English Teaching, Vol. 79, No. 3, Fall 2024, pp. 3-30
DOI: https://doi.org/10.15858/engtea.79.3.202409.3
http://journal.kate.or.kr

Facilitating Incidental Vocabulary Learning: The Effects of Bimodal Presentation and Lexical Elaboration*

Kichan Park**

Park, Kichan. (2024). Facilitating incidental vocabulary learning: The effects of bimodal presentation and lexical elaboration. *English Teaching*, *79*(3), 3-30.

To identify effective methods for boosting incidental vocabulary learning, this study examines the impacts of two tools—bimodal presentation (BP) and lexical elaboration (LE)—on vocabulary acquisition through repeated encounters with target words during meaning-focused reading. In a quiet and comfortable place conducive to full concentration on reading comprehension, 80 L2 learners of English were assigned to one of four treatment groups (BP+LE+, BP+LE-, BP-LE+, and BP-LE-) and individually read three reading passages where target words appeared four times. Surprise posttests assessed meaning recognition and meaning recall. The results of this ecologically valid experiment show significant vocabulary learning through reading, with BP notably enhancing meaning recall. Although the main effect of LE was non-significant, LE was also significantly effective for both meaning recognition and meaning recall if it was noticed by participants. These findings suggest that BP and LE can serve as valuable tools for promoting vocabulary acquisition during meaning-centered reading.

Key words: incidental vocabulary learning, bimodal presentation, lexical elaboration

*This article is based on the author's qualifying paper under the supervision of the late Mike Long for the PhD program in Second Language Acquisition at the University of Maryland, College Park. I thank the committee members, Drs. Mike Long, Kira Gor, Steven Ross, and Nan Jiang, for their constructive feedback. It was supported by the Second Language Acquisition Fund at the University of Maryland.

**Author: Kichan Park, Assistant Professor, Department of English, Republic of Korea Naval Academy; 1 Jungwon-ro, Jinhae-gu, Changwon-si, Gyeongnam, 51698, Korea; Email: park@navy.ac.kr

Received 30 June 2024; Reviewed 17 July 2024; Accepted 10 September 2024

1. INTRODUCTION

Much research has explored how L2 words are acquired, and several researchers have recommended intentional learning, at least for initial form-meaning mappings of new words, while maintaining the importance of reading for acquiring deeper knowledge and automatic processing of vocabulary (e.g., Horst, Cobb, & Meara, 1998; Laufer & Nation, 2012; Webb & Nation, 2013). While learning new vocabulary as a by-product of focusing on the meaning and function of a text may not always be the most efficient approach to vocabulary acquisition, it does not imply that studies on incidental vocabulary learning (IVL) are without value. Today, more and more institutions are adopting communicative language teaching programs such as Task-Based Language Teaching (TBLT), Content and Language Integrated Learning (CLIL) or English Medium Instruction (EMI). In these kinds of programs, learners are expected to acquire L2 forms while engaging in meaning-oriented activities. In these settings, the time available for intentional vocabulary study is sometimes limited (Hulstijn, Hollander, & Greidanus, 1996; Long, 2017). Furthermore, learners who are currently using their L2 for their academic or professional careers have possibly already learned words incidentally while reading or listening (Hulstijn, 2003; Malone, 2018; Vidal, 2011). All these points suggest that it is still worth exploring when L2 vocabulary knowledge can be acquired incidentally, and how this learning can be facilitated. In this light, several studies have been conducted to explore the process of IVL and the effects of external stimuli such as bimodal presentation (BP) of input and lexical elaboration (LE) on IVL, as will be reviewed in the following section. However, previous studies have not always shown consistent results, and still little is known about how these two stimuli work for IVL. Therefore, the present study seeks further evidence supporting the alleged facilitative roles of BP and LE in IVL and investigates how these stimuli operate during learners' meaning-centered reading.

The validity of using the term incidental learning has sometimes been questioned in the literature when it was defined as learning without attending to form (Gass, 1999; Paribakht & Wesche, 1999) or when the lack of methodological devices objectively demonstrating the absence of intention in the learner's mind was pointed out (Bruton, Garcia Lopez, & Esquiliche Mesa, 2011). However, following the tradition of second language acquisition (SLA) research (e.g., Ellis, 1994; Godfroid, Boers, & Housen, 2013; Hulstijn, 2003; Schmidt, 1990), this paper considers that intention is distinguished from attention as well as awareness, and that the incidental learning in experimental research concerns whether the context of an experiment leads the learners to intend to learn a form rather than whether learners have actually attended to (or have intended to learn) a form at the moment of encountering it. In other words, this paper operationalizes IVL as the acquisition of vocabulary occurring during engagement in a meaning-centered task without prior knowledge of subsequent vocabulary

tests.

2. LITERATURE REVIEW

2.1. Bimodal Presentation

Bimodal presentation (BP) means the presentation of an audio-recording of a native speaker reading the text simultaneously as learners read along. Several empirical studies have investigated whether BP facilitates IVL. Although Brown, Waring and Donkaewbua (2008), and Vu and Peters (2022) found no significant effect of BP on IVL, some studies (Malone, 2018; Webb & Chang, 2012, 2022) report greater gains from listening while reading than from self-paced silent reading. Due to these studies (Malone, 2018; Webb & Chang, 2012, 2022) as well as others that indirectly indicate some benefits of BP on second language acquisition (e.g., Webb & Chang, 2015), BP has been receiving lots of attention from researchers as a potential device to facilitate incidental learning (e.g., Long, 2017; Pellicer-Sánchez & Boers, 2019).

However, it needs to be noted that in some of the previous studies that report a significant effect of BP on IVL (Webb & Chang, 2012, 2022), the duration of exposure was not tightly controlled. In Webb and Chang's 2012 study, although the same amount of time was provided to the reading-while-listening and the reading-only groups for "repeated" reading of the text, the researchers could not measure exactly how many times each participant read the text during the given time. In Webb and Chang's subsequent study (2022), both the reading-while-listening and the reading-only groups read the text for about 300 minutes (i.e., in six 50-minute classes), but the audio was just 156 minutes long. Thus, some participants might have had more time to be exposed to target vocabulary after finishing their first round of reading.

In Malone's (2018) study, on the other hand, the duration of exposure was tightly controlled. However, this tight control over exposure duration could have created a more disadvantageous condition for the reading-only groups' IVL. It is reasonable to assume that all but advanced learners would read the text more slowly without audio (Webb & Chang, 2012), so the reading-only groups in Malone's (2018) study might not have had sufficient time for full reading comprehension. In addition, each line of the text suddenly disappeared for all groups when the reading time (the duration of the aural recording of the line for the listening-while-reading groups) was over. Under this environment, participants without BP would not have been able to fully focus on the details of the text due to anxiety. Therefore, a further study in which the amount of exposure is tightly controlled but participants are not rushed to skim the text is required.

Another necessity for further research is provided by our yet limited understanding of the process by which BP promotes IVL. The literature provides several hypotheses on this process. First, listening while reading could contribute to superior comprehension than reading alone because the native speaker's natural prosody could help segment the texts into larger semantic units instead of small incoherent parts (Brown et al., 2008; Pellicer-Sánchez & Boers, 2019; Vidal, 2011; Webb & Chang, 2012, 2015). If so, the mental resources saved from comprehension of the text would allow learners to attend more to unknown words. The segmentation of the text by meaningful units would also be useful for inferring their meanings. Second, the oral rendition of a text could facilitate deeper processing of vocabulary by aiding in the phonological decoding of unfamiliar words (Malone, 2018; Uchihara, Webb, & Yanagisawa, 2019; Vidal, 2011; Webb & Chang, 2012, 2015). Third, listening while reading could prevent learners from ignoring unknown words (Horst et al., 1998; Malone, 2018; Uchihara et al., 2019), a strategy frequently adopted by readers to make sense of the rest of the text without disrupting the flow of meaning caused by unknown words (Gass, 1999; Nagy, Herman, & Anderson, 1985; Paribakht & Wesche, 1999). All these hypotheses are plausible, but what is lacking is empirical evidence. The hypothesis that BP facilitates IVL by making reading comprehension easier is tested in the present study by using a post-treatment survey which asks participants about their perception of BP.

