

## *The effects of L1 and L2 e-glosses on incidental vocabulary learning of junior high-school English students\**

SI-MIN HU, VIPHAVEE VONGPUMIVITCH AND  
JASON S. CHANG

*National Tsing Hua University, Taiwan*

*(email: martinastarson@yahoo.com.tw; viphavee@mx.nthu.edu.tw;  
jason.jschang@gmail.com)*

HSIEN-CHIN LIOU

*Feng Chia University, Taiwan*

*(email: hcliou@fcu.edu.tw)*

---

### Abstract

While researchers have examined the effectiveness of various online gloss types on incidental L2 vocabulary learning, little research on online gloss languages has been conducted. Previous attempts which compared the effects of L1 and L2 glosses have reported mixed results. To fill the gaps, this study examined the effectiveness of Chinese and English e-glosses on incidental English vocabulary learning on a less-researched student group in CALL – junior high-school English-as-a-foreign-language (EFL) students. Seventy-eight students with Chinese as their first language read two online passages with either Chinese (L1) or English (L2) glosses. They were divided into four treatment groups: (1) high-proficiency students receiving L1 gloss before L2 gloss ( $n = 19$ ), (2) high-proficiency students receiving L2 gloss before L1 gloss ( $n = 19$ ), (3) low-proficiency students receiving L1 gloss before L2 gloss ( $n = 20$ ), and (4) low-proficiency students receiving L2 gloss before L1 gloss ( $n = 20$ ). After reading, the students were assessed with a vocabulary test which contained a definition-supply part and a cloze part serving as both post-tests and delayed post-tests. Repeated-measures analyses of variance were utilized to analyze the score data. Significant differences were found not only among the four groups but also between the two post-tests. Overall the high proficiency groups performed better in the post-tests, but the high proficiency group who received English glosses remembered more words in the delayed post-test than the high proficiency group who received Chinese glosses. The results show that as learners' proficiency increases, they may be able to make better use of the L2 conceptual links for word retention and learning. The conclusions support the usefulness of both Chinese and English e-glosses which should be selected based on learners' proficiency level.

Keywords: gloss, vocabulary learning, retention, gloss language effects, junior high-school students

---

\* The paper is partially funded by NSC99-2411-H007-052-MY3.

## 1. Introduction

This study examined the effectiveness of learners' use of their first language (L1) and the target or second language (L2) glosses on incidental vocabulary learning and retention in an L2 online reading environment in an EFL junior high-school context in Taiwan. The term *incidental* is defined by Hulstijn (1992) as learners' acquisition of the meanings of words when they engage in other tasks such as the comprehension of reading and listening passages. In an EFL classroom, teachers face limited instruction time and thus cannot spend much effort on directly teaching students new words. This pedagogical reality makes it reasonable to suggest that incidental vocabulary learning through reading may play a crucial role in students' language acquisition. However, incidental vocabulary learning does have limitations. First, contextual information is often ambiguous and not sufficiently reliable for learners to make the correct inference. Second, learners run the risk of failing to verify the correctness of inferences and may therefore learn words incorrectly. Glosses are proposed by previous research to be one of the solutions to aid reading comprehension and enhance incidental vocabulary learning. In CALL environments, various issues concerning the effectiveness of glosses on incidental vocabulary learning have been examined, ranging from whether or not glosses are useful to which type of gloss is more effective, for example., CALL-based or paper-based, L1 or L2 (Abraham, 2008; Chen, 2002; Davis & Lyman-Hager, 1997; Hulstijn, Hollander & Greidanus, 1996; Jacobs, Dufon & Hong, 1994; Ko, 2005; Lyman-Hager, Davis, Burnett & Chennault, 1993; Miyasako, 2002; Taylor, 2006, 2009; Watanabe, 1997; Yoshii, 2006). The current study aims to address specifically which online gloss, in the first language or the target language, is more helpful for learners of different proficiency levels by examining a less researched student sample – junior high-school EFL learners.

## 2. Literature review

In the past few years, several studies have been conducted to examine the effectiveness of computer-mediated (CALL) glosses. Three meta-analysis projects (Abraham, 2008; Taylor, 2006, 2009), in particular, have examined the effect sizes of studies contrasting CALL glosses with paper-based ones. Taylor (2006) analyzed nineteen studies which compared the efficiency of traditional L1 glosses with that of CALL L1 glosses in L2 reading and found that learners provided with CALL glosses understood significantly more text than those learners with paper-based glosses. Abraham (2008) surveyed eleven experimental studies in order to compare the effects of L2 learners' access to CALL glosses (L1 or L1 plus L2) to those without such access. Similar to the findings of Taylor (2006), Abraham's results showed that CALL glosses had an overall medium effect on learners' reading comprehension and a large positive effect on incidental vocabulary learning. To confirm the results from Taylor (2006) and Abraham (2008), Taylor (2009) conducted another meta-analysis project with 32 studies which revealed that the overall effect sizes were larger for CALL glossing studies than for non-CALL glossing studies on reading. In addition to these meta-analysis projects, single studies such as Davis and Lyman-Hager (1997) and

Lyman-Hager, Davis, Burnett and Chennault (1993) all support the efficiency of CALL-based glossing environments for reading. The next question seems to ask precisely which type of gloss is more effective for which type of learners to assist CALL designers or teachers who want to choose an appropriate gloss for their learners.

