowledge than productive knowledge (e.g., Bao, 2015; Nation & Webb, 2011).

Largely, the results of this study showed that the learners who completed reading under an ambiguous reading condition (i.e., the baseline group) acquired significantly less vocabulary knowledge than those learners who were given a purpose for reading (i.e., reading for interest; reading for passing an exam), whether it was intrinsically or extrinsically motivated. The results showed that the effect of reading purpose on IVA was pronounced in both the posttest and delayed posttest and was further shown by calculating effect sizes. Hence, the conclusion drawn from this finding is that reading with a specific purpose promotes learners' IVA, while both reading purposes, whether intrinsically or extrinsically motivated, have large effects on learners' IVA. When learners read in English for specific purposes, such as passing an exam or learning how to write, they may be motivated differently than learners who were not given a purpose.

While the results pointed to an obvious advantage of providing learners with a purpose for reading, further analysis revealed that different purposes could lead to more or less IVA. On the posttest, specifically, learners who were intrinsically motivated due to their interest in the topic acquired significantly more vocabulary knowledge than learners who were extrinsically motivated to pass an exam. This

result was shown for all three types of vocabulary knowledge assessed. The findings can partially support Laufer and Hulstijn's (2001) ILH, which claims that learners should attain better IVA when they are intrinsically motivated rather than extrinsically motivated. Meanwhile, the current study's findings supported the previous literature and further found that interest was a key factor influencing learners' vocabulary gains and retention. For example, the findings support Reynolds and Bai (2013), indicating that learners who were given the choice to read about topics they were interested in performed better in their IVA than learners who were assigned topics to read about.

The reading behavior of the intrinsically and extrinsically motivated groups could have also been affected by their purpose for reading. In Mainland China, reading exams often include multiple-choice comprehension questions. However, reading exams seldom assess vocabulary knowledge (Yan, 2015). Hence, it is possible that the learners in the exam group might have had this expectation in mind when they were reading, resulting in their attempts at gaining general comprehension knowledge without deriving the meaning of unknown words encountered when reading. These unknown words may not have been necessary for them to gain a general understanding of the article. In this way, the learners who read for the sake of passing an exam might have paid less attention to some of the unfamiliar words. On the other hand, the learners who read for interest may have paid more attention to other aspects of the article rather than just attaining general comprehension, as there was no expectation that they would have their comprehension assessed once the reading was completed. Thus, this might be the reason why learners in the exam group performed worse in their IVA than learners who read for interest.

Examination of the delayed posttest results shows a similar, although slightly different, picture of the incidental acquisition outcomes. Although the interest group received higher scores than the exam group on all three types of vocabulary assessments, these differences were statistically significant for receptive form knowledge only. There was no statistically significant difference between receptive meaning and productive meaning results for the interest group and the exam group. The delayed posttest scores showed that learners under intrinsically or extrinsically motivated conditions retained more vocabulary knowledge than the baseline group. This finding points out a difference between

incidental learning and incidental retention of vocabulary knowledge. For the retention of some types of vocabulary knowledge gained from the completion of classroom-based tasks, it may not matter if learners are intrinsically or extrinsically motivated. The results from this study further support recently published criticisms of the ILH (e.g., Bao, 2015; Hazrat & Read, 2021; Nassaji & Hu, 2012). For example, Hazrat and Read (2021) recently critically evaluated the ILH, discussing how it might not be able to accurately predict learning from classroom-based vocabulary learning tasks, as such tasks and their motivation to complete them might be affected by the classroom contexts. Instructions and task requirements provided inside classrooms may lead learners to complete the tasks in certain ways, thereby influencing their motivations to complete the task. Likewise, the findings of the current study also support Yanagisawa and Webb's (2021) meta-analysis, finding that it was possible for the ILH to explain certain types of vocabulary knowledge (e.g., receptive form) but not others (e.g., receptive meaning and productive meaning).

### **Motivation and IVA**

Looking back at the empirical studies, the ILH is often adopted by academics when investigating IVA. Some researchers (e.g., Huang et al., 2012; Nassaji & Hu, 2012) have obtained results that are consistent with and support Laufer and Hulstijn's (2001) ILH, showing that tasks with a larger involvement load score result in better IVA. However, in other studies (e.g., Hu & Nassaji, 2016; Kim, 2011), the results only partially support the hypothesis. In Kim's (2011) research, for example, tasks with higher involvement load scores were not always more effective. Kim (2011) concluded that this might have been because the components did not contribute equally to vocabulary development and that strong *evaluation* may be more significant than *need* and *search*. Similarly, Hu and Nassaji (2016) criticized the precision of the ILH, stating that tasks with the same involvement load scores but different component distributions could not equally contribute to the IVA results. Hu and Nassaji (2016) explained their findings by claiming that the ILH assigns only a narrow range of scores to tasks, making it difficult to distinguish between tasks with identical involvement load scores.