2.2. Lexical Elaboration

LE refers to providing an overt semantic clue for words potentially unfamiliar to learners. Enhanced comprehension of texts achieved through LE provides a solid basis to expect better learning. The literature proposes that when the text is readily comprehensible, learners are able to focus their attention on unknown words (Webb & Nation, 2013) while not becoming frustrated by the difficulty of comprehension (Paribakht & Wesche, 1999). In addition, LE is suspected to be particularly useful when learners are faced with texts where the information necessary for guessing the meanings of unknown words is insufficient, or when learners' comprehension of the context is limited by their low proficiency (Kim, 2006). Moreover, even for a case where there are clues hinting at the meanings of unknown words in the text before modification, increasing semantic richness is predicted to facilitate the rate of vocabulary acquisition (Kim, 2006; Rott, 1999; Toya, 1993). Supporting this, Webb (2008) demonstrated that when learners were exposed to target words for the same number of times during meaning-focused reading, the amount of information about unfamiliar words presented in treatment materials significantly accounted for the level of learners' newly obtained knowledge of word meaning.

Several studies have been conducted in attempts to directly investigate the effect of LE on the acquisition of pre-designated target words. These studies report different effects of

different types of LE. Chaudron (1982) stated that elaboration is explicit when there is an explicit indication of the relationship between the first lexical item mentioned and the following word(s) or phrase(s) (e.g., '… sewage, *which means* the polluted water, …'), whereas elaboration is implicit when there is no explicit indication of this relationship (e.g., '… sewage, *or* the polluted water, …'). The studies that adopted this distinction commonly show that *explicit elaboration* was more effective than *implicit elaboration* and *no elaboration* for learning word meaning (Kim, 2006; Toya, 1993; Vidal, 2011). However, the effect sizes of implicit LE have always been too small to reach statistical significance (Godfroid et al., 2013; Kim, 2006; Toya, 1993).

Although the research consistently reports the findings described as mentioned before, it is worth noting that in most studies the target words occurred only once (Kim, 2006) or twice (Toya, 1993). Since learners sometimes read texts where they encounter unfamiliar words more than just a few times, the effect of LE on words occurring multiple times is worth exploring. In addition, taking into consideration that learners incidentally gain knowledge of words in small increments (Horst et al., 1998; Rott, 1999; Webb, 2007), the operation of LE under the existence of other semantic clues triggers curiosity. Some researchers (Hulstijn et al., 1996; Paribakht & Wesche, 1999) claimed that the frequency effect is stronger only for words whose meanings are already clearly learned and not for others whose meanings remain obscure because only in the former case would repeated occurrences of target words reinforce the form-meaning connection in the reader's mental lexicon. In such a case, the effect of explicit elaboration would be greater when target words are encountered multiple times. However, if the widely reported frequency effect played a role such that learners could guess the word meaning even without explicitly added semantic clues, the large difference in initial gains from explicitly elaborated input and unmodified input would decrease as the frequency of occurrence of the target word increases.

The other area that needs investigation is the interaction between LE and BP. Malone (2018), and Webb and Chang (2012) showed that BP facilitates the form-meaning connection for unknown words even while reading unelaborated narratives in which the information necessary for inferring word meaning is far less than in expository prose (Nagy et al., 1985). This suggests that BP may have potential for aiding vocabulary learning even in the absence of LE. On the other hand, the literature also proposes that BP could prevent learners from skipping target words that may more frequently occur when contextual information is redundant, as it is in elaborated text (Godfroid et al., 2013; Hulstijn et al., 1996). This supports the view that BP combined with LE could be more useful for learning unknown words. These two different assumptions are both motivated by previous research findings, but to better understand the exact nature of the interaction between LE and BP, more research is necessary.

3. METHOD

3.1. Research Questions

The literature review above has identified several gaps in the literature. Regarding these gaps, this study is motivated by the following research questions:

1) Can BP or LE facilitate IVL?
2) Does the combination of BP and LE further boost IVL?
3) What are the mechanisms of BP and LE in IVL?

3.2. Participants

100 participants were recruited from 13 universities in South Korea, and 20 were randomly assigned to each of the following five conditions: (a) with both BP and LE (+BP+LE), (b) with BP but without LE (+BP-LE), (c) without BP but with LE (-BP+LE), (d) neither BP nor LE (-BP-LE) and (e) the control condition in which participants did not read any treatment text. Participants were recruited by an online flyer posted on university bulletin boards where part-time jobs were advertised. The participants were informed that this study was about the relationship between reading comprehension skills and language aptitude.

To ensure that participants had sufficient levels of proficiency required for comprehension of treatment texts, but not potential preexisting vocabulary knowledge of target words, the recruitment announcement stipulated that participants should have obtained either 550 ~ 800 on the Test of English for International Communication (TOEIC)[1] or 70 ~ 90 on the English test of the Korean College Scholastic Ability Test (CSAT) which was taken within two years of their participation in the current study. At the same time, and for the same reason, volunteers with the experience of living in an English-speaking country longer than one year were not allowed to take part in the present study.

Out of recruited 100 participants, 68 were female. The mean age of the participants was 21.2 years (SD = 2.5), and they began learning English at an average age of 7.5 years (SD = 1.9). Detailed biographical information for participants in each group is provided in Table S1 of the Supplementary Materials.[2] Welch's tests confirmed that the differences in age and

[1] TOEIC is a widely used, paper-based standardized English proficiency test. According to the Educational Testing Service (ETS), the developer of the TOEIC, TOEIC scores higher than 550 are comparable to the B1 level (Independent user - Threshold) on the Common European Framework of Reference for Languages (CEFR), and TOEIC scores higher than 785 are comparable to the B2 level (Independent user - Vantage) on the CEFR.

[2] For space purposes, Appendix S1 and Tables S1–S16 are given as supplementary materials available

age of acquisition across the groups were not significant (*p*s > .55).

3.3. Treatment Materials

12 target words were chosen from low frequency English words in consideration of their potential relevance to the topics of three reading passages. To minimize the influence of different parts of speech, only nouns were selected as target words (see Table S2 for the list of target words).

Treatment materials were developed based on three short passages from a booklet 'Speed Readings for ESL Learners: 3000 BNC (Millett, 2017)' posted on Paul Nation's website. However, to accommodate 12 target words appearing four times in each passage, the original passages were modified to a great extent. All target words appeared four times in only one passage, and not in the others. The decision to make target words appear four times was made to examine the role of LE in a text where target words appear more than twice. Four, which is not unrealistically too much greater than two, was thought to be an appropriate exposure frequency to investigate this issue. Part of the modification of the text attempted to evenly distribute the target words throughout the passage without affecting its coherence.

An examination of the passages determined that 95.06% of the final version of the unmodified texts (the texts without LE) comprised the most frequent 3,000 lemmas identified by COCA or 1,800 target lemmas included in the Korean National Curriculum for primary and secondary English education (Korean Ministry of Education, 2015), as well as semantically transparent derivations of those lemmas (e.g., 'greatly'), pronouns, loan words (e.g., 'supermarket') and the titles of the passages. This level of text coverage, which exceeded the threshold of 95%, led to the assumption that most participants would be able to understand the general meaning of the reading passages (Laufer & Nation, 2012).

After the baseline reading texts (the texts without LE) were created, the texts with explicit LE were developed. Since the literature suggests that the implicit LE is not effective for IVL, only explicit LE was provided using synonyms or definitions ('X, meaning Y'; 'X, which means Y'; 'X, that is, Y'; 'X, in other words, Y'; 'X, which is to say, Y'; 'X, put another way, Y'; 'X, by which is meant Y') (e.g., ... *sanatoriums, that is, institutions that offer medical treatment and rest* ...). Although target words appeared four times, LE was provided only once, just for their first appearance. As shown in Table S3, the sizes of the differences in length, complexity, readability and text coverage by frequent words between the two types of texts, were all very small.