One type of e-gloss design question is that of which language to use, either the learner's first language or their target language. Previous studies have shown mixed results concerning the use of L1 or L2 glosses and their effects on both reading comprehension and vocabulary learning. Jacobs, Dufon, and Hong (1994) contrasted three reading conditions: (1) reading with no gloss, (2) reading with English (L1) glosses, and (3) reading with Spanish (L2) glosses. The results showed better performance of both L1 and L2 gloss types over the non-gloss type. However, no significant difference was found between L1 and L2 glosses. Chen (2002) examined the effects of L1 and L2 glosses with 85 college EFL freshmen in Taiwan. The participants were divided into three groups: (1) L1 (Chinese) gloss, (2) L2 (English) gloss, and (3) no gloss. They read a 193-word English text, with twenty target words being glossed. Similar to Jacobs *et al.*, Chen's results also showed no significant difference between L1 and L2 glosses.

Other studies have shown the advantage of one type of glossing language over the other type. Ko (2005) investigated how different types of gloss conditions (L1 or L2) affected Korean college students' reading comprehension. One hundred and six Korean undergraduates read a 931-word article with 24 target words under one of three conditions: (1) L1 glosses, (2) L2 glosses, and (3) no gloss. The results indicated that only L2 glosses showed any effectiveness. Moreover, 62% of the learners favored L2 glosses for their reading material. Another study indicating the advantages of L2 glosses over L1 glosses was conducted by Miyasako (2002). One hundred and eighty-seven Japanese senior high school students participated in the study. The results showed that glossing languages had a relationship with the students' English ability: L2 glossing was more effective for higher-level learners while lower-level learners benefited more from L1 glossing.

While Ko (2005) and Miyasako (2002) demonstrated the superiority of L2 glosses over L1 glosses for reading comprehension, Watanabe (1997) and Yoshii (2006) confirmed the advantages of L1 glosses for incidental vocabulary learning. In Watanabe's study (1997), 24 adult ESL students from different nationalities in the USA were asked to read a 463-word text with fifteen target words. With regard to the relationship between glossing languages and English ability, the L1 group did better than the L2 group for the lower-level learners, whereas for the higher-proficiency learners the result was reversed.

Inspired by Ko (2005), Miyasako (2002), and Watanabe (1997), Yoshii (2006) continued searching for more empirical evidence on the effectiveness of L1 and L2 glosses on incidental English vocabulary learning in a multimedia environment. The study included the effect of additional pictorial cues in L1 and L2 glosses, and investigated how these additions affected vocabulary learning. One hundred and ninety-five Japanese university students read texts with different gloss types: (1) L1 text only, (2) L1 text plus picture, (3) L2 text only, and (4) L2 text plus picture. The overall results indicated significant differences between picture (text-plus-picture) and no-picture (text only) glosses. Between L1 and L2 glosses, however, no significant difference was shown. In spite of this, the L1 text only group was found to be able to

sustain their scores best. In order to explore further the better vocabulary retention that the L1 gloss group showed, Yoshii inspected the participants and discovered that they were low-intermediate or intermediate learners who were still in the early stages of English learning.

To explain his findings, Yoshii cited Kroll and Stewart's (1994) *Revised Hierarchical Model*, which proposes a developmental shift from a word association model to a concept mediation model as L2 proficiency increases. Kroll and Stewart explain that L2 learners rely on L2 word-to-L1 word links (lexical links) in the early stages, but as their L2 proficiency develops, they link L2 directly to concepts (conceptual links). Through this model, one could expect that the L1 glosses would be more effective for vocabulary learning than L2 glosses, for learners with lower proficiency, since the conceptual connections at this learning stage are stronger for L1 than for L2. On the contrary, learners with higher proficiency are expected to benefit more from L2 glosses since at their proficiency level L2 conceptual links should be stronger than L2 word-to-L1 word connections. This model explains well the findings in Yoshii's (2006) study, and recently it was also supported by Kroll, Green, Tokowicz, and Van Hell (2010).

On the whole, the relationship identified between gloss languages and learners' proficiency indicated the need to take learners' proficiency levels into consideration. As text-based glosses are more easily implemented and less intrusive to the comprehension processes in reading or listening than multimedia ones (Mohsen & Balakumar, 2011) and thus are commonly available and accessible in websites, the current study focuses on different gloss languages and an under-represented learner group – junior high-school EFL students. The current study addresses two research questions:

1. Do the effects of L1 and L2 glosses differ in incidental vocabulary learning?
2. Does the factor of learners' proficiency levels (low and high) influence the effects of L1 and L2 glosses on incidental vocabulary learning?

### **3. Methodology**

#### ***3.1 Reading materials***

A pilot study was conducted in order to test whether some pre-chosen reading materials (from Hill, 1994) were appropriate for the main study. Two experienced local English teachers (including the first author) selected three short articles as candidates for the study based on their more than ten years of teaching experience in junior high schools. The selection criteria for the target articles were as follows:

1. The rate of unknown words in the chosen article must be between 3% and 5%, no matter whether they are for high-proficiency students or low-proficiency students. This criterion is based on the literature which suggests that a person can read and understand a text independently with knowledge of about 95% of its vocabulary (Laufer, 1997; Nation, 2001; Read, 2000).
2. After being glossed, the chosen articles had to be mostly comprehensible to the students, especially for the low-proficiency students, as effective glossing should aid reading comprehension.