Apart from the studies that have included all three ILH components, previous studies were also conducted to discover the individual effect of



each component on learners' IVA (e.g., Yanagisawa & Webb, 2021; Zou, 2017). IVA studies that have used ILH have concentrated more on the cognitive components since they are easier to manipulate. However, among these studies, one crucial issue that has been overlooked is how to conceive and operationalize *need*. With respect to the *need* component, it is worth noting that Yanagisawa and Webb's (2021) meta-analysis discussed only moderate *need*, as no study to date has investigated strong *need*. Yanagisawa and Webb (2021) acknowledged the difficulty of operationalizing the *need* component by explaining that when learners complete teacher-assigned tasks, they might be intrinsically motivated to learn target words. Yanagisawa and Webb (2021) called for studies to expand upon *need* and create clearer criteria for coding this motivational component.

The inconsistency between the ILH prediction and the results in the current study might be explained by the difficulty in measuring the actual learning circumstances. That is, the ILH components are difficult to label dichotomously and consistently at times by different researchers. No *search*, for example, was operationalized by Hulstijn and Laufer (2001) as the provision of marginal glosses. However, contrary to the researchers' expectations, learners continued to search for the meaning of the unknown words encountered in the text and shifted their focus from reading to the marginal glosses. The same might also happen for the *need* component; even if the work was assigned by the teacher, learners may be intrinsically motivated to pay attention to the words. Hence, developing better guidelines and standards for how different circumstances should be coded could improve coding consistency between studies and allow for a more accurate evaluation of diverse conditions (Hazrat & Read, 2021). This idea was also supported by Bao (2015), who attempted to highlight the *need* component of ILH. Bao (2015) contended that assigning a moderate degree to external-driven *need* and a strong degree to self-driven *need* was oversimplifying. There is no doubt that moderate and strong need can be produced due to both internal and external factors, depending on the type of activity. Bao (2015) believed that a better distinction in motivation could conceal the nuanced differences between the levels of *need* when using different vocabulary learning activities.

One of the main contributions of the current study is that it supports the claim that there are several types or levels of motivation, with the most self-driven type of motivation leading to the most successful learning outcomes (Vallerand & Pelletier, 2008). Thus, to explain the

different learning outcomes caused by motivational factors, we must look at the quality of the motivation. The most successful outcomes are likely to be aroused by the self-driven type of motivation (i.e., intrinsic motivation), while the external-driven type of motivation (extrinsic motivation) is more likely to be associated with less successful outcomes. Finally, amotivation has been systematically linked to maladaptive outcomes (Vallerand & Pelletier, 2008). These results have been supported numerous times in different studies (e.g., Vallerand, 2007; Vallerand & Ratelle, 2002).

In summary, motivation is a significant factor that has often been overlooked in IVA studies. Future studies could improve ILH scoring for the *need* component and generate a consistent and detailed standard. Meanwhile, future researchers must provide clearer definitions and examples of different levels of motivation when conducting future IVA studies. More evidence is still needed regarding the relative weight of *need* and how the other components of ILH and related factors (e.g., contextual variables and the nature of tasks) affect motivation.

## **Implications of the Study**

### *Research Implications*

This study has added to the discussion of using the ILH to predict IVA outcomes from activity completion. Those studies that adopt the ILH as a framework should consider how *need* in conjunction with other factors such as task and learning context may affect IVA outcomes. Thus far, these previous discussions have not been able to conclude how the *need* component in the ILH should be better operationalized. Hence, more research should be conducted to determine the relative importance of *need* and how the motivational component should be prioritized in this study. Future studies should define the *need* component in a detailed way, as was done in the current study. Future researchers may consider why the *need* component should be quantified in a more concrete way other than just moderate and strong.

This commentary and the suggestions for further research using the ILH should not be regarded as mere criticisms. The ILH was the first attempt to inspire scholars as well as practitioners to operationalize common concepts referred to in the SLA literature “such as noticing, attention, elaboration, and motivation” (Laufer & Hulstijn, 2001, p. 1).

