Effects of Enhancement Techniques on L2 Incidental Vocabulary Learning

Shiping Duan[1]

[1] School of Foreign Languages, Zhongyuan University of Technology, Zhengzhou, China

Correspondence: Shiping Duan, School of Foreign Languages, Zhongyuan University of Technology, No.41 Zhongyuan Road (Middle), Zhengzhou, Henan, 450007, China. Tel: 86-371-6250-6978. E-mail: dsproyce@zut.edu.cn

Abstract

Enhancement Techniques are conducive to incidental vocabulary learning. This study investigated the effects of two types of enhancement techniques-multiple-choice glosses (MC) and L1 single-gloss (SG) on L2 incidental learning of new words and retention of them. A total of 89 university learners of English as a Freign Language (EFL) were asked to read the same reading texts with the two types of glossing and no glossing. Vocabulary acquisition was measured with the vocabulary knowledge scale (VKS). The results indicated that there were obvious vocabulary gains for both MC and SG groups. MC glossing is more conducive to incidental vocabulary learning than SG glossing in both immediate and delayed vocabulary post test. What's more, learners with larger vocabulary size demonstrated much more significant gains than those with small ones.

Keywords: incidental vocabulary learning, enhancement techniques, multiple-choice glosses, VKS, task-induced involvement

1. Introduction

Due to the impossibility of teaching all the words in English, incidental vocabulary learning has long been brought to the front and become a hot issue in the field of second or foreign language acquisition. Not surprisingly, teachers as well as learners have always shown a keen interest in the possibilities in incidental vocabulary learning through reading. However, many studies indicate that the process of L2 incidental vocabulary learning (IVL) is slow and error-prone with small vocabulary gains (Read, 2004; Laufer, 2005; Peters, Hulstijn, Sercu, & Lutjeharms, 2009；Khezrlou, Ellis, & Sadeghi, 2017). What's more, in formal language learning environment, learners receive insufficient exposure to target language. Therefore, L2 IVL calls for lots of efforts invested to explore its enhanced techniques.

Research on learning and memory and the literature on L2 vocabulary learning in particular, show that successful L2 IVL is contingent on three factors: noticing and processing of the meaning of unfamiliar words, and repetition of the form-meaning mapping (Hulstijn, Sercu, & Lutjeharms, 2009, p. 114). Related studies suggest that elaborate processing of lexical information is conducive to word learning, which can find strong evidence from the following theories as noticing, depth of processing, and task-induced involvement (Craik, 2002; Craik & Tulving, 1975; Hulstijn, 2001, Reynolds, 2014; Teng, 2017).

Thus, many enhanced techniques-glossing, bolding, italicizing, color-coding, or word-focused exercises, multi-media annotation--have been employed to increase the effectiveness of incidental vocabulary learning while reading. Among these techniques, marginal glossing, multimedia glossing in particular, have been commonly used recently to enable language learners to process unknown words elaborately and thus facilitate vocabulary learning (Hulstijn, Sercu, & Lutjeharms, 2009; Khezrlou, Ellis, & Sadeghi, 2017; Sun, 2017; Rassaei, 2017). But concerning printed reading materials, a traditional and major format of reading, its marginal glosses has been relegated to be a Cinderella, a secondary position, which remains to be explored.

Considerable studies show that vocabulary growth can be enhanced by comprehensible input in the form of reading accompanied by enhanced techniques, such as glosses. Marginal glosses, one of the main enhanced techniques can be presented in different formats-single-gloss type (i.e. providing L 1 equivalent for the unfamiliar word in the reading text) and multiple-choice glosses (in which a correct synonym and two distractors

for each target word in L1 are presented for each unfamiliar word), yet which gloss type can achieve better effect is still unclear. In the meanwhile, however, L1 single-gloss remains a common and acceptable aid for L2 reading materials. That practice is especially popular in China.

Given the fact that marginal glossing enables, and in a sense encourages, learning of vocabulary, it is important for language researchers and teachers to know clearly about the usefulness of each type of glossing, specifically, to know if single glossing is superior over no glossing, and multiple-choice glossing over single glossing. Although the question of the role of glossing in L2 vocabulary acquisition does not seem new, it is actually far from being a well-settled issue. The fact that this topic is familiar to many people just adds to the importance and practical value of the research on it. To gain further insights into IVL through glossing, more studies are needed.

In view of the above theoretical and practical considerations, the present study aims to examine closely the unanswered question of the effect of marginal glossing on vocabulary learning, which, in a sense, still stays at a "common sense" level among L2 pedagogy practitioners, and which is still a controversial issue for L2 acquisition researchers.

2. Literature Review

2.1 Depth of Vocabulary Knowledge and Vocabulary Knowledge Scale

Research on vocabulary learning requires a definition of vocabulary knowledge. No unequivocal consensus has so far been reached as to the nature of lexical knowledge. Most researchers in this domain prefer to regard it as a continuum consisting of a variety of dimensions of knowledge. Vocabulary Knowledge Scale (VKS) is so far the best and the most convenient measurement instrument ever developed to measure depth of vocabulary knowledge.

Previous research suggests that lexical knowledge can be explored from various dimensions. Notably, there are two aspects of the problem involved. Breadth of knowledge (vocabulary size) is concerned with the question "How much vocabulary does an L2 learner need?" (Nation & Waring, 1997). Yet, "knowing a word requires more than just familiarity with its meaning and form" (Schmidt & McCarthy, 1997). Hence, depth of vocabulary is also at issue. It is related to the quality of the learner's vocabulary knowledge. Many researchers view the representation of lexical knowledge as a continuum. And the notion of continua of vocabulary acquisition has been converted into 'scales of vocabulary knowledge', which can be found in numerous recent studies. Rob Waring (2002) points out, the theoretical construct underlying such scales assumes that word knowledge is not bipolar in nature i.e. "known" or "unknown", but linear, involving several stages of acquisition which can be measured by degrees.