For reading with BP, the passages were audio-recorded by a female native English speaker.

for online access through the IRIS online repository (retrieved from https://www.iris-database.org/). All the materials such as reading passages and measurements are also found on the IRIS website.

The texts were read at the average rate of 125 words per minute (wpm) (range: 110 ~ 145 wpm). This was slower than L1 English speakers' "rapid" speech rate (160 ~ 170 wpm, Pawlas, Ramig, & Countryman, 1996, p. 85, as cited in Malone, 2018) but faster than the rate native speakers talk to low-level non-native speakers (100 wpm, Chaudron, 1988, p. 64–69) (c.f., the reading speed of the audio for the listening-while-reading was 93 wpm in Brown et al., 2008; 120 ~ 140 wpm in Malone, 2018; 120 wpm in Webb & Chang, 2022). However, it was predicted that participants' silent reading speed would be slower than the natural reading speed of the native-speaking recorder. Therefore, because the same time should be given to the BP+ and the BP- groups, short pauses (1~2 seconds) were inserted between sentences in the audio to ensure enough reading time for the BP- groups. These pauses did not greatly impede the natural flow of the audio recording. Providing sufficient reading time to all participants was indispensable because otherwise all findings could be just the function of the levels of understanding of the texts within a given time. In a post-treatment survey, out of 80 participants in the treatment groups, only one from the BP-LE+ group (1%) replied that the reading time had often not been enough. The other participants responded that the reading time had sometimes been not enough (21%), rarely not enough (44%) or always enough (34%). Perceived sufficiency of reading time was not significantly different across the four treatment groups, $F(3, 42.19) = 1.22$, $p < .32$ (See Table S4 for detailed descriptive statistics).

During the treatment, parts of the text were presented on timed PowerPoint slides (*M*: 114 words per slide, range: 65 ~ 151 words per slide), and participants could not go back to a previous part that had already been read. The audio for the BE+ groups started as soon as a part of the text was presented, and it ended when the audio recorder completed reading that part. When the audio ended, the slide was automatically turned over.

A timer was presented at the top-right corner of the screen for the BP- groups to help participants monitor the time remaining for reading the part of the text presented on each slide. This manipulation was necessary for a valid study. Without the timer, the BP- groups were not able to carefully read the text and would hurry to read while skipping some words due to their anxiety because they did not know when the reading time would suddenly be over.

15 comprehension questions were given in the middle of the reading passages to draw participants' attention to meaning. Each comprehension question was presented on a PPT slide following a part of the text, so participants could not read the text when they responded to the comprehension questions. Participants' high accuracy rates on the reading comprehension questions (*M*: 92%, range: 66 ~ 100%) indicated that most participants understood the key contents of the reading passages. An after-treatment survey also showed that most participants felt that the texts were not difficult. Group comparisons for reading comprehension scores and survey responses suggested that the four groups were not

significantly different from one another regarding the levels of understanding of the texts, perceived difficulty of the texts, and familiarity with the topics of the three reading passages (*p*s > .11) (see Table S4 for descriptive statistics).

3.4. Instrumentation

3.4.1. Vocabulary posttests

Three types of tasks were used to measure different aspects of new vocabulary knowledge acquired through reading the treatment passages.

First, since recognition of a new form is the starting point in the long and complex process of vocabulary acquisition (e.g., Webb, 2007), a form recognition task was employed. During this test, participants were asked to choose one option out of two for each target word (e.g., (a) 'yes' if they had seen the word in the treatment texts; (b) 'no' if they had not seen the word) (see Appendix for examples of the three types of vocabulary posttests used for the study). In addition to the 12 target words, another 12 low frequency words that had not appeared in the reading passages were also used. When scoring, one point was given for each correct response. The control group did not carry out this task because this group had not read the treatment passages and could not answer the questions. The internal consistency of the task performed by the four treatment groups ($k = 80$) was low (Cronbach's alpha = .53) because this task was too easy for all participants after encountering a small number of target words ($n = 12$) four times during reading (see Elgort, Brysbaert, Stevens, & Van Assche, 2018; Pellicer-Sánchez, 2016 for the evidence of quick word form learning). Due to its low internal consistency, we do not report the results of this task in the main text.[3]

Second, a meaning recognition task was used to detect initial, partial knowledge of word meaning obtained through IVL. In this task, participants were asked to choose one sentence out of four options with the correct use of a target word. Regarding the nature of this test, it was judged possible that participants could choose the correct answer not based on their vocabulary knowledge but on their familiarity with the contexts of the four options. For example, a participant could choose the correct answer for *ailment* just because they read a text about disease ('Malaria'), and the context of one option retaining this word was also about disease. Therefore, to minimize the unwanted influence of 'familiarity with context,' the contexts of all distracters were designed to be somewhat related to the treatment texts. This was achieved by creating correct sentences with other words that had appeared in the texts as distracters and then replacing them with the tested word (e.g., for a distracter for the

[3] In analysis for form recognition, a significant effect of BP on form recognition was found ($z = 2.20$, $p = .028$), and neither other main effects nor interactions between them were significant (see Table S5 for descriptive statistics of the results and Table S6 for specific model outputs).

word "germination": *He was injured in a Saturday-night fracas* [another target word that appeared in a treatment passage] *downtown → He was injured in a Saturday-night germination downtown*). If there were only target words in this multiple-choice task, some participants could randomly choose an answer without paying any attention to the question because they did not know any of the asked words. To prevent this situation, in addition to the 12 target words, another 12 words that also appeared in the reading passages but were predicted to be known to most participants were also included and randomly distributed in the task by Qualtrics. When scoring, one point was given for each correct response. According to Cronbach's alpha, the internal consistency of the task with 24 items (k = 100) was .90.

Third, to measure the extent of the development of semantic knowledge of novel words during meaning-oriented reading, an L2→L1 translation task was also employed as a measure of meaning recall. In this task, target words were presented with a blank, and participants were asked to type the L1 translation equivalent of each target word in the blank. When scoring, a score of 1 was given if a participant's answer was completely correct (e.g., the Korean translation of 'hospital' or 'a medical facility' for the target word 'infirmary'), and partial credit was not provided for responses with meanings close to the target words. Two raters scored responses separately. The percentage agreement between the two raters was 98.3% (Cohen's Kappa was .96), and all disagreements were resolved through discussions between the two raters. The internal consistency of the test was then examined. Cronbach's alpha coefficient for this test with 12 items (k = 100) was .73.

3.4.2. Retrospective vocabulary pretest

To investigate vocabulary learning during meaning-centered reading without any expectations about a subsequent vocabulary test, we decided not to use a pretest for target words. This was because a pretest could give participants the impression that the current study has something to do with vocabulary and draw their attention to unfamiliar words in the reading passages. At the same time, use of pseudowords was also avoided for ecological validity. Instead, real low-frequency words were used as target words. For these reasons, to check the treatment groups' prior knowledge of target words, following Kim (2006) who used real words as the learning targets, questions on participants' preexisting knowledge of the target words were given in a survey that was administered right after the vocabulary posttests. Specifically speaking, the treatment groups were presented with a list of 12 target words and were asked to indicate whether they had known their meanings.

The treatment groups' survey responses were then compared with their answers to the meaning recall task. If these answers were at least partially correct, the words which were indicated to have been known were treated as words with prior knowledge. This scoring

criterion is different from the one employed for the meaning recall task that was used as the posttest. This relatively looser criterion for the pretest was employed since one of the purposes of the present study is to investigate the effects of BP and LE on acquisition of incomplete initial semantic knowledge of target words. If only target words with completely correct L1 translations had been treated as known words, learning gains measured by the meaning recognition task could have been inflated. This is because participants' meaning recognition performance could be affected not only by participants' newly acquired knowledge but also by their already existing incomplete knowledge which was not identified by a pretest with the strict criterion.