This was a clear step in the right direction for understanding how different tasks lead to vocabulary acquisition. Nevertheless, the link between ILH and these constructs should be closely scrutinized by researchers. In regard to explaining why one reading condition is more effective than another, learners' self-directed motivation, awareness of unknown words, self-efficacy, and other contextual factors are also likely to affect IVA outcomes. As a result, future studies should consider how there may be varying levels of *need* that are not currently brought to the surface using the current ILH framework.

Last, from a methodological perspective, researchers can consider collecting both online and offline data to understand the relationship between motivation and learners' IVA from reading. If motivation does affect reading behavior, then this can be shown when using eye-tracking methods of data collection (Dirix et al., 2019). These online data, combined with the offline data collected from traditional vocabulary assessments, can be useful for researchers to better understand the effect motivation has on learners' reading behavior and therefore IVA outcomes (Elgort et al., 2018; Godfroid et al., 2018; Pellicer-Sánchez, 2016).

### *Teaching Implications*

The importance of vocabulary in language learning cannot be overstated. The question of how to make vocabulary instruction more engaging has always been debated and investigated (Nation, 2013). Educators must select effective techniques to teach and encourage vocabulary learning (Nation, 2013). To improve the effectiveness of vocabulary learning, educators should first recognize the importance of IVA in increasing learners' vocabulary size (Nation, 2013). Teachers should advise learners to make good use of both intentional and incidental vocabulary learning to expand their vocabulary sizes (Schmitt, 1998). As the current study results showed that there were sizable vocabulary learning outcomes from just reading a single article, educators could consider doing a similar task in their classrooms with their learners. After showing the learners how they were able to incidentally acquire some words just from reading, this can clearly illustrate to the learners the importance of reading for vocabulary growth. Such a simple task can show learners that they can learn vocabulary just from reading texts that they enjoy (Krashen, 1981).

Moreover, educators should provide learners with reading materials and tasks that they are more interested in to further encourage IVA. The findings in the current study revealed that learners who were intrinsically motivated learned more vocabulary than those who were extrinsically motivated or were not told why they were engaging in reading. Thus, when giving classroom-based tasks, teachers may consider allowing the learners some say in selecting the topics they read about.

Finally, one important point needs to be made regarding classroom assignments. The results of the current study actually found that it matters less whether the learners read intrinsically or extrinsically to retain vocabulary knowledge. The students were clearly oriented prior to engaging in reading. Learners must understand the relationship between the completion of a task and any consequences to know how seriously they should take such learning opportunities. An example of this in the current study was the baseline group. Hence, teachers are encouraged to always explain why any task is given to their students and why it is useful.

### *Limitations*

The current study examined the effect of reading purpose on learners' IVA to contribute to theory and practice. This study has filled a research gap by investigating how learners' reading purposes affects IVA. Nevertheless, as with all studies, some limitations were present.

First, although the frequency of target words was strictly controlled in the current study, this frequency of occurrences was set at three. This was necessary to ensure that the reading material was comprehensible and that the lexical load was suitable. The study did show that there was a sizable amount of learning that took place from just reading a single article. If given exposure to more occurrences of the target word or more target words, it is likely that the potential of IVA would have increased (Ellis, 2002; Pulido, 2004; Zahar et al., 2001). Therefore, three exposures to an unknown word may not have been enough to show robust acquisition outcomes.

Furthermore, since the study took place in a lower grade of a junior secondary school, the participants' reading abilities and vocabulary sizes were limited. The length of the reading material had to be controlled. Thus, the participants recruited for the current study read only 576

tokens and met only six unknown target words in the given text. Future researchers may consider conducting similar studies with different participant populations and using text types other than expository, as was done in the current study.

Additionally, the current study did not control the learners' background knowledge. Researchers have argued that learners' background knowledge, such as topic familiarity, has a significant effect on their IVA (Pulido, 2003; Zhao et al., 2016). In this study, although we distributed a questionnaire that included questions about the participants' English learning backgrounds, some influences of topic familiarity on learners' IVA from reading could have occurred. Future studies may also investigate the potential effect that motivation in connection with background knowledge may have on learners' IVA.

Finally, given that the three groups in the current study were comparable and homogeneous, one can assume that the participants in the baseline group were potentially interested in the same topics as the interest group. Hence, researchers are encouraged to think more clearly about how reading conditions can be operationalized in future classroom-based studies. Additionally, some participants in the baseline group might have expected a reading comprehension test or even a vocabulary test during their reading, which might have altered their reading behavior. Future related studies might try to include a post-reading questionnaire or interview to explore the reading behavior of the baseline group, as done by Godfroid et al. (2018) in their eye-tracking study. At the same time, normally a true control group should be recruited to take the tests without any reading. However, since the current study was undertaken in a secondary school context, this was not permitted due to ethical concerns prompted during the ethics review. One important addition to future research, if possible, is the inclusion of a true control group that takes all the immediate and delayed tests without being exposed to any treatment.