Table 1 *Basic Information of Three Candidate Articles*

Article	length	glossing	Flesch reading ease	Flesch-Kincaid Grade level
A	361	5%	82.5	5.4
B	354	5%	84.2	5.0
C	381	5%	81.7	5.2

A: Asking the Time (adapted from Hill, 1994)

B: The Only Girl (adapted from Hill, 1992)

C: Beginning to Enjoy the Violin (adapted from Hill, 1992)

The Microsoft Word program was used to calculate readability scores of the three articles by giving an ease score and a grade level. Taking into consideration the average sentence length and the average number of syllables per word, the Flesch Reading Ease score rates the articles on a 100-point scale, with higher scores indicating easier texts, and the Flesch-Kincaid Grade Level score rates the articles on a US grade-school level. Results showed that all three articles were easy and an average fifth grader in the US should be able to understand all three articles (see Table 1). To pilot test the three articles, six students with different English proficiency levels who were not involved in the main experiment but who had the same background as those participants were instructed to read all three articles in the paper version. They also circled the words they did not understand, and expressed how much they understood the articles after they had been offered with Chinese glosses. Based on the students' performance, Articles A and C were chosen as the articles for the treatment and were modified to contain only about 5% unknown words (Article A and its glosses are shown in Appendix A). To limit the variables affecting the results of this study, only unknown content words were glossed. Altogether, 33 content words were marked as unknown by the pilot-study participants and were therefore treated as the glossed target words in Articles A and C.

### 3.2 *Participants*

Seventy-eight ninth-grade students in six classes from a public junior high school in Taiwan participated in this study. Originally, the study started with 198 ninth-graders. However, 51 students were excluded because they either did not have full attendance or had difficulty understanding the 35 keywords that were included in the L2 glosses in this study (this set of 35 keywords used in the L2 glosses was different from the 33 glossed target words which appeared in the reading texts). The remaining students were grouped into three proficiency levels based on their scores in the latest intra-school English test, which had a total score of 80. Because the participants' intra-school test scores were not normally distributed, the median and quartiles rather than the mean were used to aid the grouping of students. Thirty-eight students who were at or above the 75th quartile (obtaining a score higher than 62) were regarded as high-level learners, whereas 40 students who were at or below the 25th quartile (obtaining a score lower than 37) were identified as low-level learners. The remaining 69 students whose scores fell between 37 and 62 were considered

intermediate-level and were excluded from the study in order to contrast the difference between learners' proficiency levels.

The 78 participants were classified into four groups to reflect the two independent variables in this study: the two gloss languages (Chinese and English), and the students' English proficiency levels (low and high). The first group, "Low L2→L1" (n = 20), was a group of low-level students who read Article A with English glosses (L2) in the first experiment, then read Article C with Chinese glosses (L1) in the second experiment. The second group, "Low L1→L2" (n = 20), was a group of low-level students who read Article A with Chinese glosses (L1) before reading Article C with English glosses (L2) in the second experiment. The third group, "High L2→L1" (n = 19), was a group of high-level students who received the same treatment as group 1. And the fourth group, "High L1→L2" (n = 19), was a group of high-level students who received the same treatment as group 2. The dependent variable was students' scores measured by the immediate and delayed vocabulary post-tests.

### 3.3 Instruments

**3.3.1 Pre-tests.** A pre-test was designed to make two contributions. On the one hand, the results of the pre-test were intended to enhance the validity of the study by excluding those participants who had difficulty reading the L2 glosses (e.g., glossing words in English) offered in this study. On the other hand, the results of the pre-test ensured the equivalence of the groups' previous knowledge of the target words. Only the words which could not be recognized by any of the participants were regarded as target words and included in the post-tests. To achieve these goals, the pre-test had two parts. The first part included 35 key words and phrases which were important for readers to understand the L2 glosses in this study. The second part contained 33 words taken from Article A and Article C which were marked as unknown words by the students in the pilot study. The items in the pre-test were designed for the participants to check each word in the word list to see whether or not they could recognize it and provide a short written explanation in Chinese if they knew its meaning. A sampled pre-test item is like this:

英文單字 English word	不認識請打 X (Put a X, if you do not know the word)	若認識請寫出字義 (If you know the word, write its meaning in either Chinese or English.)
<b>had</b>		

Note: The English translation was provided under the original Chinese instruction

The pre-test result showed that out of the 33 target words, 24 words from the two articles (12 in each) were unknown to all participants. Thus, these 24 content words were regarded as target words in the two online reading experiments.

**3.3.2 Post-tests.** The participants received two kinds of post-tests: the immediate post-test and the delayed post-test. Each post-test contained two groups of test items: the definition-supply items and the cloze items, each with twelve items respectively

(see Appendix B for an example). The format of the definition-supply test was designed similarly to the format used in the pre-test (shown above). The cloze test provided the contexts and glosses as hints with the target word's beginning and ending letters for participants to spell the word out. A sample of the cloze item is shown below:

- 請依括弧中的提示，拼寫出在文章中出現的單字(注意時態要正確)。  
(Spell the word in the correct form according to the context and the hints)
1. Bill p\_\_\_\_\_d (**had/擁有**) a small shop on a street corner in town.

The definition supply part is operationalized as a recognition ability of the word sense in terms of vocabulary knowledge and skill, whereas the cloze part is a productive capacity in relation to word form. The delayed post-test was exactly the same as the immediate post-test but the order of the test items was different.

### 3.4 The treatment

After the articles for the current study were chosen, the treatment, namely online reading with the aid of glosses, was conducted in a computer lab. There were two online experiments in the treatment. The second experiment came one week after the first one. The rationale behind the two experiments was to ensure the credibility of the results derived from the first experiment. Because the duration of each online experiment only took about fifteen minutes, the result would be more convincing if the two experiments yielded the same outcome.