Virtually, these stages are the embodiment of vocabulary acquisition continuum, which starts with a vague familiarity with the word, and ends with the ability to use the word correctly in the production, while during this process, different aspects of word knowledge are growing in a linear way (Laufer & Paribakht, 1998). Therefore, these scales can be used to assess the degree of vocabulary knowledge held by a learner.

Paribakht and Wesche (1997) proposed the 5-point scale Vocabulary Knowledge Scale (VKS), which has gained significant circulation in L2 vocabulary assessment. The particular aim of the VKS is to construct a "practical instrument for use in studies of the initial recognition and use of new words" (p. 29). The VKS roughly corresponds to the continuum, ranging from 'passive real' to 'active real' vocabulary. "This instrument uses a 5-point scale combining self-report and performance items to elicit self-perceived and demonstrated knowledge of specific words in written form. The scale ratings range from total unfamiliarity, through recognition of the word and some idea of its meaning, to the ability to use the word with grammatical and semantic - accuracy in a sentence" (ibid, p. 179). Paribakht and Wesche identify five stages of vocabulary knowledge, which are listed below (see Figure 1).

	Self-reported Categories
I	I don't remember having seen this word before.
II	I have seen this word before but I don't know what it means.
III	I have seen this word before and I think it means _____.
IV	I know this word. It means _____.
V	I can use this word in a sentence. e.g. _____.

Figure 1. VKS elicitation scale: self-report categories

Obviously these vocabulary knowledge scales entail declarative knowledge. In vocabulary assessment, these scales are most often used by learners to self-assess their knowledge of a list of words. Proponents of these tests of vocabulary therefore suggest that a single scale can assess both breadth (how many words) and depth (how well known) aspects of vocabulary knowledge (Rob, 2002).

2.2 L2 Incidental Vocabulary Learning and Enhanced Techniques

Several researches indicate that both guessing and dictionary use while reading are conducive to word retention, yet readers using guessing often make erroneous inferences (Hulstijn, 1992), and dictionary use is time-consuming, for it often interferes with text comprehension. Therefore, learning vocabulary incidentally through reading alone is a very slow and error-prone with lower pick-up rate, even more, it is difficult to predict which words will be learned, when and to what degree (Coady & Huckin, 1997; Paribakht & Wesche, 1997; Laufer, 2005; Read, 2004, Peters, Hulstijn, Sercu, & Lutjeharms, 2009, p. 114).

To facilitate IVL processes of L2 learners, empirical research in SLA has turned to investigate the effects of enhancement techniques - textual glosses, i.e, provision of L1 translation or brief explanation of assumed unfamiliar words. Such an aid may take the forms of marginal glosses. These can be of two types: single glosses and multiple-choice glosses.

Marginal L1 single-gloss, though in a way enhancing text comprehension and IVL, requires a little effort when processing unfamiliar words (Hulstijn, 1996, p. 328). The result with word retention may not be as good as inference and dictionary use, in that marginal glosses do not even demand the effort of searching and much 'involvement' in processing the new word, then choosing the appropriate meaning out of several possible ones, which is required by dictionary look-up.

A study by Watanabe (1997) found that both single and multiple-glosses groups performed significantly better on the vocabulary post-tests than the appositives and text-only groups. Furthermore, the gloss group yielded higher mean scores than the multiple-choice group, though the difference was not statistically significant. Hulstijn (1992, Experiment III) examined the relative effectiveness of a multiple-choice approach and a single synonym approach in a within-subject design. The experiment shows a higher retention effect for the multiple-choice group than for the single-synonym procedure on post-test, which contradicts the results of Watanabe's study (1997).

The contradiction lies in that both Hulstijn's study (1992) and Watanabe's study (1997) used printed text without the provision of feedback whether students' selected answer was correct. Thus, mistaken answers would go unnoticed. Therefore, Multiple-choice glossing, while encouraging deeper processing, suffers (in the printed form) from the lack of immediate feedback to student errors.

The problem mentioned above can be addressed by providing immediate feedback concerning students' selections. Nagata (1999) solved the problem by using a computer to provide on-going, immediate feedback regarding mistaken selections. The result of his study suggests that multiple-choice glossing format is significantly more effective than single glossing format for recalling target vocabulary.

Thus, compared with single glossing, multiple-choice glossing (MC) with immediate feedback has proved to be more conducive to IVL. This should be mainly attributable to the fact that MC glossing can stimulate the cognitive processes involved in such notions as "noticing", "awareness", and "mental effort". The success of multiple-choice glosses may be explained by the following factors:

(a) Multiple-choice glosses are easier to use than a dictionary.

(b) They draw learners' attention to target words, supporting the notion of "consciousness-raising" and "input enhancement".

(c) They help to connect words to meanings immediately, contributing to the "meaning-form connection" approach.

(d) They encourage learners to pass back and forth between target words and glosses, stimulating more of the learners' mental effort to perform lexical processing, eventually contributing to the retention of the words, which is in line with the "depth of processing" hypothesis (Craik & Lockhart, 1972; Craik and Tulving, 1975).

2.3 L2 Input Enhancement and Depth of Processing

Krashen's comprehensible input hypothesis (1985) tell us that vocabulary acquisition needs input modification to speed up the rate of acquisition, through stimulating the cognitive processes involved in such notions as 'noticing', 'awareness' and 'mental effort'. In a word, the more L2 readers experience favorable conditions for lexical input to become intake, the more likely incidental vocabulary acquisition is to occur.

Input modification in the form of elaboration, is actually a kind of input enhancement. Thus arise two questions: can input modification or enhancement facilitate acquisition? If so, in what way can input be modified or enhanced to facilitate acquisition?