### *Suggestions for Future Study*

In summary, the findings gained from the current study cannot totally support Laufer and Hulstijn's (2001) ILH. Although the baseline group and the exam group were not intrinsically motivated to read and were given the same involvement load score, their vocabulary learning outcomes were found to be significantly different. This is because their

motivation to complete the reading was different. To truly understand how reading purpose affects learners' IVA from reading, motivation must be clearly defined in future studies. As an important factor that has often been overlooked in IVA studies, motivation type may have led to different reading behaviors and learning outcomes in those studies. Therefore, future researchers who investigate IVA from reading must provide a clearer definition of motivation, along with examples of how different levels of motivation are investigated in these studies. Meanwhile, there are still uncertainties about the importance of the role that motivation plays in connection with other variables. Thus, IVA researchers should also consider other motivational theories when they are aiming to improve the use of the ILH.

### Acknowledgments

The authors would like to acknowledge the anonymous reviewers' feedback provided on previous versions of this paper.

### Note

- 1 We use the terms "receptive" and "productive" based on Nation's (2013) taxonomy of vocabulary knowledge. We believe that using these terms focuses attention on the aspect of knowledge that was assessed and not the type of assessment instrument created.

### References

- Arseven, S. (2013). The effects of marginal glosses and online dictionary use on incidental receptive and productive vocabulary acquisition through reading. In L. G. Chova, A. L. Martinez, & I. C. Torres (Eds.), *5<sup>th</sup> International Conference on Education and New Learning Technologies* (pp. 6441–6449). AUC Knowledge Fountain.
- Bao, G. (2015). Task type effects on English as a foreign language learners' acquisition of receptive and productive vocabulary knowledge. *System*, 53, 84–95.
- Cobb, T. (2007). Computing the vocabulary demands of L2 reading. *Language Learning & Technology*, 11(3), 38–63.
- Cobb, T. (2008). Commentary: Response to McQuillan and Krashen (2008). *Language Learning & Technology*, 12(1), 109–114.

- Cobb, T. (2015). *Range for Texts v.5.1*. Compleat Lexical Tutor. <https://www.lextutor.ca/cgi-bin/range/texts/>
- Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 104(3), 268–294.
- Dirix, N., Beken, H. V., De Bruyne, E., Brysbaert, M., & Duyck, W. (2019). Reading text when studying in a second language: An eye-tracking study. *Reading Research Quarterly*, 55(3), 371–397.
- 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.
- 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.
- Elgort, I., & Warren, P. (2014). L2 vocabulary learning from reading: Explicit and tacit lexical knowledge and the role of learner and item variables. *Language Learning*, 64(2), 365–414.
- Ellis, N. C. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition*, 24(2), 143–188.
- Godfroid, A., Ahn, J., Choi, I., Ballard, L., Cui, Y., Johnston, S., Lee, S., Sarkar, A., & Yoon, H. J. (2018). Incidental vocabulary learning in a natural reading context: An eye-tracking study. *Bilingualism: Language and Cognition*, 21(3), 563–584.
- Gribbons, B., & Herman, J. (1997). True and quasi-experimental designs. *Practical Assessment, Research, and Evaluation*, 5(14), 1–3.
- Hazrat, M., & Read, J. (2021). Enhancing the involvement load hypothesis as a tool for classroom vocabulary research. *TESOL Quarterly*, 56(1), 387–400.
- Hu, H. C. M., & Nassaji, H. (2016). Effective vocabulary learning tasks: Involvement Load Hypothesis versus Technique Feature Analysis. *System*, 56, 28–39.
- Huang, S., Willson, V., & Eslami, Z. (2012). The effects of task involvement load on L2 incidental vocabulary learning: A meta-analytic study. *Modern Language Journal*, 96(4), 544–557.
- Huckin, T., & Coady, J. (1999). Incidental vocabulary acquisition in a second language. *Studies in Second Language Acquisition*, 21(2), 181–193.
- Hulstijn, J. H., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning*, 51(3), 539–558.
- Jin, Z., & Webb, S. (2020). Incidental vocabulary learning through listening to teacher talk. *Modern Language Journal*, 104(3), 550–566.
- Joe, A. (1998). What effect do text-based tasks promoting generation have on incidental vocabulary acquisition? *Applied Linguistics*, 19(3), 357–377.