In each experiment, the participants were instructed to read on the computer screens a chosen article which included 5% of the words highlighted and glossed. Figure 1 shows how the glosses were presented on the screen. As the students clicked on a highlighted word, a gloss would appear above the word. It is important to highlight here that the glosses are hidden behind each word on the screen while each learner goes through the words in the passage. Only when they clicked on a specific

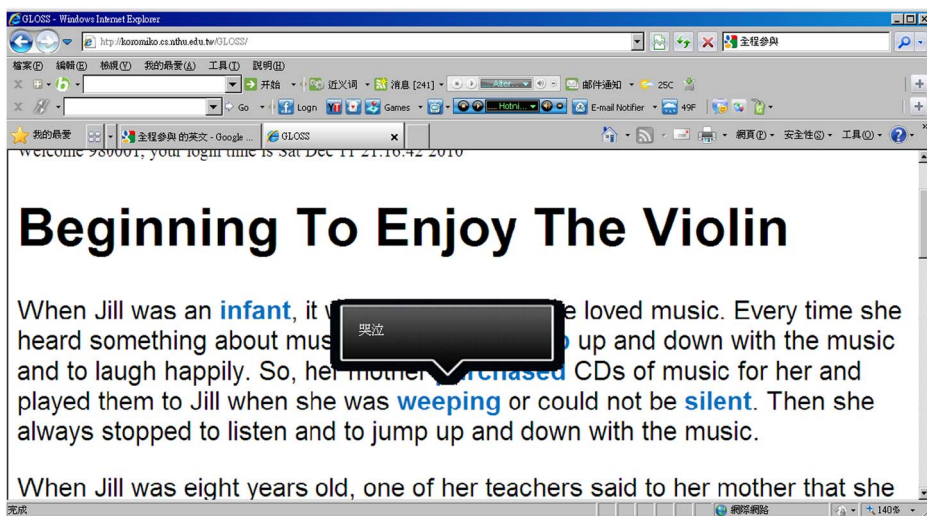


Fig. 1. An Example of Computer-enhanced Glossing in the Treatment.

word would the English or Chinese gloss of that word appear. Thus, the gloss showed up only upon the user's request. This method of online glossing is different from the traditional method in which a fixed set of either Chinese or English glosses is always presented with the reading passage regardless of learners' needs.

A management module was designed with a database system that kept track of the gloss languages each participant made use of, and the frequency with which each particular glossed word was clicked on when all participants were working online. The program was written in Django and SQLite3. It ran on a windows server, with a 64 bit CPU and 1 Gigabyte RAM. Students had a PC Windows XP client when the experiments were running and they used Internet Explorer to access the system.

### 3.5 Procedure

Following Yoshii (2006), this study took place over a two-week period. The participants took the pre-test one week prior to the treatment without knowing that they would be given any vocabulary tests in advance. The post-test was conducted immediately after the treatment and the delayed post-test came two weeks later. This study took four class periods (each lasting 40 minutes) with the instructors' permission: the first for explanation and the pre-test, the second for the first experiment and the immediate post-test, the third for the second experiment and the second immediate post-test, and the fourth for the delayed post-test (with test items of the two immediate post-tests in the two experiments). The answers in the pre- and post-tests were counted as correct if the students gave at least one answer for a word with multiple meanings. In the cloze post-test, however, the students got the right answer only when they spelled the whole target word correctly. The students received one point for each correct answer. Two raters, both native Chinese speakers who are junior high school English teachers, examined the answers. Repeated measures analyses of variance (ANOVA) were conducted on the definition-supply and the cloze post-tests. When the results showed significant differences among these tests, *post-hoc* analyses further examined the differences across the four groups.

## 4. Results

### 4.1 Definition-supply post-tests

*4.1.1 The first experiment.* Table 2 presents descriptive statistics for the definition-supply items of the post-tests in the two experiments and also the delayed post-test. The mean scores of all four groups were low in general (the full score of each test was 12). In the delayed post-test, all four groups suffered from a serious loss of retention and the mean scores are much lower than those in the immediate post-test.

We then used repeated measures ANOVA to examine the differences. The results showed significant differences not only between the two post-tests ( $F = 257.82$ ,  $df = 1$ ,  $p < .01$ ), but also among the four groups ( $F = 20.17$ ,  $df = 3$ ,  $p < .01$ ) (see Table 3). Because there was also a significant interaction effect between the post-tests and the groups ( $F = 19.68$ ,  $df = 3$ ,  $p < .01$ ), *post-hoc* analyses were needed to further examine the interaction effect.



Table 2 Descriptive Statistical Results of Definition-supply Posttests in the Two Experiments

Group		Def Immediate 1	Def Delayed 1	Def Immediate 2	Def Delayed 2
High L1 → L2 (n = 19)	M	<b>6.63</b>	M <b>0.89</b>	M 8.21	M 2.16
	SD	2.89	SD 1.05	SD 2.84	SD 2.54
	Med	7	Med 1	Med 9	Med 1
	IQR	2 ~ 11	IQR 0 ~ 3	IQR 3 ~ 12	IQR 0 ~ 8
High L2 → L1 (n = 19)	M	<b>6.32</b>	M <b>1.42</b>	M 6.84	M 1.42
	SD	2.96	SD 1.57	SD 3.34	SD 1.68
	Med	6	Med 1	Med 7	Med 1
	IQR	2 ~ 12	IQR 0 ~ 6	IQR 2 ~ 12	IQR 0 ~ 5
Low L1 → L2 (n = 20)	M	<b>2.65</b>	M <b>0.25</b>	M 1.8	M 0.25
	SD	1.81	SD 0.72	SD 1.19	SD 0.44
	Med	3	Med 0	Med 1.5	Med 0
	IQR	0 ~ 5	IQR 0 ~ 3	IQR 0 ~ 4	IQR 0 ~ 1
Low L2 → L1 (n = 20)	M	<b>1.75</b>	M <b>0.25</b>	M 2.85	M 0.45
	SD	1.07	SD 0.55	SD 1.63	SD 0.75
	Med	1.5	Med 0	Med 2.5	Med 0
	IQR	0 ~ 4	IQR 0 ~ 2	IQR 1 ~ 5	IQR 0 ~ 2

Note: IQR = Inter-quartile Range

The total score = 12, one point for each correct answer.