The first question can be answered by Krashen's Input Hypothesis and Schmidt's Noticing Hypothesis (1990), which claim that input must be noticed and comprehensible in order to be acquired. Input modification must be done to enable learners to "notice" the input, and to enhance comprehension. Then acquisition can eventually be facilitated. Here take input enhancement for instance. It can be achieved through various ways, such as increasing frequency, increasing salience (e.g., highlighting), providing additional information (e.g., marginal glossing). This is also consistent with the 'mental effort' hypothesis (Hulstijn, 1992), which suggests that information that has been attained with more mental effort can later be better retrieved and recalled than information that has been attained with less mental effort. Thus, the retention of an inferred word meaning is predicted to be higher than the retention of a given word meaning (Craik & Tulving, 1975).

The depth of processing theory (DOP), one of the major long-term memory encoding theories in cognitive psychology (Craik & Lockhart, 1972), suggests that levels of processing depth which occurred at the time of learning produce an important effect on the memory duration and strength or knowledge learned---the deeper the process, the better memory becomes (Hulstijn & Laufer, 2001). In other words, the enhancement of memory is determined by elaboration.

Influential as it is, DOP theory was challenged by Eysenk (1978) and Nelson (1977) by questioning that the concepts of deep processing or elaboration are hard to formalize and operationalize. In answer for this, Hulstijn and Laufer (2001) proposed the Task-Induced Involvement Load Hypothesis for L2 vocabulary learning, which is a motivational-cognitive construct, consisting of three basic components: need, search and evaluation. The need component is the motivational, non-cognitive dimension of involvement, whereas search and evaluation are the two cognitive dimensions of involvement, contingent upon allocating attention to form-meaning relationships. According to the Hypothesis, the retention of unfamiliar words, generally, depends upon the degree of involvement in processing these words. Namely, words that are processed with higher involvement load will be retained better than words that are processed with lower involvement load. The superiority of Involvement Load over the Depth of Processing resides in the fact that the Involvement Load is easily observable and measurable, and thus can be operationalized and empirically investigated. The meta-analytic study conducted by Huang, Willson, and Eslami (2012) provided a strong evidence for the effectiveness of Involvement Load Hypothesis.

In view of above-discussed psycholinguistic point of view, researchers have investigated different gloss types through raising noticing and involvement load of difficult vocabulary and increasing the mental effort learners invest in trying to infer the meaning of unfamiliar words. Contrasted with "given" information, like definition or translations (that is, single-gloss), the most common example of input modification containing "inferred" information are multiple-choice glosses, which require readers to match L1 and L2 meanings. Therefore, the present study is designed to investigate the effectiveness of the two types of glossing with an eye to offering some suggestions for reading curricular design.

3. Aims and Research Questions

Given the fact that marginal glossing enables, and in a sense encourages, learning of vocabulary, it is important for language researchers and teachers to know clearly about the usefulness of each type of glossing, specifically, to know if single glossing is superior over no glossing, and multiple-choice glossing over single glossing. To gain further insights into IVL through glossing, more studies about effects of glossing types on incidental vocabulary learning are needed.

To this end the following research questions were addressed:

(a) Which gloss type, multiple-choice glosses, L1 single-gloss, no-gloss (text only), best facilitates incidental vocabulary learning in immediate VKS in immediate vocabulary test?

(b) Is there any change in the results of the delayed test?

(c) Does vocabulary size have an effect on IVL gains and retentions under glossing conditions?

4. Method

4.1 Participants

The participants were three intact groups of intermediate university learners of English as a Foreign Language at *Henan University of Science & Technology*, one group randomly assigned to each task. The three groups were parallel classes of second-year English majors, whose average vocabulary size is assumed to be above 3,000

word families. We consider them parallel because they were similar in proficiency, age, etc. Most importantly, the target words were new to all the participants.

The overall number of students taking part in the study was 89, Group I (N=30) read the text with multiple-choice glosses for Target Words (TWs), Group 2 (N=30), the one with single glossing for TWs, and Group 3 (N=29), the one with no glossing for TWs. Some students, however, dropped out of the delayed post-test. Because the test was not announced in advance (owing to the incidental learning nature of the study), we could not ensure that all the students who took part in the study would be present at the delayed test session.

4.2 Design

The present study was designed mainly to test the effect of marginal glossing types on incidental vocabulary learning, that is, the retention of the words the participants had encountered in reading. The participants were divided into three groups: Multiple-Choice glossing (MC) group who were to read a passage with multiple-choice glosses; Single-Glossing (SG) group who were to read the same passage with single L1 gloss; and Control (C) group, reading the same passage' but not receiving any type of glossing. Students were instructed to read the passage and prepare to translate it, i.e. write out the content of the passage in great detail. Eleven words were carefully selected as targets. After reading the text, students were required to do two posttests: the immediate test and delayed test for their both receptive and productive knowledge of target words.

4.3 Materials

4.3.1 Reading Passage

The title of the passage that students read was "Ring a ling a ling about those cell phones, I do sing" from *China Post*. It is an essay concerning the use of cell phones. The text is about 570 words in length including 11 unknown words. It is a relatively easy text, which second-year English majors should be able to read without much difficulty. This estimation was based on extensive piloting. Prior to the experiment, a pilot test directing "read and write out the content afterwards" was administered with non-English majors. The follow-up interview indicated that the text was appropriate for them. For English majors, therefore, the text should be rather easy.

The reason why we chose the 570-word passage with 11 unfamiliar words was based on some research concerning vocabulary; threshold for reading comprehension. A definite answer has not been offered as to the optimal ratio of unknown words to known words in a text. Yet according to Laufer (1992), a minimum of 95% text coverage will lead to the transfer of reading strategies. Hirsh and Nation (1992), however, argue that around 98% coverage of vocabulary is needed for learners to gain unassisted comprehension of a fiction text. Thus, 11 unknown words in a 570- word passage were appropriate for the participants to guess in reading. The new words accounted for less than 2% of the total number. To test the glossing effects, a short passage as of 570 words would work well.