- 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), 1–16.
- Kim, Y. (2011). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. In P. Robinson (Ed.), *Task-based language learning* (pp. 100–140). Wiley-Blackwell.
- Krashen, S. D. (1981). The fundamental pedagogical principle in second language teaching. *Studia Linguistica*, 35(1–2), 50–70.
- Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. Routledge.
- Laufer, B. (1989). A factor of difficulty in vocabulary learning: Deceptive transparency. *ALIA Review*, 6(1), 10–20.
- Laufer, B. (1997). The lexical plight in second language reading: Words you don't know, words you think you know, and words you can't guess. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy*. Cambridge University Press.
- Laufer, B. (2005). Focus on form in second language vocabulary learning. *EUROSLA Yearbook*, 5(1), 223–250.
- Laufer, B. (2020). Lexical coverages, inferencing unknown words and reading comprehension: How are they related? *TESOL Quarterly*, 54(4), 1076–1085.
- Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in second language: The construct of task-induced involvement. *Applied Linguistics*, 22(1), 1–26.
- Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, 22(1), 15–30.
- 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.
- Liang, H. (2018). The factors influencing the efficacy of incidental vocabulary acquisition in English reading. In R. Green, Y. Zhang, I. Rumbal, & M. J. Zhang (Eds.), *5<sup>th</sup> International Conference on Education, Language, Art and Inter-cultural Communication (ICELAIC)* (pp. 445–448). Atlantis Press.
- Liu, D. Y., Zheng, W. Q., & Nunan, D. (2012). *Yiwu jiaoyu jiaokeshu: Yingyu* [Compulsory education textbook: English]. People's Education Press.
- Liu, L. L. (2020). *Xinkebiao chuzhong yingyu huati yuedu* [New standard junior high school English reading by topics]. Modern Education Press.
- McLaughlin, B., Rossman, T., & Mcleod, B. (1983). Second language learning: An information-processing perspective. *Language Learning*, 33(2), 135–158.

- McQuillan, J., & Krashen, S. D. (2008). Commentary: Can free reading take you all the way? A response to Cobb (2007). *Language Learning & Technology*, 12(1), 104–108.
- Meara, P. (2002). The rediscovery of vocabulary. *Second Language Research*, 18(4), 393–407.
- Montero Perez, 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.
- Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly*, 20(2), 233–253.
- Narvaez, D., & Van den Broek, P. (1999). The influence of reading purpose on inference generation and comprehension in reading. *Journal of Educational Psychology*, 91(3), 488–496.
- Nassaji, H., & Hu, H. C. M. (2012). The relationship between task-induced involvement load and learning new words from context. *International Review of Applied Linguistics in Language Teaching*, 50(1), 69–86.
- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, 63(1), 59–82.
- Nation, I. S. P. (2013). *Learning vocabulary in another language*. Cambridge University Press.
- Nation, I. S. P., & Webb, S. (2011). *Researching and analyzing vocabulary*. Boston: Heinle.
- Nguyen, C. D., & Boers, F. (2019). The effect of content retelling on vocabulary uptake from a TED talk. *TESOL Quarterly*, 53(1), 5–29.
- 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.
- Peters, E., Hulstijn, J. H., Sercu, L., & Lutjeharms, M. (2009). Learning L2 German vocabulary through reading: The effect of three enhancement techniques compared. *Language Learning*, 59(1), 113–151.
- Plonsky, L., & Oswald, F. L. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning*, 64(4), 878–912.
- Pulido, D. (2003). Modeling the role of second language proficiency and topic familiarity in second language incidental vocabulary acquisition through reading. *Language Learning*, 53(2), 233–284.
- Pulido, D. (2004). The relationship between text comprehension and second language incidental vocabulary acquisition: a matter of topic familiarity? *Language Learning*, 54(3), 469–523.
- Qian, J. (2019). An empirical study on online reading and incidental vocabulary acquisition. In K. Cermakova & J. Holmanova (Eds.), *8<sup>th</sup> Teaching and Education Conference* (pp. 315–322).