The control group's prior knowledge was measured with the meaning recall task. In other words, this group performed the same meaning recall task twice both as a pretest (before meaning recognition) and posttest (after meaning recognition). When scoring the control group's responses in the pretest, target words for which at least a partially correct meaning in L1 was provided were treated as known words.

3.4.3. Questionnaire surveys

To ensure that the results from the four treatment groups were not influenced by other confounds and to understand the mechanisms of the BP and LE in IVL, two brief surveys were used.

Just after finishing their reading, participants were provided with the first questionnaire asking their perception of the reading passages. The survey asked (a) how difficult the passages were, (b) whether the reading time was sufficient and (c) how familiar they were with the topics of the reading passages. The BP+ groups were also asked about (d) whether they felt the BP was helpful for their reading comprehension.

The second questionnaire was provided after the three vocabulary posttests. This survey included questions for the retrospective vocabulary pretest which was described in the previous section. The LE+ groups were also asked about whether they had noticed LE for each target word when reading the treatment texts. In these questions, example sentences with LE that had appeared in the reading passages were also given to help participants' recollection of memory.

3.4.4. Proficiency measure

Although TOEIC and the English test of CSAT scores were used for screening of participants during recruitment, because participants' proficiency was not measured by the same instrument, a cloze test (Brown, 1980) was administered to accurately measure their proficiency at the moment of the experiment. When scoring, one point was given for each

correct response. According to Cronbach's alpha, the internal consistency of the task with 50 items (k = 100) was .81.

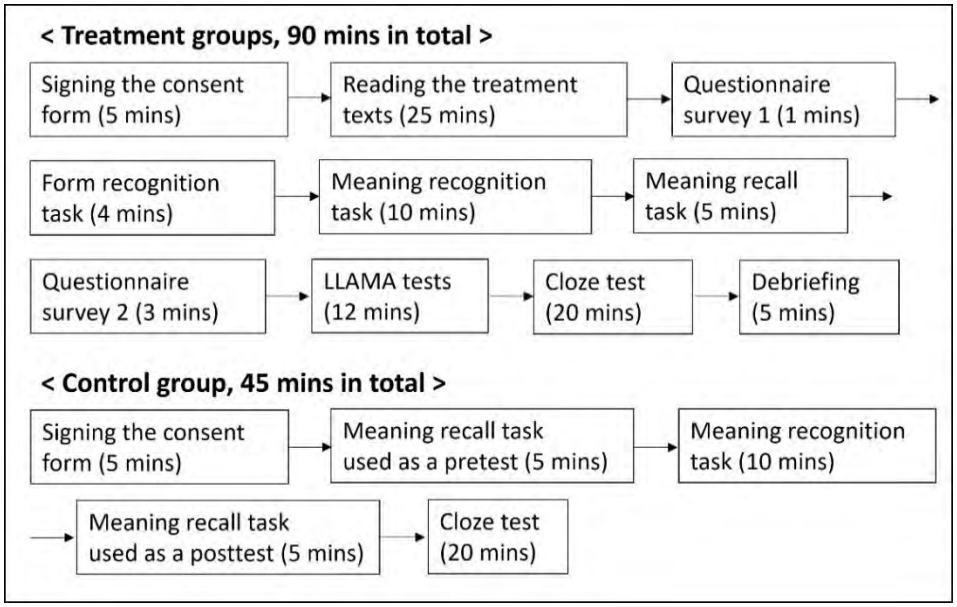
3.5. Procedure

The overall procedure of this project is presented in Figure 1.

First, participants indicated available times for participation in the study on a quick survey that followed the consent form. The researcher then met them one by one on Zoom and implemented the experiment. During the experiment, participants were asked to turn on their camera to ensure that they were located in a quiet place without any distractors around them as required by the flyer and consent form. They were also asked to share their screen with the researcher on Zoom during the three vocabulary posttests and cloze test to prevent cheating.

FIGURE 1

The Overall Procedure and Average Time Spent for Each Task



Participants in the treatment groups first read treatment texts under one of the four experimental conditions: BP+LE+, BP+LE-, BP-LE+ or BP-LE-. Before reading the texts, participants were told that the researcher intended to test their ability to accurately read passages written in English within a limited time, and that comprehension questions would be given in the middle of the passages. For the treatment, a paragraph-length part of each

text that the researcher shared using Zoom was presented to participants for a limited time (i.e., the duration of the audio recording of that part of the text). Simple reading comprehension questions were provided at five points along each passage. These questions required participants to answer by saying "yes" or "no" on Zoom. Instantaneous oral feedback (i.e., correct / wrong) was also given by the researcher to encourage them to keep attending to the content of the passages. So as not to draw overt attention to the target words, none of the comprehension questions required understanding them or was directly related to them. Three treatment passages were preceded by three paragraphs from another reading passage and three comprehension questions for practice. The BP+ groups were instructed to listen to the audio while reading the text. Just after finishing their reading, the first questionnaire was provided. After that, the three types of vocabulary posttests and the second survey were administered.[4] LLAMA D and B aptitude subtests then followed, and participants also performed the cloze test. The results from the LLAMA subtests are not reported in this article because they are beyond the scope of the present study. The time spent completing the experiment varied from 75 to 100 minutes across individuals in the treatment groups.

Participants in the control group first performed the meaning recall task as a pretest. They then carried out this task once again after completing the meaning recognition task. They also took the cloze test. The primary purpose of the control group was to gauge the impact of the meaning recognition task on the treatment groups' meaning recall performance. It took about 45 minutes for participants in the control group to complete the experiment.

3.6. Analysis

To account for variances resulting from random population sampling and varying levels of difficulty in measurement items, logistic mixed-effects models were employed for statistical analyses. These models are advantageous because they can accommodate different types of random effect structures as well as the fixed effects of independent variables (Baayen, Davidson, & Bates, 2008; Cunnings, 2012). The categorical variables, BP and LE, were coded as 0.5 (for BP+ or LE+) and -0.5 (for BP- or LE-) using contrast coding; the continuous variable Proficiency was Z-transformed and will be called ClozeZ hereafter. For preexisting vocabulary knowledge, which will be called PreKnow, treatment coding ('0' for

[4] Participants carried out the meaning recognition task before performing the meaning recall task because this study attempted to measure participants' initial yet unstable semantic knowledge of unfamiliar words before they drew conclusions about the definitions of these words. In SLA research, implicit knowledge has sometimes been measured before explicit knowledge (e.g., Pellicer-Sánchez, 2016; Suzuki & DeKeyser, 2017). In addition, the control group showed that the probability of vocabulary acquisition only through the meaning recognition task was minimal (see Table S10 and Figure 2).

unknown and '1' for known words) was used.

All models were implemented through the "lme4" software package in R (Bates, Mächler, Bolker, & Walker, 2015). Barr, Levy, Scheepers and Tily (2013) suggested that random slopes of all fixed effects and their interactions should be included in mixed-effects models for confirmatory studies. However, in the present study, models with the most complex random-effect structures either did not converge or produced outputs in which the intercept highly correlated with a random slope (r = -1 or 1). Since such a high correlation signifies model overparameterization (Baayen et al., 2008; Bates, Kliegl, Vasishth, & Baayen, 2015), these maximal models were simplified. Specifically, the model simplification process started from dropping the highest-order interaction term out of the most complex random effects structure. The other random slopes were then dropped one by one from the smallest variance component until a sign of over-specified random effects was no longer observed (Bates, Kliegl, Vasishth, & Baayen, 2015).

4. RESULTS

4.1. Proficiency

The five groups' (i.e., BP+LE+, BP+LE-, BP-LE+, BP-LE- and control groups) cloze test scores were first compared. The descriptive statistics of the cloze test scores are presented in Table S7. The result of a one-way ANOVA revealed that the five groups' scores were not significantly different, $F(4, 95) = 1.00$, $p = .42$, $\eta 2 = 040$.