In the present study, the three groups read the same passage with the same reading instructions, but under different glossing types, that is, Group 1 under Multiple-Choice glossing; Group 2, under single glossing; group 3, under no glossing. Two examples of glossing versions are as follows. Example (a) shows the single-glossing; and example (b) illustrates the multiple-choice glossing. For example,

(a) Single-glossing1. flummoxed:困惑的

(b) Multiple-Choice glossing 1. flummoxed: A. 困惑的   B. 激动的   C. 伤心的

4.3.2 Target Words

The rationale for selection of target words should be that target words are unknown to participants. To achieve such an effect, we took the following steps:

(a) We checked these target words in Collins COBUILD English Dictionary to ensure that most of them are low-frequency words. Because we estimated them to be unfamiliar to most second-year students, and the dictionary indicates that 95% words are in five frequency bands, the remaining 5% that have no frequency band are words which you will probably read or hear rather than words which you need to use yourself. Of the eleven words, four words (*flummoxed, stodgy, a smattering of, and whopping*) are in no frequency band; four words (*ubiquitous, tote, anecdotal and infringe upon*) in the bottom (i.e. the lowest-frequency) band; two words (*anonymous, query*) in the bottom two (i.e. the lower-frequency) band, and only one (plunge) in the middle band. Thus, the 10 low frequency words make up 90% of the target words.

(b) We checked these target words in *Lexicon for English Majors - A Supplement to the English Curriculum* published by the Shanghai Foreign Language Education Press (2001). And most of them are not included in it.

(c) These words were tried on 4 English majors of the same English proficiency, who did not participate in the experiment. They were asked to mark the unknown words. And then an interview was conducted to make sure whether these unknown words were perceived relevant to their understanding of the passage.

(d) We presented the passage to 4 experienced teachers and asked them to mark the unknown words. They were asked to indicate whether each boldfaced word (assumed target words) was known to only good students, or unknown to all students. Every word marked at least three times as unknown to all students or only known to good students was considered appropriate for the experiment. We then selected 12 target words.

(e) A pilot test was conducted with non-English majors to further ensure the validity of these target words. The compound word earshot was found conducive to students' memory. And also based on students' response and my supervisor's suggestion, the word earshot was eliminated. Finally, 11 words were selected.

(f) The final selection ensured that not too many topic-specific words were chosen, and that the target words consisted of six adjectives, three verbs and two nouns. The part of speech of each word was indicated in parentheses. All the target words and their part of speech were glossed in the margin of the reading text. And to make them salient, all of them were boldfaced in the text.

4.4 Tests

4.4.1 Immediate and Delayed Vocabulary Tests

To evaluate incidentally learned word knowledge, Vocabulary Knowledge Scale (VKS), slightly adapted from Paribakht and Wesche's VKS, 1997) was adopted. This instrument is complex to operationalize, yet it does give rich insights into participants' receptive and productive aspects of word knowledge in that this measure was supposed to be sensitive enough to detect small learning increments. Therefore, we chose to use VKS as the testing measure.

This instrument in our study used a 4-point scale combining self-perceived and demonstrated knowledge of specific words in written form. The scale ratings range from total unfamiliarity, through recognition of the word and some idea of its meaning, to the ability to use the word with grammatical and semantic accuracy in a sentence. Participants were presented with a list of target words and asked to indicate their level of knowledge. This instrument is presented in Figure 2.

categories	Meaning
(1)	I don't remember having seen this word before.
(2)	I have seen this word before but I don't know what it means.
(3)	I understand the word when I see it in a sentence, but I do not use it in my own speaking or writing. It means _____; (synonym or Chinese translation)
(4)	I can use this word in a sentence. e.g. _____.

Note: adapted from Parbakht and Wesche (1997).

Figure 2. Adapted VKS categories

After the reading tasks, participants were required to do two tests: immediate vocabulary test and delayed vocabulary test. Actually, they were the same. The delayed test aimed to test IVL retention, which was held 7 days later.

4.4.2 Test for Receptive Vocabulary Size

Various measurement instruments have been proposed to measure vocabulary knowledge. For the receptive vocabulary size, the notable one is the Vocabulary Levels Test (Nation, 1990). This test is based on words from 5 word-frequency levels (2,000, 3,000, 5,000, the University Word List, the 10,000 words) and tests the participants' word knowledge out of context. The participants are required to match the words in the left column with their corresponding definitions in the right column (see appendix 1). All the participants were required to do the test.

4.5 Procedure

Step 1 One reading plus testing session was held with second-year English majors: Group 1 and Group 2 with 30 students respectively, Group 3 with 29 students. The three groups were randomly assigned to Multiple-Choice

Glossing, Single-Glossing and Control conditions. They were given the following information: (1) Participants would first read a text, and then write out the content of the text in Chinese in great detail; (2) Participants would not have the text at their disposal when writing out the content; (3) 20 minutes would be allotted for reading the passage; (4) Participants would not be allowed to use the dictionary.

Moreover, participants were not told in advance that they would be tested later on their knowledge of the glossed words; instead, they were told that they would have to write out the content of the passage without looking back at the text (for the text had been collected after participants read it). This was administered to create conditions conducive incidental vocabulary learning: participants' attention was diverted from *specific unknown words*, and directed towards an understanding of the passage *as a whole*.

Step 2 After the collection of the texts, the target word tests were handed out. Adjacent students received different versions of the same test. According to the meaning each number indicates, the students were told to choose the right one which suited most. The instructions were written in both English and Chinese to ensure students fully understand the requirements.

The administration of immediate vocabulary test took about 15 minutes. The entire session (reading plus vocabulary levels test) lasted 35 minutes. No mention of delayed vocabulary test was made.

Step 3 Seven days later, the vocabulary delayed test, which has the same content as "immediate test", was conducted to test the retention of incidental vocabulary learning. The students were allowed 10 minutes to finish the test. Adjacent students received different versions of the same test.