- Rahul, D. R., & Ponniah, J. (2020). Understanding the robustness of incidental vocabulary acquisition through reading: Qualitative insights from biolinguistics. *Rupkatha Journal on Interdisciplinary Studies in Humanities*, 12(5), 1–12.
- Reynolds, B. L. (2016a). The effect of L2 learners' vocabulary size on the incidental acquisition of vocabulary through the reading of a novel. In National Chiayi University Department of Foreign Languages (Ed.), *NCYU inquiry in applied linguistics: The 2015 issue—Theme: Linguistics applied across borders* (pp. 1–22). Crane.
- Reynolds, B. L. (2016b). Investigating the effects of target word properties on the incidental acquisition of vocabulary through reading. *TESL-EJ: Teaching English as a Second or Foreign Language*, 20(3), 1–31.
- Reynolds, B. L. (2020). Situated incidental vocabulary acquisition: The effects of in-class and out-of-class novel reading. *Applied Linguistics Review*, 1–29.
- Reynolds, B. L., & Bai, Y. L. (2013). Does the freedom of reader choice affect second language incidental vocabulary acquisition? *British Journal of Educational Technology*, 44(2), E42–E44.
- Reynolds, B. L., & Ding, C. (2021). Effects of word-related factors on first and second language English readers' incidental acquisition of vocabulary through reading of an authentic novel. *English Teaching: Practice and Critique*, 21(2), 171–191.
- Schmidt, R. (1994). Deconstructing consciousness is search of useful definitions for applied linguistics. *ALIA Review*, 11, 4–19.
- Schmitt, N. (1998). Tracking the incidental acquisition of second language vocabulary: A longitudinal study. *Language Learning*, 48(2), 281–317.
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Palgrave Macmillan.
- 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.
- Song, T., & Reynolds, B. L. (2022). The effects of lexical coverage and topic familiarity on the comprehension of L2 expository texts. *TESOL Quarterly*, 56(2), 763–774.
- Swanborn, M. S. L., & de Glopper, K. (2002). Impact of reading purpose on incidental word learning from context. *Language Learning*, 52(1), 95–117.
- Tekmen, E. A. F., & Daloglu, A. (2006). An investigation of incidental vocabulary acquisition in relation to learner proficiency level and word frequency. *Foreign Language Annals*, 39(2), 220–243.
- Teng, F. (2018). Incidental vocabulary acquisition from reading-only and reading-while-listening: A multi-dimensional approach. *Innovation in Language Learning and Teaching*, 12(3), 274–288.

- Vallerand, R. J. (2007). A hierarchical model of intrinsic and extrinsic motivation for sport and physical activity. In M. S. D. Hagger & N. L. D. Chatzisarantis (Eds.), *Self-determination theory in exercise and sport* (pp. 255–279). Human Kinetics.
- Vallerand, R. J., & Pelletier, L. G. (2008). Reflections on self-determination theory. *Canadian Psychology, 49*(3), 257–262.
- Vallerand, R. J., & Ratelle, C. F. (2002). Intrinsic and extrinsic motivation: A hierarchical model. In E. L. Deci & R. M. Ryan (Eds.), *Handbook of self-determination research* (pp. 37–64). University of Rochester Press.
- Van den Broek, P., Lorch, R. F. J., Linderholm, T., & Gustafson, M. (2001). The effects of readers' goals on inference generation and memory for texts. *Memory and Cognition, 29*(8), 1081–1087.
- Van Zeeland, H., & Schmitt, N. (2013). Incidental vocabulary acquisition through L2 listening: A dimensions approach. *System, 41*(3), 609–624.
- Webb, S. (2008). The effects of context on incidental vocabulary acquisition. *Reading in a Foreign Language, 20*(2), 232–245.
- Webb, S., Sasao, Y., & Ballance, O. (2017). The updated vocabulary levels test: Developing and validating two new forms of the VLT. *International Journal of Applied Linguistics, 168*(1), 33–69.
- Wei, R., Hu, Y., & Xiong, J. (2019). Effect size reporting practices in applied linguistics research: A study of one major journal. *SAGE Open, 9*(2), 1–11.
- Wilkins, D. A. (1972). Grammatical, situational and notional syllabuses. In G. Nickel (Ed.), *3<sup>rd</sup> International Congress of Applied Linguistics* (pp. 254–265). Julius Gross Verlag.
- Xie, H. R., Zou, D., Wang, F. L., & Wong, T. L. (2017). A review on recent development of the involvement load hypothesis. In S. Cheung, L. F. Kwok, W. Ma, L. K. Lee, & H. Yang (Eds.), *10<sup>th</sup> International Conference on Blended Learning (ICBL)* (pp. 447–452). Springer.
- Yan, C. M. (2015). “We can’t change much unless the exams change”: Teachers’ dilemmas in the curriculum reform in China. *Improving Schools, 18*(1), 5–19.
- 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. L., & Cao, X. F. (2021). Effects of task involvement load on L2 vocabulary acquisition and their association with language aptitude. *Asia-Pacific Education Researcher, 30*(5), 421–430.
- Zahar, R., Cobb, T., & Spada, N. (2001). Acquiring vocabulary through reading: Effects of frequency and contextual richness. *Canadian Modern Language Review, 57*(4), 541–572.
- Zhang, S., & Duke, N. K. (2008). Strategies for internet reading with different reading purposes: A descriptive study of twelve good internet readers. *Journal of Literacy Research, 40*(1), 128–162.