Med: median

First, a one-way ANOVA was conducted to compare the four groups for the definition-supply “immediate” post-test in the first experiment. As reported in Table 2, the *High L1 → L2* group had the highest mean score (6.63), followed by the *High L2 → L2* group (6.32), the *Low L1 → L2* group (2.65), and the *Low L2 → L1* group (1.75). In other words, simply looking at the means, L1 gloss seemed to work better for both the high and low groups. A significant difference was found among the performance of the four groups in the definition-supply immediate post-test ( $F = 22.92$ ,  $df = 3$ ,  $p < .01$ ; Table 5). Tukey HSD and Scheffe tests were then used to examine the differences across the pairs. We found that the means of the two high groups were actually not statistically different from each other ( $p = 0.97$ ), nor were the means of the two low groups ( $p = 0.61 > 0.01$ ). However, both high groups performed better than both low groups (mean differences ranging from 3.66 to 4.88,  $p = .00$ ). In other words, there did not seem to be a “gloss-language” effect because the high groups, regardless of which glosses they received first, always performed better than the low groups. Instead, we could only see a “proficiency” effect.

Second, a one-way ANOVA was done to compare the four groups for definition-supply “delayed” post-test in the first experiment. As reported in Table 2, the *High L2 → L1* group had the highest mean score (1.42), followed by the *High L1 → L2* group (0.89). The two low groups had the same mean score (0.25). Again, a significant difference was found among the performance of the four groups in the definition-supply delayed post-test ( $F = 5.77$ ,  $df = 3$ ,  $p < .01$ ). The Tukey HSD and Scheffe *post-hoc* tests did not show the two high groups to differ from each other, and only the *High L2 → L1* group had a significantly higher mean than the two low

Table 3 Repeated Measures ANOVA to Compare the 4 Groups for Definition-supply Immediate and Delayed Posttests Scores in the First Experiment

Tests of Within-Subjects Effects						
Source		Type III Sum of Squares	df	Mean Square	F	Sig.
<b>Posttests 1</b>	Sphericity Assumed	514.38	1	514.38	<b>257.82</b>	<b>0.00*</b>
<b>Posttests1 * group</b>	Sphericity Assumed	117.80	3	39.27	<b>19.68</b>	<b>0.00*</b>
Error (Def1)	Sphericity Assumed	147.64	74	2.00		
Tests of Between-Subjects Effects						
Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept		990.32	1	990.32	225.41	0.00*
<b>Group</b>		265.86	3	88.62	<b>20.17</b>	<b>0.00*</b>
Error		325.11	74	4.39		

Note: \*  $p < .01$

groups (mean difference = 1.17,  $p = .00$ ). Thus, it seemed that the *High L2→L1* group retained the words better than all other groups. Accordingly, perhaps we could say that there was a “gloss-language” effect for the high group; that is, L2 glosses seemed to help the high level learners maintain words in their memory better than L1 glosses.

*4.1.2 Results of the second experiment.* The same analyses were conducted on the data collected in the second experiment in which students who received L1 glosses in the first experiment now received L2 glosses and vice versa. The repeated measures ANOVA revealed significant differences not only between the two post-tests ( $F = 343.97$ ,  $df = 1$ ,  $p < .01$ ) but also among the four groups ( $F = 23.53$ ,  $df = 3$ ,  $p < .01$ ), with a significant interaction effect between the post-tests and the groups ( $F = 28.37$ ,  $df = 3$ ,  $p < .01$ ). These were followed by *post-hoc* analyses. A significant difference was found among the performance of the four groups in the definition-supply immediate post-test ( $F = 32.39$ ,  $df = 3$ ,  $p < .05$ ). Tukey HSD and Scheffe tests yielded the same findings as those derived from the definition-supply immediate post-tests in the first experiment. That is, the two high groups and the two low groups did not differ from each other, but both high groups performed better than both low groups (mean differences ranging from 7.02 to 3.99). Thus, there did not seem to be a “gloss-language” effect. Instead, we could only see the “proficiency” effect. This finding is consistent with the pattern in the first experiment.

A significant difference was also found in the delayed post-test among the performance of the four groups ( $F = 6.23$ ,  $df = 3$ ,  $p < .05$ ). The Tukey HSD and Scheffe *post-hoc* tests revealed that the two high groups still did not differ from each other, and nor did the two low groups. But only the *High L1→L2* group was significantly better than the two low groups in the delayed post-test. Recall that in the first experiment’s delayed post-test, this group (the *High L1→L2*) did not have a significantly higher mean than the two low groups. Instead, it was the *High L2→L1* group who did so. But now that the *High L1→L2* group received L2 glosses in the second experiment, their delayed post-test mean was significantly better than the two low groups while the *High L2→L1* group’s was not. Thus, it seems that whoever received L2 glosses in the experiment would perform better than the two low groups in the delayed post-test. We can therefore say with more certainty that there was a “gloss-language” effect for the high groups. L2 glosses seemed to help the high level learners maintain words in their memory better than L1 glosses in this study.