Step 4 After the collection of the delayed vocabulary test sheets, the test booklets for receptive vocabulary size were handed out, which contained 5 word-frequency levels (2,000, 3,000, 5,000, the University Word List, the 10,000 words). It took no more than 20 minutes.

4.6 Scoring

Gains in students' vocabulary knowledge were measured using VKS. The possible scores for a word on this instrument and their relationship to the categories are given in Figure 3.

Categories	possible scores	Meaning of scores
① →	0	The word is not familiar at all.
② →	1	The word is familiar but its meaning is not known.
③ →	2	A correct synonym or translation is given.
④ →	3	The word is used with semantic appropriateness and grammatical accuracy in a sentence.

Figure 3. Adapted VKS scoring categories and meaning of scores

As is illustrated in Figure 3, answers on the vocabulary scale receive scores ranging from 0 to 3. A four-point scale is chosen to allow for partial word knowledge. And VKS scoring accepts word knowledge of categories (1) and (2) for scores of 0 and 1, and requires a demonstration of knowledge for higher scores. A score of 2 indicates that appropriate synonym or translation has been given for category (3). A score of 3 for Category (4) reflects both semantically and grammatically correct use of the target word, even if the other part of the sentence contains errors.

A wrong answer or an answer that does not comply with the context in Categories (3) or (4) will lead to a score of 1. A Score of 2 is given in Category (4), if the word is used demonstrating the learner's knowledge of meaning in that context but with inaccurate grammar (e.g. a target noun query used as a verb, "She queries me some questions").

5. Results

5.1 Impact of Glossing Types on IVL

Table1 and Table 2 present the descriptive statistics for the impact of three gloss types on IVL immediate vocabulary test and delayed vocabulary test.

As these tables indicate, participants in both glossing conditions significantly outperformed no glossing conditions in both immediate and delayed vocabulary tests. Among the glossing conditions, the mean scores for the MC glossing condition are higher than that of SG condition.

Table 1. Glossing effect: descriptives of the immediate vocabulary test

Group	Number	Mean	Std. Deviation
MC Glosses	30	21.90	6.150
SG Glosses	30	17.27	5.146
No Glosses	29	10.59	4.625

Table 2. Glossing effect: descriptives of the delayed vocabulary test

Group	Number	Mean	Std. Deviation
MC Glosses	26	15.04	4.40
SG Glosses	25	10.36	3.99
No Glosses	28	5.07	3.04

To compare the effects of different glossing types on incidental vocabulary learning, we chose one-way ANOVA to analyze the results. Regarding the immediate vocabulary test, one-way ANOVA indicated a statistically significant difference among the groups, $F_{(2, 86)} = 33.24$, $p < .001$. That is to say, students in both multiple-choice glossing group and single glossing group gain significantly deeper word knowledge than the control group. Most important of all, students in multiple-choice glossing group performed much better than those in single-glossing group. Therefore, significant mean difference, which is 4.63, $p= .001$, exists between Group 1 and Group 2. Likewise, the analysis of the delayed vocabulary test indicated statistically significant differences among the groups, $F_{(2, 76)} = 45.72$, $p =. 000< .001$. That is to say, the multiple-choice group still performed best in the retention in the delayed vocabulary post-test, even if all the three groups lost IVL gains, after 7 days.

5.2 Effect of Vocabulary Size on IVL

In terms of effects of vocabulary size on IVL, independent samples T-test is used here to present the effect of vocabulary size upon IVL gains and retentions.

According to learners' performance on receptive vocabulary levels test, all the participants were divided into two categories: students with high vocabulary level (for the number of their correct answers >57.01 mean); students with low vocabulary level (for the number of their correct answers <57.01 mean). Thereby, 38 participants were labeled as students with low vocabulary level, and 41 participants were students with high vocabulary level.

Tables 3 and 4 present descriptive statistics for the vocabulary size on IVL in immediate and delayed vocabulary tests. As Table 3 indicates, in immediate vocabulary test, the mean score for 41 students with high vocabulary level is higher than that for 38 students with low vocabulary level. Regarding the retention of IVL in the delayed vocabulary post-test, the mean score for the high vocabulary level participants is also higher than that for the low vocabulary level participants. In addition, the delayed post-test also shows that both groups lost some IVL gains.

In order to test the mean difference between the two groups is obviously significant or not, independent samples t-tests were run to investigate each group's vocabulary size performance from immediate test to delayed tests. Both high group learners' immediate scores, $t= 5.78$, $p=.000< 0.05$, and delayed scores, $t=3.42$, $p =.001< 0.05$, were found to be significantly higher than those of low group learners', which means that, the effects of vocabulary size on IVL gains are obviously significant. That is to say, the high vocabulary level students gained much more than the low vocabulary level ones in IVL. Namely, there exist significant differences between the effect of small vocabulary on IVL retentions and that of large vocabulary on IVA retentions.

Table 3. Vocabulary size effect: group statistics of the immediate vocabulary test

Vocabulary level	Number	Mean	Std. Deviation	Std. Error Mean
High	41	20.02	6.49	1.00
Low	38	12.37	5.20	.84

Table 4. Vocabulary size effect: group statistics of the delayed vocabulary test

Vocabulary level	Number	Mean	Std. Deviation	Std. Error Mean
High	41	11.98	5.66	.8838
Low	38	7.92	4.80	.7797

5.3 Receptive Vocabulary Size

The above-listed tables indicate that, the learner's vocabulary size played an important role on IVL gains and retentions. Now two questions may be raised:

(i) Why was not the test for receptive vocabulary size conducted prior to the reading plus vocabulary tests?

ii) Were the groups balanced on vocabulary size? If not, the validity of the conclusion would be in doubt that the glossing types had effect on IVA.

As to the first question, we administered it technically in this order. Considering the nature of incidental vocabulary learning, we put the test for receptive vocabulary size at the end of reading tests. If we had initially conducted the test for receptive vocabulary size, this would have induced participants to sense that the following reading tests were related to vocabulary, and participants would focus their attention on target words. Thus, the performance of participants would go against the nature of IVL. That is, to divert their attention from vocabulary, we conducted the test for receptive vocabulary size at the end, then contributing to ensure the nature of IVL.