4.2. Preexisting Vocabulary Knowledge

To verify participants' self-assessed, retrospective, preexisting vocabulary knowledge, the treatment groups' responses to survey questions that asked their prior knowledge of each target word were compared with their meaning recall responses. Out of 280 "I had known the word" responses in the survey, participants' prior knowledge was evident in only 245 responses (88.5%) for which at least partially correct L1 meaning recall answers were provided. For the other 35 responses, participants provided incorrect L1 translations, and many of the incorrect responses (23 responses out of 35) appeared to be due to form confusion (e.g., "infirmity" for the target word "infirmary") or dependence on deceptively transparent spellings (e.g., "something about germ" for the target word "germination") (Laufer, 1997). Those 35 responses were not treated as known words below.

Another finding worth noting is the different sizes of the gaps between self-reported and verified preexisting vocabulary knowledge across the treatment groups. The gaps seemed

much bigger in the BP- groups (11 words in BP-LE+ group and 19 words in BP-LE- group) than in the BP+ groups (4 words in BP+LE+ group and 2 words in BP+LE- group) as illustrated in Table 1. A *Post-hoc* statistical test was performed using a logistic mixed-effects model with the 280 items that participants thought they had known before the experiment. The dependent variable was "match" between self-reported and verified prior knowledge (coded as '0' if they did not match and '1' if they matched), and the fixed effects were BP and LE. To account for the effect of proficiency, ClozeZ was also fitted as a covariate. Participant and Item were random effects. The results (see Table S8 for specific model output) show that BP was a significant predictor of the match ($z = 3.18$, $p = .001$), but none of the other predictors or interactions between them was. Put differently, the BP+ groups were significantly less likely to wrongly perceive target words as other words than the BP- groups, but there was no significant difference between the LE+ and LE- groups.

TABLE 1

The Gaps between the Perceived and Verified Preexisting Vocabulary Knowledge

Group	No. of Perceived Known Words	No. of Verified Known Words	Gap
BP+LE+	74	70	4
BP+LE-	71	69	2
BP-LE+	70	59	11
BP-LE-	64	46	18

Note that the survey asking participants' preexisting vocabulary knowledge was administered right after they had read the text. Therefore, the gaps suggest that even after encountering target words four times in the reading passages, some participants still did not realize the incompatibility of wrongly retrieved word meanings with the general meaning of the reading passages. However, when BP was provided, this phenomenon was significantly less likely to occur. This suggests that without BP, the text was processed more shallowly such that correction of wrong perception of target words occurred less frequently.

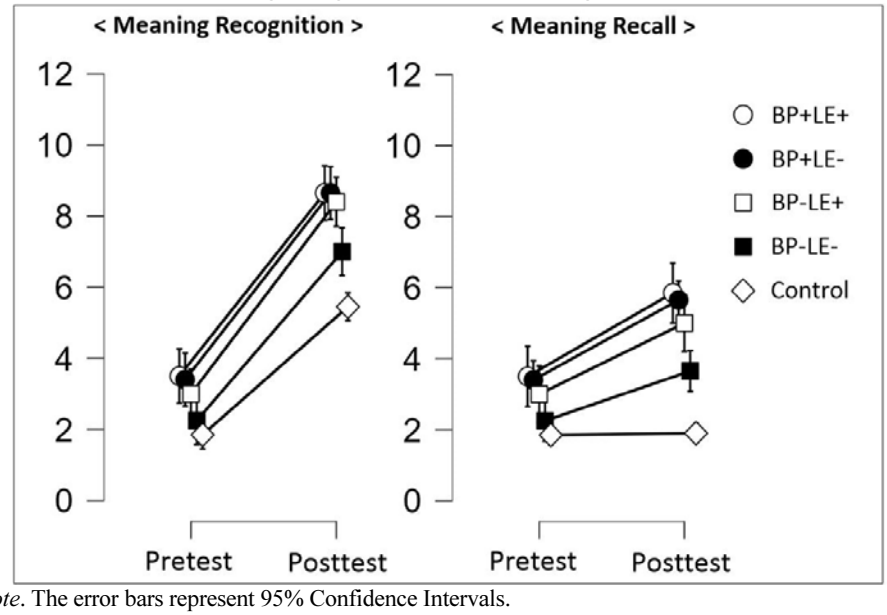
After having verified participants' responses on the retrospective vocabulary pretest, the five groups' preexisting vocabulary knowledge of target words was compared. The results of a Welch test showed that the five groups' prior knowledge was significantly different, $F(4, 46.84) = 5.12$, $p = .002$ (see Figure 2 and Table S9 or S10 for descriptive statistics). In all the following analyses, the difference in prior knowledge between the groups was statistically controlled for by adding PreKnow as a covariate to statistical models. Treating PreKnow as a covariate in models with the posttest score as the dependent variable was also expected to be helpful for minimizing the influence of *regression to the mean* compared to using the subtraction of the pretest score from the posttest score as the dependent variable in statistical models (Bonate, 2000).

4.3. Effect of Reading

The descriptive statistics of the results of the pretest and posttests (meaning recognition and meaning recall tasks) as well as estimated learning gains (i.e., posttest score - pretest score) are presented in Figure 2, Table S9 and Table S10.

FIGURE 2

The Average Numbers of Correct Responses on the Pretest, Meaning Recognition Task and Meaning Recall Task



Note. The error bars represent 95% Confidence Intervals.

The gains of the control group indicate the average numbers of posttest responses that could be correctly answered by chance. To see whether vocabulary learning had occurred through reading, the estimated gains between the control and treatment groups were compared. A mixed-effect logistic regression model was fitted with the correctness of participants' responses on each posttest item as the dependent variable. Reading (coded as '-0.5' for the control group and '0.5-' for the treatment groups) was the independent variable, and preexisting vocabulary knowledge (PreKnow) and proficiency (ClozeZ) were covariates. Participant and Item were random effects.

The results revealed that, when holding PreKnow and ClozeZ constant, the treatment groups showed significantly higher scores than the control group both in meaning recognition ($z = 4.16$, $p < .001$) and meaning recall ($z = 4.69$, $p < .001$) (see Table S11 for specific model outputs). The control and the BP-LE- groups' posttest performance was also

compared by replacing the variable Reading with the dummy-coded variable Group in the same model. The results showed that when holding PreKnow and ClozeZ constant, the BE-LE- group performed significantly better both in meaning recognition (z = 2.18, p = .029) and meaning recall (z = 3.55, p < .001) than the control group (see Table S12 for specific model outputs). This suggests that acquisition of semantic knowledge occurred through reading even without BP or LE.

4.4. Effects of BP and LE

To examine whether BP or LE made a difference in IVL (Research Question 1) and whether the combination of BP and LE further facilitated IVL (Research Question 2), the four treatment groups' performances on the posttests were compared with each other after excluding the data from the control group.

The results of the meaning recognition and meaning recall tasks were separately analyzed. The correctness of participants' responses on each posttest was inserted as the dependent variable in models, and BP and LE were entered as the fixed effects. PreKnow and ClozeZ were covariates, and Participant and Item were random effects.

A summary of the statistical analyses is presented in Table 2 (see Table S13 for specific model output). When controlling for preexisting vocabulary knowledge, ClozeZ was a significant predictor both for meaning recognition (z = 3.18, p = .001) and meaning recall (z = 4.34, p < .001). This finding suggests that a good understanding of the text is crucial for IVL. In terms of the effects of BP and LE, neither significantly improved participants' meaning recognition. However, in meaning recall, although the main effect of LE was not significant, a significant main effect of BP (z = 2.42, p = .016) was observed.