- Zhao, A., Guo, Y., Biales, C., & Olszewski, A. (2016). Exploring learner factors in second language (L2) incidental vocabulary acquisition through reading. *Reading in a Foreign Language*, 28(2), 224–245.
- 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.

*Ms. Xiaoyan MA received her MPhil degree in Curriculum and Instruction (English Language Education) from the University of Macau, MCSAR. She currently works as a secondary school English teacher in a public secondary school in Zhuhai, China.*

*Dr. Barry Lee REYNOLDS is Associate Professor of English Language Education in the Faculty of Education at the University of Macau. His research interests lie in the areas of Incidental Vocabulary Acquisition, L2 Writing Instruction, CALL, and English Teacher Education. More information on his research and teaching interests can be found at the Applied Linguistics, Language, and Literacy website: <https://alll.fed.um.edu.mo>*

## Appendix A: List of 50 Topics

### Discovering Your Favorite Topics

Student Number: \_\_\_\_\_

This questionnaire is used to help you discover the topics that you are interested in. You are going to choose an English article in a book or magazine to read with a topic you like. Please select the reading topic(s) below that you may have interest in, or you may choose (you may have multiple choices):

本問卷旨在幫助你尋找感興趣的閱讀話題。假如您將在書籍或雜誌中自由選擇你喜歡的英文話題來閱讀，以下列出了各種有趣的話題，請在下面勾選出您感興趣的主題，或者您可能選擇的文章主題（選擇數量不限）：

No.	Reading Topics	I will choose this topic
1	Friends 交朋友	
2	Things around you 身邊的事物	
3	Colors 顏色	
4	The family 家庭與家人	
5	Things in the classroom 教室裏的東西	
6	Things around the house 房子周圍有甚麼	
7	Hobbies 愛好	
8	Food 食物	
9	Shopping 購物	
10	Important dates 重要的日期	
11	School subjects 學習科目	
12	Joining a club 加入一個社團	
13	Daily routines 日常習慣	
14	Transportation 交通	
15	Rules 規則與規矩	
16	Animals in a zoo 動物園裏的動物	
17	The weather 天氣與節氣	
18	The neighborhood 鄰居	
19	Physical appearance 外貌特徵	
20	Trips and vacations 旅行	
21	Weekend activities 週末活動	
22	Holidays 假期	
23	Personal traits 個人性格特點	
24	City introduction 城市介紹	
25	Entertainment 娛樂活動	

No.	Reading Topics	I will choose this topic
26	Life goals 人生目標	
27	Cooking 烹飪	
28	Decision making 決策與決定	
29	Life in the future 未來生活	
30	Invitations 邀請	
31	Health and first aid 健康與急救	
32	Volunteering and charity 志願活動與慈善	
33	Chores and permission 家務	
34	Interpersonal communication 人際溝通	
35	Unforgettable events 難忘的事件	
36	Legends and stories 傳奇故事	
37	Facts about the world 關於世界的事實	
38	Movies, cartoons and music 電影、動畫片與音樂	
39	Literature 文學	
40	Fun places 有趣的地方	
41	Living environment 生活環境	
42	Learning how to learn 學會如何學習	
43	Festivals 節日	
44	Customs 傳統習俗	
45	Mysteries 神秘的事物	
46	Getting around 出行	
47	Inventions 發明創造	
48	How we have changed 生活變化	
49	Protecting the environment 保護環境	
50	Feelings and emotions 感受與情緒	



## **Appendix B: Reading Material**

### **Something You Need to Know about the UK**

If you want to have a wonderful trip to the UK, you cannot miss this helpful article! The following consists of the history of the UK, the weather, the sports in the UK, and the best time to travel to the UK.

#### *About the Country*

England, Scotland, Wales, and Northern Ireland together make the UK. The UK consists of many islands, big or small land areas with water around. Scotland and Wales are the most mountainous parts of the UK. The central government is in London. London is Europe's largest city in which many people enjoy living. London has many beautiful places. For example, Big Ben is one of the most well-known places people like to visit. There are also some lovely streets and rivers that are worth traveling to. Don't miss them!