As for the second question, the following tables show the balance of the three groups on vocabulary size. Table 5 is descriptive statistics of group initial balance on vocabulary size made by one-way ANOVA.

One-way ANOVA indicated that statistically there is no significant difference between the three groups. (F= .164, p= .849>.05). Now we can conclude that the three groups were balanced initially on receptive vocabulary size. A further analysis of their interrelationship will be displayed by the univariate ANOVA (See Table 7; Table 8).

Table 5. Descriptives of group initial balance on vocabulary size

Group	Number	Mean	Std. Deviation	Sta. Error
MG Glosses	26	56.38	9.667	1.896
SG Glosses	25	57.80	8.062	1.612
No Glosses	28	56.89	8.942	1.690

5.4 Interaction between Glossing Type and Vocabulary Size on IVL

From above-mentioned analysis, we get a detailed and separate view of how the two factors (the glossing types and vocabulary size) influence the IVL gains and retentions. However, the interactions should not be ignored. Therefore, the following three tables present the interaction among factors and the intercept effect on IVL gains and retentions.

Table 6 shows the univariate analysis of variance (ANOVA), with glossing group and vocabulary level as the between-subject factors. Table 7 shows the effects between-participants in the immediate vocabulary post-test. It reveals that, a significant glossing group effect (F=78.253; p= .000) was obtained and, obviously, a significant vocabulary level effect (F =107.205; p = .000), as well as a significant 'interaction of glossing group x vocabulary level (F = 3.732; p= .029). The same analyses were conducted on the data of delayed vocabulary post-test (in Table 8), not yielding completely identical pattern of results. The significant glossing group and vocabulary level effects were found, whereas the interaction of glossing groups and vocabulary level was found

not significant (F, = 1,356; p=.264). The insignificant group x vocabulary level interaction may have resulted from ten students who did not do the delayed test (for we were not to inform the participants of the test).

Table 6. Between-subjects factors

Value Label			
GROUP	1	MC	26
	2	SG	25
	3	C	28
LEVEL	1	H	41
	2	L	38

Table 7. Tests of between-subjects effects: dependent variable-IMMEDIATE

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	3033.784	5	606.76	52.11	.000
intercept	20861.810	1	20861.81	1791.69	.000
Group	1822.302	2	911.15	78.25	.000
LEVEL	1248.258	1	1248.26	107.21	.000
GROUP*LEVEL	86.904	2	43.45	3.73	.029
Error	849.988	73	11.64		
Total	24981.000	79			
Corrected Total	3883.772	78			

Note: a R Square = .781 (Adjusted R Square =.766).

Table 8. Tests of between-subjects effects: dependent variable-DELAYED

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1734.09	5	346.82	34.88	.000
intercept	7871.83	1	7871.83	791.67	.000
Group	1394.117	2	697.06	70.10	.000
LEVEL	370.294	1	370.29	37.24	.000
GROUP*LEVEL	26.962	2	13.48	1.36	.264
Error	725.860	73	9.94		
Total	10400.00	79			
Corrected Total	2459.949	78			

6. Discussion

The first research question of the current study concerned which gloss type - multiple-choice glosses, L1 single-gloss, and no gloss (text only), - best facilitates incidental vocabulary learning. As the results indicated, MC Glosses in a text have a more significant impact on the gains and retention of IVL than single-gloss, which is in line with Hulstijn's findings (Hulstijn, 1992; Laufer & Hulstijn: 2001) that inferred meanings are more likely to be retained than meanings provided by 11 single-gloss.

The practice of multiple-choice glossing, virtually, is a kind of combination of the advantages of inferring and

glosses. Compared with single-glossing, multiple-choice glossing, on the one hand can reduce difficulties presented by insufficient context as well as the possibility of incorrect inferences; on the other hand, can require to pay more attention to target words and to make greater effort to interpret them. Take MC group, for instance. To select a correct interpretation among the three alternatives, the participants in this were encouraged to pass back and forth between target words glosses, and to invest some degree of mental effort and attention, whereas the single-gloss group was provided with the correct interpretation only. Apparently, the multiple-choice procedure required active involvement from the students, thereby increasing their lexical processing, while the single-gloss procedure made the students more passively receptive. Therefore, compared with single-gloss, multiple-choice glosses induced a deeper level of processing, and consequently enhancing subsequent more word recall and retention.

Multiple-choice glosses, a kind of salience enhancement of unknown words, draw learners' attention to target words, supporting the notion of "consciousness-raising" and "input enhancement". When processing multiple-choice glosses, learners first allocate attention to the search for meaning by referring to the glossing options. Next, they evaluate the different meanings and make a decision as to which fits the TW context best. In other words, multiple-choice glosses trigger a search for lexical meaning and an interaction between the various glossing options and the context provided by the reading text. The use of multiple-choice glosses may increase the chances of establishing form-meaning connections as compared to single L I gloss or text only, resulting in lexical acquisition and retention, which was proved in above-listed results. Therefore, it is safe to say that the performance on multiple-choice glosses is a conscious decision making process.

As mentioned above, multiple-choice glosses induce some degree of mental effort and attention. This kind of effort, virtually, also involves some degree of *search* and *evaluation,* the major components in the construct of *Involvement Load.* Based on it, the words that are processed with higher involvement load will be retained better than words that are processed with lower involvement load.