TABLE 2

Significant Predictors of Meaning Recognition and Meaning Recall

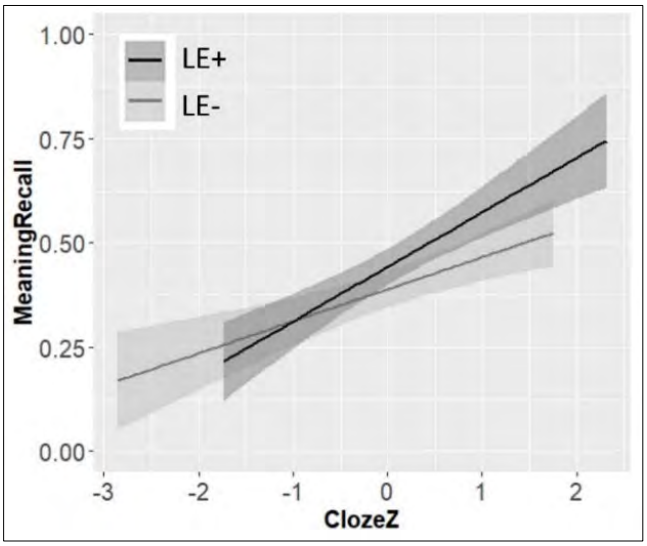
Posttest	Predictors
Meaning recognition	PreKnow (z = 3.41, p < .001), ClozeZ (z = 3.14, p = .002)
Meaning recall	PreKnow (z = 10.97, p < .001), ClozeZ (z = 4.34, p < .001), BP (z = 2.42, p = .016), LE:ClozeZ (z = 2.20, p = .028)

No further facilitative effect was observed for the treatment combining BP and LE as the interaction between them was non-significant. However, a significant interaction between LE and ClozeZ was found (z = 2.20, p = .028) in meaning recall. A plot was fitted to understand the nature of this interaction, and this plot (Figure 3) shows that the more proficient participants were, the more effective LE was. A suspected reason for this finding was that only proficient participants had the knowledge necessary to notice LE (e.g., the knowledge of the grammar rules of relative clauses, or formulaic sequences such as 'in other

words' or 'put another way'). This possibility will be inspected below.

FIGURE 3

The Interaction between LE and ClozeZ in Meaning Recall



4.5. Relationship between Perception of BP and Vocabulary Learning

To confirm the hypothesis that BP promotes IVL by making reading comprehension easier (Research Question 3), right after the treatment session, the BP+ groups (n = 40) were given a survey question asking how helpful they felt the audio support was for their reading comprehension. As shown in Table S14, the results revealed that a substantial number of participants felt that the audio support was very (5%) or somewhat interfering (27.5%). To examine if these participants' perception of BP was determined by their levels of proficiency, a Spearman correlation analysis was performed. The results of this nonparametric correlation analysis showed that participants' helpfulness ratings on BP were not systematically related to their proficiency (R_s[40] = -.09, p = .58).

Next, we investigated whether participants who felt the audio was helpful for their reading comprehension acquired greater vocabulary knowledge than those who felt it was interfering. As shown in Table S14, 13 participants who responded that the audio was very or somewhat interfering showed numerically higher proportions of correct responses to previously unknown words than the other 18 participants who responded that the audio was a little or very helpful in both meaning recognition (70% vs. 64%) and meaning recall (36% vs. 25%). This finding is not consistent with the claim that BP is helpful for vocabulary acquisition by making reading comprehension more comfortable (e.g., Brown et al., 2008; Webb & Chang, 2012).

4.6. Effect of Noticed LE

To investigate the operation of LE for IVL (Research Question 3), post-treatment survey questions asked participants whether they had noticed LE for each target word while reading. Although noticing is sometimes operationalized by the amount of attention (e.g., Godfroid et al., 2013), in this study it was operationalized as participants' awareness of the existence of an overt semantic clue (LE) for a target word in the text (Long, 2017; Schmidt, 1990). Out of 480 cases of LE for individual target words (40 participants × 12 target words), 384 cases (80%) were reported to be noticed by participants. Put another way, although only explicit LE was provided, not every LE was noticed. To determine which factors contributed to noticing LE, a logistic mixed effects model was fitted using only data from the LE+ groups (480 responses from 40 participants). In this analysis, noticing (whether or not they had noticed LE) was the dependent variable. PreKnow, ClozeZ, and BP were fixed effects, and Participant and Item were random effects.

The results (see Table S15 for specific model output) showed that preexisting vocabulary knowledge ($z = 2.47$, $p = .014$) and BP ($z = 2.53$, $p = .011$) were significant predictors of noticing LE. However, ClozeZ did not significantly predict the probability of noticing LE. This means that no evidence was found for the assumption that the significant interaction between LE and ClozeZ in meaning recall that was observed in an earlier analysis (see Figure 3) was because low-proficiency participants did not have the knowledge required to notice LE (e.g., knowledge of relative clause or formulaic phrases).

To understand the role of noticed LE in word meaning acquisition, the variable 'LE' was replaced with 'Noticed LE' in the models that had been used for earlier analyses (from which the outputs presented in Table 2 were produced). For this variable, both unnoticed LE and the absence of LE were coded as '-0.5', and only noticed LE was coded as '+0.5'. Except for this variable, the same variables and model building procedure were used.

TABLE 3

Significant Effects Found in Models for Examining the Role of Noticed LE in IVL

Outcome Measure	Predictors
Meaning Recognition	Prior Knowledge ($z = 6.12$, $p < .001$), ClozeZ ($z = 3.38$, $p < .001$), Noticed LE ($z = 2.25$, $p = .024$), Proficiency:Noticed LE ($z = 2.23$, $p = .026$)
Meaning Recall	Prior Knowledge ($z = 10.64$, $p < .001$), ClozeZ ($z = 4.45$, $p < .001$), BP ($z = 2.01$, $p = .045$), Noticed LE ($z = 2.96$, $p = .003$), Proficiency:Noticed LE ($z = 2.73$, $p = .006$)

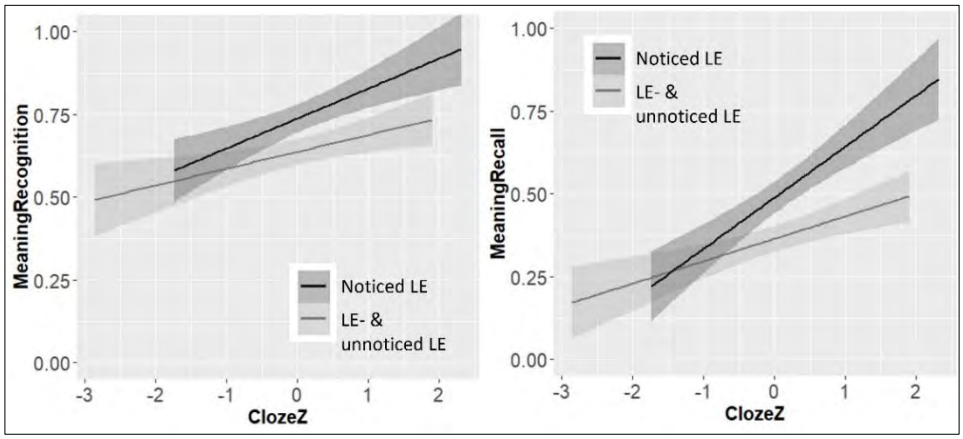
The results are summarized in Table 3 (see Table S16 for specific model output). Noticed LE was a significant predictor of meaning recognition ($z = 2.25$, $p = .024$) and meaning recall

($z = 2.96, p = .003$). This indicates that although LE did not significantly affect word meaning learning, once LE was noticed, it significantly facilitated meaning acquisition.

Another important finding was a significant interaction between ClozeZ and noticed LE in both meaning recognition ($z = 2.23, p = .026$) and meaning recall ($z = 2.73, p = .006$). This interaction suggests that the more proficient participants were, the more effective noticed LE was (see Figure 4). This implies that the significant interaction between ClozeZ and LE for meaning recall addressed above was not because proficiency determined the probability of noticing LE, but because high-proficient participants were more successfully able to develop further semantic knowledge while reading *after* noticing LE compared to lower-proficiency participants. This interpretation is in line with the idea that incidental vocabulary learning is incremental in nature.