#### *Weather*

When talking about the weather, people often think about its climate. It is always wet because of its unique location. People in the UK may always feel that their clothes are wet. Also, the temperature is at times changeable. That's why people say that everyone talks about the weather in the UK! In the morning the weather is warm, just like in spring. After an hour, it may start to rain hard, and the weather gets a little colder. In the later afternoon, the sun may come out again, and it will be like summer at this time of the day. The weather is mostly warm in the summer and more enjoyable in autumn. Because of this unique climate, always remember to bring an umbrella with you when traveling outside!

#### *Food*

The UK is also the home of some particular food. Each of the four countries in the UK has its unique food. The most well-known English meal is fish and chips. This food consists of potatoes and fish, which can be seen even in almost every family. Every year, UK shops sell over 250 million fish and chips meals. The UK also has races for this meal, and

many shops join and compete for the best fish and chips shop. People are also enthusiastic about sharing their food around the world!

### *Sports*

Lots of different sports are played in the UK. In the 2016 Olympic Games in Brazil, the UK team competed in almost every sports race and won over 65 competitions. Among them, the most well-known sport in the UK is football. People are enthusiastic about football and the teams they support. When the local people are free, they like to play football with their friends and compete against other teams. If you travel to well-known cities for their football team such as Manchester, the third-largest city in England, don't miss the football game!

### *When is the best time to travel to the UK?*

In the UK, temperatures may not get very warm or cold because of its climate. However, rain and cool weather may often be seen year-round. The best time to come to the UK is usually in the summer months. However, late spring and early autumn can be good as well. Every year, many people come to the UK to travel, and most of these people are Chinese. Local people are always enthusiastic and kind to the people traveling to their country. In short, it is very suitable to go to the UK anytime because you will feel you are in a different place each time the temperature changes!

## Appendix C: Receptive Form and Productive Meaning Test

1. Where is the central government of the UK?

- A. New York      B. London      C. Beijing      D. Paris

2. What is the most memorable meal in the UK?

- A. French fries      B. Soup      C. Fish and chips      D. Hamburger

3. What is the most well-known sport for the local people in the UK?

- A. Basketball      B. Volleyball      C. Table tennis      D. Football

4. According to the article, when is the best time to visit the UK?

- A. Spring      B. Summer      C. Autumn      D. Winter

5. Which word did you see in the reading? If possible, write the Chinese translation of the word in the blank.

- A. comsist  
B. consist  
C. konsist  
D. komsist

Chinese Translation: \_\_\_\_\_

6. Which word did you see in the reading? If possible, write the Chinese translation of the word in the blank.

- A. conpete  
B. conpate  
C. compete  
D. compate

Chinese Translation: \_\_\_\_\_

7. Which word did you see in the reading? If possible, write the Chinese translation of the word in the blank.

- A. unique  
B. uniqua  
C. uneque  
D. unequa

Chinese Translation: \_\_\_\_\_

8. Which word did you see in the reading? If possible, write the Chinese translation of the word in the blank.

- A. enthusiestic  
B. enthusiastique  
C. enthusiestique  
D. enthusiastic

Chinese Translation: \_\_\_\_\_

9. Which word did you see in the reading? If possible, write the Chinese translation of the word in the blank.

- A. climate
- B. climete
- C. clemate
- D. clemete

Chinese Translation: \_\_\_\_\_

10. Which word did you see in the reading? If possible, write the Chinese translation of the word in the blank.

- A. tamperature
- B. temperature
- C. tamparature
- D. temparature

Chinese Translation: \_\_\_\_\_

## Appendix D: Receptive Meaning Test

Please choose the correct meaning for the word among the four options:

1. consist:

- A. 檢舉                      B. 剔除                      C. 組成                      D. 堅持

2. climate:

- A. 日曆                      B. 氣候                      C. 遠足                      D. 證人

3. unique:

- A. 誘人的                      B. 唯一的                      C. 美麗的                      D. 難忘的

4. compete:

- A. 完成                      B. 寫作                      C. 競爭                      D. 烹飪

5. temperature:

- A. 氣溫                      B. 時間                      C. 季節                      D. 節氣

6. enthusiastic:

- A. 冷漠的                      B. 特色的                      C. 中立的                      D. 熱情的

7. article:

- A. 文章                      B. 報紙                      C. 論文                      D. 詩歌

8. northern:

- A. 東部的                      B. 西部的                      C. 南部的                      D. 北部的

9. location:

- A. 簡介                      B. 位置                      C. 機遇                      D. 交通

10. potato:

- A. 番茄                      B. 番薯                      C. 土豆                      D. 黃瓜