Since the Involvement is easy to observe and to operationalize, we compare the two different glossing types under the same direction in the form of *Task-Induced Involvement Load* (Laufer & Hulstijn, 2001). Let's look at the three basic components of the Involvement: *need, search and evaluation* respectively. In MC group, the participants are asked to read a passage and write out the content afterwards under the multiple-choice glossing conditions. The task in the MC group induces a moderate *need* (imposed by the teacher), strong *search* and strong *evaluation* because the target words are glossed in three options (one correct *option,* two distractors), and learners have to search the right option, that is, the meaning of unknown words, and evaluate it against the other two options and TW context, according to Laufer and Hulstijn (2001), the task in MC group can be described in terms of an involvement index, where the absence of a factor is marked as 0, a moderate presence of a factor as I, and a strong presence as 2, the involvement index of the task is 5 (1 + 2 + 2). Whereas, in the task under the single glossing conditions, the meaning of an unknown word is provided directly in L1 equivalent, and the learners lot to search strongly and evaluate against the TW context. Its involvement index is 2 (I + 1+ 0). Hence, MC group induces a greater involvement load than the SG group. These can be summarized in Table 9.

Table 9. Task-induced involvement load

Task	Components of Involvement		
Conditions	Need	Search	Evaluation
Multiple-Choice glosses	+	++	++/+(depending on subjects and options)
Single glossing	+	+	---

Note: absence(-), moderate(+), strong(++).

From the above involvement table, we can see that the major difference between the two conditions lies in absence of the evaluation in the SG condition, and the presence of evaluation in the MC condition. In both conditions there is a moderate need, induced by the researcher.

Virtually, IVL probabilities are subject to numerous factors, among which, vocabulary size is one that cannot be ignored. With respect to the second research question, the findings in the study provide the evidence that the learners with large vocabulary size demonstrate much more significant gains than those with small vocabulary size in word knowledge in both immediate vocabulary test and delayed test. This is consistent with Sternberg and Powell's findings (1983) that learning vocabulary from context tasks differentiates between able and less

able readers. Because the number of known words surrounding unknown words affects the successful derivation and learning of the unknown words, poor readers, who encounter more unknown words, consequently, will be less able to successfully derive and learn the meanings of unknown words, whereas, good readers are able to learn more new words.

Another issue which merits attention is the superiority of VKS over other vocabulary test. As tables in this chapter indicate, the present study did yield our expected results. VKS---the measurement for vocabulary gains and retentions adopted in our experiment really proved an effective instrument. To be more specific, the obvious and significant gains and retentions of IVL through reading under multiple-choice glossing conditions, should partially be attributable to using VKS, which is sensitive to capture the small learning increments in IVL. If we took word definition as the measurement in our study, the results of the MC group's performance would not be so obvious, and the scores for SG group and C group would also be reduced accordingly. Studies have shown that incidental vocabulary learning grows incrementally as the word is encountered in different contexts. The effective measurement should embody learner's partial word knowledge. Since it scores partial knowledge and is associated with higher learning outcomes than a dichotomous scoring method, VKS, which embodies learner's overall word knowledge and the scoring of which ranges from 0 to 3, is of significant value to future L2 vocabulary testing.

Another key originality in this research lies in the provision of immediate feedback in response to learners' selection, which, in a sense, facilitates learners' gains and retentions of IVL through reading. Just as Hulstijn's research (1992, Experiment III) shows, although multiple-choice glosses encourage deeper lexical processing, the printed form of multiple-choice glosses cannot correct learners' errors without providing feedback on whether student's selected answer was correct or not. However, this issue was aptly addressed in our study by providing immediate feedback in response to learners' selections.

7. Conclusions

7.1 Findings and Implications

Based on the results of our experiment and discussion, several main findings can be summarized as follows: (1) The provision of any given type of marginal glosses, compared with reading without any gloss, does result in much better IVL and retention, which is consistent with the previous findings that the presence of marginal glosses was shown to enhance vocabulary learning, when compared with the absence of marginal glosses. (2) Students of multiple-choice glosses group do demonstrate greater receptive and productive knowledge than those of single-gloss group. (3) Learners with high vocabulary level gain deeper receptive and productive vocabulary knowledge than those with low vocabulary level.

The research into vocabulary is vital and in many respects necessary antecedent to the effective teaching of vocabulary. The present research on the effects of marginal glosses on IVL has yielded findings that have direct implications for teaching and learning English vocabulary. In what follows, we are to correct some generally held misconceptions with regard to the common practice-provision of LI single-gloss, and offer two pieces of suggestions for language teachers and course designers.

One implication is about learning vocabulary through reading. Language teachers face a dilemma in current language teaching situation. They admit the importance of vocabulary learning in English teaching and want to teach a lot of it to their students. But it is impossible for them to spend most of the class time on vocabulary teaching alone. Therefore, it is of great importance to pursue the possibilities in incidental vocabulary learning in L2 situation. In the literature on FL/L2 instruction, it is a generally accepted principle that extensive L2 reading is conducive to vocabulary enlargement. However, reading for global meaning alone will not work. The enhancement of IVL can be achieved by modified lexical input in the form of reading with the provision of MC glosses, which has been confirmed in the present study.

Moreover, just as the results in the present study indicate, vocabulary size also plays an important role in IVL gains and retentions. Teachers, therefore, should be aware that, especially for low-ability readers, simply wide reading without further aid will not lead automatically to significant gain in vocabulary size. Therefore, care should be taken to help low-ability readers by supplying them with carefully chosen, adapted and glossed texts. Previous research suggests that creating informative contexts has been known to raise the amount of words learned incidentally for all ability levels. Reading provided with multiple-choice glosses followed by immediate feedback is a case in point. Although it is not a guarantee for successful learning of new word meanings, one can at least ensure that low-ability readers are provided with maximum opportunity to derive word meanings by making contexts more transparent.

Another implication is about provision of Multiple-Choice Glosses for texts. Based on above discussion, we can conclude from the present study that provision of multiple-choice glosses for texts and reading passages is of crucial importance in L2 course books and reading materials. Therefore, it is suggested here that more are to be done to make language teachers and learners, and textbook designers in particular, aware of the superiority of multiple-choice glosses over single-gloss, in order to improve the efficiency of vocabulary teaching and learning.