## 4.2 The cloze post-tests

In Table 4, the means of the cloze tests of all four groups were lower, compared with those of the definition-supply tests. This is because cloze tests required not only word recognition but also productive knowledge of the words which included spellings. Thus, it was not surprising that the students would find the cloze more challenging than the definition supply tests. The same statistical analyses were conducted on the cloze data as we did on the definition-supply data. Although the means were lower, significant differences were still found not only between the two post-tests in the first cloze experiment ( $F = 30.43$ ,  $df = 1$ ,  $p < .01$ ) but also among

Table 4 One-way ANOVA to Compare the 4 Groups for Definition-supply Immediate Posttest in the First Experiment

Tests of Between-Subjects Effects						
Dependent Variable: Definition-Supply Immediate 1						
Source	Type III Sum of Squares	Df	Mean Square	F	Sig.	
Corrected Model	364.969a	3	121.66	22.92	0.00*	
Intercept	1466.08	1	1466.08	276.18	0.00*	
<b>Group</b>	364.97	3	121.66	<b>22.92</b>	<b>0.00*</b>	
Error	392.83	74	5.31			
Total	2188.00	78				
Corrected Total	757.79	77				

Note: \*  $p < .01$

the four groups ( $F = 23.04$ ,  $df = 3$ ,  $p < .01$ ). *Post-hoc* analyses confirmed the same pattern of only the “proficiency” effect but no “gloss-language” effects. In other words, the two high groups performed significantly better than the two low groups. This is the same result as that derived from the definition-supply immediate post-tests in both experiments. The delayed cloze post-tests’ results ( $F = 16.45$ ,  $df = 3$ ,  $p < .05$ ) also showed exactly the same pattern.

The second cloze experiment showed the same results as the first. Repeated-measures ANOVA revealed significant differences not only between the two post-tests ( $F = 49.82$ ,  $df = 1$ ,  $p < .01$ ) but also among the four groups ( $F = 21.07$ ,  $df = 3$ ,  $p < .01$ ). *Post-hoc* analyses showed different performances across proficiency levels but not between the two groups of the same proficiency level. The delayed post-test also showed exactly the same pattern ( $F = 12.78$ ,  $df = 3$ ,  $p < .05$ ).

To sum up, throughout the study we consistently found “proficiency” effect rather than “gloss-language” effect; that is, in both the immediate and delayed post-tests, the high proficiency groups always performed better than the low groups regardless of the gloss language. In the delayed post-tests of the definition-supply items, the high group that received L2 glosses did significantly better than the low groups, while the high group that received L1 glosses did not. However, in the delayed post-tests of the cloze tests, both high groups did significantly better than the low groups. It seemed that L2 glosses helped the high level students retain words better in their memory for the definition-supply items, but this gloss-language effect was not evident for the cloze items where both high groups were better than the two low groups.

## 5. Discussion

In this study, we see from the descriptive statistics for both definition-supply and cloze post-tests that the mean scores of all four groups were low in general. This means that even after the learners had read two articles with the help of online L1/L2 glosses, the vocabulary gains as measured in the definition-supply or the cloze items

Table 5 *Post-hoc Analysis Using Tukey HSD and Scheffe Tests to Examine the Interaction Between Groups and Definition-supply Immediate Posttest Scores in the First Experiment*

Multiple Comparisons							
Dependent Variable: Definition-Supply Immediate 1							
	(I) groupmem	(J) groupmem	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Tukey HSD	<u>Low L2 -&gt; L1</u>	Low L1 -> L2	-0.90	0.73	0.61	-2.82	1.02
		<u>High L1 -&gt; L2</u>	-4.8816*	0.74	<b>0.00*</b>	-6.82	-2.94
		<u>High L2 -&gt; L1</u>	-4.5658*	0.74	<b>0.00*</b>	-6.51	-2.63
	<u>Low L1 -&gt; L2</u>	Low L2 -> L1	0.90	0.73	0.61	-1.02	2.82
		<u>High L1 -&gt; L2</u>	-3.9816*	0.74	<b>0.00*</b>	-5.92	-2.04
		<u>High L2 -&gt; L1</u>	-3.6658*	0.74	<b>0.00*</b>	-5.61	-1.73
	<u>High L1 -&gt; L2</u>	<u>Low L2 -&gt; L1</u>	4.8816*	0.74	<b>0.00*</b>	2.94	6.82
		<u>Low L1 -&gt; L2</u>	3.9816*	0.74	<b>0.00*</b>	2.04	5.92
		High L2 -> L1	0.32	0.75	0.97	-1.65	2.28
	<u>High L2 -&gt; L1</u>	<u>Low L2 -&gt; L1</u>	4.5658*	0.74	<b>0.00*</b>	2.63	6.51
		<u>Low L1 -&gt; L2</u>	3.6658*	0.74	<b>0.00*</b>	1.73	5.61
		High L1 -> L2	-0.32	0.75	0.97	-2.28	1.65
Scheffe	Low L2 -> L1	Low L1 -> L2	-0.90	0.73	0.68	-2.98	1.18
		High L1 -> L2	-4.8816*	0.74	0.00*	-6.99	-2.77
		High L2 -> L1	-4.5658*	0.74	0.00*	-6.68	-2.45
	Low L1 -> L2	Low L2 -> L1	0.90	0.73	0.68	-1.18	2.98
		High L1 -> L2	-3.9816*	0.74	0.00*	-6.09	-1.87
		High L2 -> L1	-3.6658*	0.74	0.00*	-5.78	-1.55

Table 5 *Continued*

High L1 → L2	Low L2 → L1	4.8816*	0.74	0.00*	2.77	6.99
	Low L1 → L2	3.9816*	0.74	0.00*	1.87	6.09
High L2 → L1	High L2 → L1	0.32	0.75	0.98	-1.82	2.45
	Low L2 → L1	4.5658*	0.74	0.00*	2.45	6.68
	Low L1 → L2	3.6658*	0.74	0.00*	1.55	5.78
	High L1 → L2	-0.32	0.75	0.98	-2.45	1.82

Note: \*  $p < .01$

were still quite low. This might be related to the nature of incidental learning. Learners pick up meanings incidentally as they encounter words. However, the depth of learning is shallow and thus fragile as time goes by (Nation, 2001; Yoshii, 2006). Even for the L1 glossing group, it is still difficult to make conceptual links between the unknown target words and the explanations provided by the Chinese glosses in such a short period of time, given their limited first exposure to the words through online reading. EFL junior high-school students in Taiwan tend to take close reading and memorization of vocabulary items as study skills for English, and more on-task time is thus required for them to synthesize the knowledge into their language system.