FIGURE 4

The Interaction between Noticed LE and ClozeZ in Meaning Recognition (Left) and Meaning Recall (Right)



5. DISCUSSION

5.1. Bimodal Presentation

The current study first observed a null effect of BP, as well as that of LE, on meaning recognition. This was presumably because considerable word meaning acquisition had occurred even without BP or LE as shown in the "Effect of Reading" section. This indicates that when repeatedly encountering L2 words in semantically rich contexts, learners can successfully develop semantic knowledge of those words up to a certain level, and BP and

LE do not make significant differences. However, this study shows that BP significantly boosts meaning recall of unfamiliar words. This suggests that BP is beneficial for obtaining more concrete semantic knowledge of unfamiliar words. The finding is important since it has been pointed out that one of the biggest limitations of contextual vocabulary learning is its inefficiency for gaining stable and accurate semantic knowledge (e.g., Laufer, 1997, Laufer & Nation, 2012).

As summarized in the literature review section of this article, previous research suggests that the advantages of BP for IVL would be due to more obvious prosodic connections between novel words and semantic clues, a deeper processing of the text accompanied by a simultaneous stream of information, or the ease of phonological decoding. All these accounts convincingly explain how BP could produce greater learning outcomes in the current study. However, what this study newly discovered is that it may not always be the case that BP helps overall comprehension of the text or makes L2 learners comfortable while reading. In the present study, while many participants reported that BP was helpful for their reading comprehension (45%), a substantial proportion of participants responded that BP interfered with it (33%) (see Vu & Peters, 2022 for a similar finding). This indicates individual differences in L2 learners' preferences about BP. Moreover, it would be reasonable to predict that their perception of BP would also be affected by the characteristics of the text such as its complexity or readability. However, an interesting finding is that in terms of IVL, BP was as effective for participants who felt that it interfered with their reading comprehension as for those who felt that it helped their reading.

In addition to the suggested mechanisms, the effect of BP can also be explained by increased attention. Researchers have been interested in finding a way to facilitate incidental learning (e.g., Godfroid et al., 2013; Pellicer-Sánchez, 2016), and the findings of the current study supports the proposed hypothesis that BP might be useful for achieving this goal (Long, 2017; Pellicer-Sánchez & Boers, 2019; Webb & Chang, 2015). Participants without BP might have overlooked the forms of target words and semantic clues related to them when their attention was oriented to following the main ideas of the text. On the contrary, when BP was provided, their attention might have been maintained when they encountered target words and details of the reading passages, as shown by L2 learners' greater regression counts (i.e., eye movements back to a previously read part of a text) during reading-while-listening than during reading-only in a previous eye-tracking study (Conklin, Alotaibi, Pellicer-Sánchez, & Vilkaitė-Lozdienė, 2020). In the present study, the evidence of increased attention through BP is found in the result that the BP- groups significantly more frequently than the BP+ groups misperceived unknown target words as known words, even after four encounters with target words. In addition, when BP was provided, LE in the text was more likely to be noticed.

Finally, before being overoptimistic about the effect of BP, it is necessary to note that if

target words occur very frequently within short intervals as in the present study, it is possible that the effect of BP diminishes. For example, Brown and his colleagues (Brown et al., 2008) report that the number of exposures to target words was more influential for IVL than BP was. This suggests that with a great number of occurrences, target word meanings would be learned successfully even without BP. However, it is also predictable that when target words appear frequently with BP, BP can contribute to the establishment of more stable and fully specified semantic representations beyond an observable level. If that is the case, drawing a conclusion about the null effect of BP after using only measures of shallow semantic knowledge should be avoided.

5.2. Lexical Elaboration

In the present study, the main effect of LE was not found in meaning recognition or meaning recall. This contrasts with previous research reporting significant gains in word meaning knowledge with the use of explicit LE (Kim, 2006; Toya, 1993; Vidal, 2011). However, in most previous studies, target words occurred only once (Kim, 2006) or twice (Toya, 1993) in reading passages. Therefore, it can be reasoned that LE makes a difference when target words are rarely repeated. These two studies also used only sensitive instruments that could measure an initial state of newly acquired semantic knowledge (e.g., a meaning recognition task). Therefore, no evidence was found supporting the assumption that just one encounter with LE can lead to the acquisition of specific word meaning through memorization of the LE. The significant interaction between proficiency and noticed LE observed in the present study further suggests that the mechanism LE contributes to word meaning acquisition would be mostly by assisting learners' understanding of the word meaning. Once novel words are understood, the understood meaning of these words would be retrieved by necessity, such as at a vocabulary posttest in which learners would try to recall incidentally collected information about the words in order to respond to questions. However, when target words occur multiple times in meaningful contexts within short intervals, the effect of LE would be less significant because target words can be understood even without LE (Pellicer-Sánchez, 2016). In a similar vein, the significant effect of explicit LE observed in Vidal's (2011) study is suspected to have been influenced by other confounds. In that study, the frequency of target word occurrence (1 ~ 6 occurrences) or the levels of predictability of L2 word meaning (based on an L1 word with a similar form) was not tightly controlled, and only the overall effect of LE on all target words was reported.

Does this mean that LE is unnecessary if target words appear multiple times? Before answering this question, we should recall that in the present study, once LE was noticed it significantly predicted successful meaning recognition and meaning recall. Moreover, for proficient participants, the effect of LE on meaning recall was greater. These results suggest

that although not guaranteed, LE has a potential to facilitate at least some learners' vocabulary acquisition.

5.3. Limitations of the Study

This study has several limitations. First, noticing was measured by quick survey questions in a retrospective manner. Therefore, participants' responses could have been influenced by memory decay. Objective measures of noticing (e.g., eye-tracking) in future research would help overcome this limitation.

Second, the treatment group's prior knowledge was asked through a survey, so the level of precision of the measure could be questioned. Indeed, some researchers (e.g., Huibregtse, Admiraal, & Meara, 2002; Zhang, Liu, & Ai, 2020) have warned that learners tend to overestimate their prior vocabulary knowledge in self-assessment. However, we claim that in the current study, the treatment group's overestimation of their prior knowledge was accounted for at least to some extent since their responses on the retrospective vocabulary pretest were verified through their meaning recall responses.

Another potential problem of the post-treatment retrospective vocabulary pretest may be found from the observation that although the five groups had comparable levels of proficiency, these groups showed different levels of preexisting vocabulary knowledge of target words. Specifically, the treatment groups' prior knowledge was greater than the control group's knowledge, and the BP+ groups also showed higher pretest scores compared to the BP- groups (see Appendix S1 for specific results of statistical tests). This seems to indicate that since the treatment groups' preexisting knowledge was not measured *a priori*, it might have been strengthened over the treatment session. In other words, unlike in typical vocabulary learning studies that utilize an actual vocabulary test as a pretest, even the words with very limited prior knowledge before the treatment session, which are usually categorized as 'unknown words', might have been counted as 'known words' in the current study. Although these results may indicate a lack of precision in the measurement from a strict perspective, it should be noted that the treatment groups showed higher posttest scores than the control group even after accounting for the former groups' higher pretest scores. This was also the same for the BP+ groups who showed higher meaning recall scores than the BP- groups, even after accounting for the former groups' greater prior knowledge. Put differently, if prior knowledge of target words had been measured before the treatment was provided (i.e., before prior knowledge was strengthened), the observed effects of reading and BP on IVL would have been greater than the ones observed in the current study. Therefore, the use of the post-treatment retrospective vocabulary pretest does not threaten the validity of the observed positive effects of reading and BP on IVL. On the other hand, a stronger effect of LE might also have been observed if preexisting vocabulary knowledge of

target words was measured before strengthening, so this is a limitation of this study.