7.2 Limitations and Suggestions for Further Research

There were some limitations to this study. First, the small size of the sample population (N=89) sheds some doubt on the validity of the observed significance. A replication study with a greater number of participants is needed in order to obtain more reliable results.

A more detailed picture of the relationship between MC glossing and IVL has been depicted in this research, but much is yet to be known as to the effects of reappearance or occurrence of target words in IVL. It is universally acknowledged that frequency of occurrence of new words positively establish form-meaning relation of a target word. It is, therefore, necessary to explore the probability of MC glosses compensating for the frequency of occurrence of target words. This would at least partially provide curriculum designers and teachers with a better understanding of what accounts for students' success in acquiring vocabulary. Moreover, VKS, the effective measurement of learners' word knowledge, needs to be further examined on its extension to L2 or FL vocabulary testing.

Acknowledgements

References

Coady, J., & Huckin, T. (1997). *Second language vocabulary acquisition*. Cambridge: Cambridge University Press (CUP).

Craik, F., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology, 104*, 268-294. https://doi.org/10.1037/0096-3445.104.3.268

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671-684. https://doi.org/10.1016/S0022-5371(72)80001-X

Hirsh, D., & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a foreign language, 8*, 689-689.

Huang, S., Willson, V., & Eslami, Z. (2012). The Effects of Task Involvement Load on L2 Incidental Vocabulary Learning: A Meta-Analytic Study. *The Modern Language Journal, 96*(4), 544-557. https://doi.org/10.1111/j.1540-4781.2012.01394.x

Hulstijn, J. (1992). Retention of inferred and given word meanings: Experiments in incidental vocabulary learning. In P. Arnaud, & H. Bejoint (Eds.), *Vocabulary and applied linguistics* (pp. 113-125). London: MacMillan. https://doi.org/10.1007/978-1-349-12396-4_11

Hulstijn, J., Holander, M., & Griedenus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *Modern Language Journal, 80*, 327-339. https://doi.org/10.1111/j.1540-4781.1996.tb01614.x

Hulstijn, J. H., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning, 51*(3), 539-338. https://doi.org/10.1111/0023-8333.00164

Khezrlou, S., Ellis, R., & Sadeghi, K. (2017). Effects of computer-assisted glosses on EFL learners' vocabulary acquisition and reading comprehension in three learning conditions. *System, 65*, 104-116. https://doi.org/10.1016/j.system.2017.01.009

Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. Addison-Wesley Longman Ltd.

Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *Modern Language Journal, 73*(4), 440-463. https://doi.org/10.1111/j.1540-4781.1989.tb05325.x

Laufer, B. (1992). How much lexis is necessary for reading comprehension? In P. J. I. Arnaud, & H. Bejoint (Eds.). *Vocabulary and Applied Linguistics*. London: Macmillan. https://doi.org/10.1007/978-1-349-12396-4_12

Laufer, B. (2005). Focus on form in second language vocabulary learning. In S. H. Foster-Cohen, M. Garcia-Mayo, & J. Cenoz (Eds.), *Eurosla yearbook* (Vol. 5, pp. 223-250). Amsterdam : Benjamins. https://doi.org/10.1075/eurosla.5.11lau

Laufer, B., & Paribakht, T. S. (1998). The relationship between passive and active vocabulary: effects of language learning context. *Language learning, 48*(3), 365-391. https://doi.org/10.1111/0023-8333.00046

Laufer, B., & Hulstijn. J. H. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics, 22*(1), 1-26. https://doi.org/10.1093/applin/22.1.1

Nagata, N. (1999). The effectiveness of computer-assisted interactive glosses. *Foreign language Annals, 32*(4), 469-479. https://doi.org/10.1111/j.1944-9720.1999.tb00876.x

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. New York: Newbury House Publishers.

Nation, P., & Waring, R. (1997). Vocabulary size, text coverage and word lists. In N. Schmitt, & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 6-19).

Paribakht, T., & Wesche, M. (1997). Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition. In. J. Coady, & T. Huckin (Eds.), *Second language vocabulary acquisition*. (pp. 174-200). Cambridge: Cambridge University Press.

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. https://doi.org/10.1111/j.1467-9922.2009.00502.x

Rassaei, E. (2017). Computer-mediated textual and audio glosses, perceptual style and L2 vocabulary learning. *Language Teaching Research*. https://doi.org/10.1177/1362168817690183

Read, J. (2004). Research in teaching vocabulary. *Annual Review of Applied Linguistics, 24*, 146-161. https://doi.org/10.1017/S0267190504000078

Reynolds, B. L. (2014). Evidence for the task-induced involvement construct in incidental vocabulary acquisition through digital gaming. *The Language Learning Journal*, 1-19.

Rob, W. (2002). Scales of Vocabulary Knowledge in Second Language Vocabulary Assessment Appeared in Kiyo, The occasional papers of Notre Dame Seishin University.

Sternberg, R. J., & Powell, J. S. (1983). Comprehending verbal comprehension. *American psychologist, 38*(8), 878. https://doi.org/10.1037/0003-066X.38.8.878

Sun, C. H. (2017). The value of picture-book reading-based collaborative output activities for vocabulary retention. *Language Teaching Research, 21*(1), 96-117. https://doi.org/10.1177/1362168816655364

Teng, F. (2017). The Effects of Task-induced Involvement Load on Word Learning and Confidence Judgments Mediated by Knowledge and Regulation of Cognition. *Educational Sciences: Theory & Practice, 17*(3). https://doi.org/10.12738/estp.2017.3.0167

Watanabe, Y. (1997). Input, intake, and retention: Effects of increased processing on incidental learning of foreign language vocabulary. *Studies in Second Language Acquisition, 19*(3), 287-307. https://doi.org/10.1017/S027226319700301X

Wesche, M., & Paribakht, T. (1996). Assessing vocabulary knowledge depth vs. breadth. *Canadian Modern Language Review, 53*(1), 13-40.