With regard to the differential effects of L1 and L2 glosses on incidental vocabulary learning, if we look at the descriptive statistics of the two groups that have the same proficiency level, the high-level learners almost always performed better with the help of L2 glosses while the low level learners consistently performed better when utilizing L1 glosses. However, despite such a pattern from the descriptive statistics, the groups with the same proficiency level were not statistically different. Instead, ANOVA results only showed “proficiency” effects rather than “gloss-language” effects in all post-tests. That is, based on the ANOVA results, high level learners always got higher scores than low level learners no matter whether they used L1 or L2 glosses. Thus, learners with a better English level to start with performed consistently better in both the recognition task (the definition supply items) and the productive vocabulary task (the cloze items), compared with those with a weaker English level.

In spite of this, *post-hoc* analyses still disclosed a “gloss-language” effect in the definition-supply delayed post-tests in both experiments; that is, high level learners were found to retain the words better whenever they read with L2 glosses. These results, however, did not apply to the cloze items which required productive vocabulary ability rather than simply word recognition. The English glosses thus seemed to be more effective for high-level learners in word retention than the Chinese glosses did, if the assessment targets recognition of word sense only and not productive ability. In this case, the effects of L1 and L2 glosses would differ, depending on the proficiency of the learners as well as the type of the tests (e.g., the kind of vocabulary knowledge required) for long-term retention of words.

The results of this study supported what has been found in previous studies (Jacobs *et al.*, 1994; Ko, 2005; Miyasako, 2002) and confirmed Kroll and Stewart’s *Revised Hierarchical Model* (1994) which claimed the effectiveness of L2 glosses in higher level participants, for at their proficiency level, L2 conceptual links should be stronger than L2 word-to-L1 word connections. For low-level learners, however, their performance was the same no matter which gloss languages they received and whether they were tested immediately after reading or after a week.

Taking both the gloss languages and the learner proficiency levels into consideration, it seems that there might be a threshold level for either online gloss language (Chinese or English) as a help option to take effect for incidental vocabulary learning when learners are at different stages. For the seventy-eight EFL junior high students in this study, the Chinese online gloss would still be of help when they were at the beginning stage of learning the English language. However, for students whose language proficiency is moving towards intermediate and high-intermediate levels, they could benefit much more from English glosses for their vocabulary learning while reading.

## 6. Conclusion

This study examined the effectiveness of different gloss languages, L1 or L2, on incidental vocabulary learning with a particular focus on the influence of junior high school learners' proficiency levels. Four groups of junior high-school EFL students in Taiwan representing high and low proficiency level groups read online texts with either Chinese or English glosses and performed two tasks – definition-supply and cloze. The four groups were found to perform significantly differently. Regardless of the gloss language, the high groups performed better than the low groups both immediately and one week after reading. Further comparisons indicated the effectiveness of L2 glosses for high level learners in word retention for the definition-supply items, which supported the *Revised Hierarchical Model* (Kroll & Stewart, 1994) which claims that L1 conceptual links are stronger than L2 conceptual links in the early stages of L2 acquisition (i.e., lower proficiency level), and that a direct link to concepts from L2 words can be possible only with increasing proficiency (e.g., higher proficiency level). In other words, the Chinese conceptual links help more than the English ones for the weaker students (if they read the texts with the Chinese glosses); the English gloss help only works while reading online when the students' English becomes better.

The results of this study have implications for vocabulary learning and teaching. First, e-glosses are useful whether in Chinese or English for enhancing vocabulary learning, and we should continue to utilize e-glosses to help students understand reading materials. Second, the effects of Chinese and English glosses may depend on junior high school learners' English proficiency level. To meet individual learners' needs, therefore, both L1 and L2 glosses should be offered for students to choose from in CALL environments. In Asian EFL settings such as Taiwan, students rely heavily on their L1 reference resources; in view of the findings of this study, instructors can encourage more proficient learners to use English glosses to achieve longer-term effects and also encourage most learners to try English annotations as their proficiency level progresses beyond the beginning level.

Attention must be drawn to some limitations. First, this study conducted the online experiment twice to ensure the reliability of the results. While we did our best to control the readability of the two reading materials as well as to restrict the target words in the two experiments to content words, we failed to limit the difficulty of the target words in the two experiments to the same degree. As a result, a future study needs to control for this factor in order to better achieve the goal of the repeated treatment. Second, regarding the delayed post-test, we originally intended to use a two-week span following similar studies (Chun & Plass, 1996; Kost, Foss & Lenzini, 1999; Yoshii & Flaitz, 2002; Yoshii, 2006). But because of time limitations, the delayed post-test for the second experiment was combined with the delayed post-test for the first experiment. That is, the delayed post-test for the second experiment was conducted only one week after the treatment, which could have a negative influence on the design of the repeated treatment as well as the result of the study.

Despite its limitation, the significance of the study lies in the fact that a CALL-enhanced gloss reading program can provide additional input for EFL junior high students to boost their learning of vocabulary, even when the gain is limited



(considering the online time was only 30 minutes in total). More importantly, different gloss languages are shown to have different impacts on learners of different proficiency levels. Future studies can continue to examine the conditions in which the L1 or L2 glosses benefit different types of learners.