6. CONCLUSION

The study found that BP is helpful for developing concrete semantic knowledge of unfamiliar words during reading when these words appear multiple times in the text. It also found that LE can promote word meaning acquisition if it is noticed. These findings suggest that BP and LE can be usefully employed in materials development.

Applicable levels: Early childhood, elementary, secondary, tertiary

REFERENCES

Baayen, R. H., Davidson, D. J., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390-412.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255-278.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). *Parsimonious mixed models*. https://arXiv:1506.04967

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1-48.

Bonate, P. L. (2000). *Analysis of pretest-posttest designs*. Boca Raton, FL: Taylor & Francis Group.

Brown, J. D. (1980). Relative merits of four methods for scoring cloze tests. *The Modern Language Journal*, *64*(3), 311-317.

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

Bruton, A., Garcia Lopez, M., & Esquiliche Mesa, R. (2011). Incidental L2 vocabulary learning: An impracticable term? *TESOL Quarterly*, *45*(4), 759-768.

Chaudron, C. (1982). Vocabulary elaboration in teachers' speech to L2 learners. *Studies in Second Language Acquisition*, *4*(2), 170-180.

Chaudron, C. (1988). *Second language classrooms: Research on teaching and learning*. New York: Cambridge University Press.

Conklin, K., Alotaibi, S., Pellicer-Sánchez, A., & Vilkaitė-Lozdienė, L. (2020). What eye-tracking tells us about reading-only and reading-while-listening in a first and second language. *Second Language Research*, *36*(3), 257-276.

Cunnings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research*, *28*(3), 369-382.

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

Ellis, N. C. (1994). Vocabulary acquisition: The implicit ins and outs of explicit cognitive mediation. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 211-282). New York: Academic Press.

ETS (English Testing Service). *Mapping table*. Retrieved on August 15, 2023, from https://etswebsiteprod.cdn.prismic.io/etswebsiteprod/515ff84d-fdd0-4e7f-8d79-e19b4d19e9db_TOEIC+Listening+and+Reading+Mapping+Table+CEFR-092021.pdf

Gass, S. M. (1999). Incidental vocabulary learning. *Studies in Second Language Acquisition*, *21*(2), 319-333.

Godfroid, A., Boers, F., & Housen, A. (2013). An eye for words: Gauging the role of attention in L2 vocabulary acquisition by means of eye-tracking. *Studies in Second Language Acquisition*, *35*(3), 483-517.

Horst, M., Cobb, T., & Meara, P. (1998). Beyond a clockwork orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language*, *11*(2), 207-223.

Huibregtse, I., Admiraal, W., & Meara, P. (2002). Scores on a yes-no vocabulary test: Correction for guessing and response style. *Language Testing*, *19*(3), 227-245.

Hulstijn, J. H. (2003). Incidental and intentional learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 349–381). Oxford, England: Blackwell.

Hulstijn, J., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *The Modern Language Journal*, *80*(3), 327-339.

Kim, Y. (2006). Effects of input elaboration on vocabulary acquisition through reading by Korean learners of English as foreign language. *TESOL Quarterly*, *40*(2), 341-373.

Korean Ministry of Education. (2015). *National curriculum for English*. Retrieved on January 21, 2020, from https://ceri.knue.ac.kr/pds/2015_02_english.pdf

Laufer, B. (1997). The lexical plight in second language reading. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 20-34). Cambridge, England: Cambridge University Press.

Laufer, B., & Nation, I. S. P. (2012). Vocabulary. In S. M. Gass, & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 163-176). New York: Routledge.

Long, M. H. (2017). Instructed second language acquisition (ISLA): Geopolitics, methodological issues, and some major research questions. *Instructed Second Language Acquisition*, *1*(1), 7-44.

Malone, J. E. (2018). Incidental vocabulary learning in SLA: Effects of frequency, aural enhancement, and working memory. *Studies in Second Language Acquisition*, *40*(4), 651-675.

Millett, S. (2017). *Speed readings for ESL learners: 3000 BNC*. School of Linguistics and Applied Language Studies, Victoria University of Wellington. Retrieved on October 10, 2019, from https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/speed-reading-and-listening-fluency.

Nagy, W. E., Herman, P. A., & Anderson, P. C. (1985). Learning words from context. *Reading Research Quarterly*, *20*(2), 233-253.

Paribakht, T. S., & Wesche, M. (1999). Reading and "incidental" L2 vocabulary acquisition: An introspective study of lexical inferencing. *Studies in Second Language Acquisition*, *21*(2), 195-224.

Pellicer-Sánchez, A. (2016). Incidental L2 vocabulary acquisition from and while reading: An eye-tracking study. *Studies in Second Language Acquisition*, *38*(1), 97-130.

Pellicer-Sánchez, A., & Boers, F. (2019). Pedagogical approaches to the teaching and learning of formulaic language. In A. Siyanova-Chanturia & A. Pellicer-Sánchez (Eds.), *Understanding formulaic language* (pp. 153-173). New York: Routledge.

Rott, S. (1999). The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition and retention through reading. *Studies in Second Language Acquisition*, *21*(4), 598-619.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*(2), 129-158.

Suzuki, Y., & DeKeyser, R. (2017). The interface of explicit and implicit knowledge in a second language: Insights from individual differences in cognitive aptitudes. *Language Learning*, *67*(4), 747-790.

Toya, M. (1993). Form of explanation in modification of listening input in L2 vocabulary learning. *University of Hawai'i Occasional Papers Series*, *23*, 1-168.

Uchihara, T., Webb, S., & Yanagisawa, A. (2019). The effects of repetition on incidental vocabulary learning: A meta-analysis of correlational studies. *Language Learning*, *69*(3), 559-599.

Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning*, *61*(1), 219-258.

Vu, D. V., & Peters, E. (2022). Learning vocabulary from reading-only, reading-while-listening, and reading with textual input enhancement: Insights from Vietnamese EFL learners. *RELC Journal*, *53*(1), 85-100.

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, *28*(1), 46-65.

Webb, S. (2008). The effects of context on incidental vocabulary learning. *Reading in a Foreign Language*, *20*(2), 232-245.

Webb, S., & Chang, A. C-S. (2012). Vocabulary learning through assisted and unassisted repeated reading. *Canadian Modern Language Review*, *68*(3), 267-290.

Webb, S., & Chang, A. C-S. (2015). Second language vocabulary learning through extensive reading: How does frequency and distribution of occurrence affect learning? *Language Teaching Research*, *19*(6), 667-686.

Webb, S., & Chang, A. C.-S. (2022). How does mode of input affect the incidental learning of collocations? *Studies in Second Language Acquisition*, *44*(1), 35-56.

Webb, S., & Nation, I. S. P. (2013). Teaching vocabulary. In C. Chappelle (Ed.), *Encyclopedia of applied linguistics* (pp. 5670-5677). Oxford, England: Wiley-Blackwell.

Zhang, X., Liu, J., & Ai, H. (2020). Pseudowords and guessing in the Yes/No format vocabulary test. *Language Testing*, *37*(1), 6-30.

APPENDIX

Examples of the Three Types of the Measurements

A. Form Recognition Task

Direction: Select “Yes” if you saw the word in the texts you read. Select “No” if you did not see the word.

- Sanatorium: Yes ○ No ○

B. Meaning Recognition Task

Direction: Choose a sentence in which you think the use of the underlined word is the most appropriate.

- Sanatorium
  a. Students can drink water, low-fat milk, fruits or sanatorium juice.
  b. He knows how to cook a sanatorium.
  c. Her mother is a Sanatorium.
  d. He was put in a sanatorium when it was found that he was suffering from mental illness.

C. Meaning Recall Task

Direction: Please write what you think each of the following English words mean in Korean. If you do not know an exact Korean translation, provide a synonym or describe the meaning of a word as specifically as possible.

- Sanatorium:
  ( )