### Acknowledgements

Thanks go to the involved participants and Huang Sheng-Ting, who wrote the online reading-gloss program under the guidance of Prof. Jason Chang. Special thanks also go to the anonymous reviewers who provided comments on improving earlier versions of this article.

### References

- Abraham, L. B. (2008) Computer-mediated glosses in second language reading comprehension and vocabulary learning: A meta-analysis. *Computer Assisted Language Learning*, **21**(3): 199–226.
- Chen, H. (2002). *Investigating the effects of L1 and L2 glosses on foreign language reading comprehension and vocabulary retention*. Paper presented at the annual meeting of the Computer-Assisted Language Instruction Consortium, Davis, CA.
- Chun, D. M. and Plass, J. L. (1996) Effects of multimedia annotations on vocabulary acquisition. *The Modern Language Journal*, **80**(2): 183–198.
- Davis, J. N. and Lyman-Hager, M. (1997) Computers and L2 reading: Student performance, student attitudes. *Foreign Language Annals*, **30**(1): 58–72.
- Hill, L. A. (1994). *Happy readers 3a*. Efsthadiadis Group S. A.
- Hulstijn, J. H. (1992) Retention of inferred and given word meanings: Experiments in incidental vocabulary learning. In: Arnaud, P. J. and Bejoint, H. (eds.), *Vocabulary and applied linguistics*. London: Macmillan, 113–125.
- Hulstijn, J. H., Hollander, M. and Greidanus, T. (1996) Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *The Modern Language Journal*, **80**(3): 327–339.
- Jacobs, G., Dufon, P. and Hong, F. C. (1994) L1 and L2 glosses in reading passages: Their effectiveness for increasing comprehension and vocabulary knowledge. *Journal of Research in Reading*, **17**(1): 19–28.
- Ko, M. H. (2005) Glosses, comprehension, and strategy use. *Reading in a Foreign Language*, **17**(2): 125–143.
- Kost, C. R., Foss, P. and Lenzini, J. J. (1999) Textual and pictorial glosses: Effectiveness of incidental vocabulary growth when reading in a foreign language. *Foreign Language Annals*, **32**(1): 89–113.
- Kroll, J. F., Green, D. W., Tokowicz, N. and Van Hell, J. G. (2010) The revised hierarchical model: A critical review and assessment. *Bilingualism: Language and Cognition*, **13**(3): 373–381.
- Kroll, J. F. and Stewart, E. (1994) Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, **33**(2): 149–174.
- Laufer, B. (1997) The lexical plight in second language reading: Words you don't know, words you think you know, and words you can't guess. In: Coady, J. and Huckin, T. (eds.), *Second language vocabulary acquisition: A rationale for pedagogy*. New York: Cambridge University Press, 20–34.

- Lyman-Hager, M., Davis, N., Burnett, J. and Chennault, R. (1993) *Us Vie de Boy: Interactive reading in French*. In: Borchardt, F. L. and Johnson, E. M. T. (eds.), *Proceedings of CALICO 1993 Annual Symposium on Assessment*. Durham, NC: Duke University, 93–97.
- Miyasako, N. (2002) Does text-glossing have any effects on incidental vocabulary learning through reading for Japanese senior high school students? *Language Education & Technology*, **39**: 1–20.
- Mohsen, M. A. and Balakumar, M. (2011) A review of multimedia glosses and their effects on L2 vocabulary acquisition in CALL literature. *ReCALL*, **23**: 135–159.
- Nation, P. (2001) *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Read, J. (2000) *Assessing Vocabulary*. Cambridge: Cambridge University Press.
- Taylor, A. M. (2006) The effects of CALL versus traditional L1 glosses on L2 reading comprehension. *CALICO Journal*, **23**(2): 309–318.
- Taylor, A. M. (2009) CALL-based versus Paper-based glosses: Is there a difference in reading comprehension? *CALICO Journal*, **27**(1): 147–160.
- Watanabe, Y. (1997) Effects of single and multiple-choice closes on incidental vocabulary learning. *JACET Bulletin*, **27**: 177–191.
- Yoshii, M. (2006) L1 and L2 glosses: Their effects on incidental vocabulary learning. *Language Learning & Technology*, **10**(3): 85–101.
- Yoshii, M. and Flaitz, J. (2002) Second language incidental vocabulary retention: The effect of picture and annotation types. *CALICO Journal*, **20**(1): 33–58.

## Appendix A

### Reading Materials

#### Article A: Asking the Time (adapted from Hill, 1994)

Bill **possessed** a small shop on a street corner in town, and he **vended** newspapers, comic books, candy, chocolates, and **numerous** other things in that shop. There was a **public telephone** beside his shop, and when people wanted to use the phone, they often came into the shop to **request** small money. Often they ask for the **change** and they will also **purchase** something from him. So, he was happy to have the telephone box so near him. Also, he often had too much small money in his shop, so he was **delighted** to help people in this way.

One evening, a young man came into the shop to ask for small money because he wanted to telephone someone. From the way the man spoke, Bill **realized** that he was an **outlander**. Bill changed the money for him, but the man did not buy anything.

After ten minutes, the man **returned** to Bill's shop.

'Ah,' Bill **considered**, 'he feels that he should buy something from me because I changed his money for him.'

But he was wrong: the man said, 'I'm sorry, but I need some more **change**.' He gave Bill one hundred dollars. Bill **offered** him small money for it and the young man **dashed** out of the shop.

'Well, that is the end of him,' Bill thought, but after ten more minutes the man came back again. 'This time he will buy something for sure,' Bill thought happily.

But he was wrong again. 'I'm sorry to ask you for some more small money,' the young man said, 'but last night I **encountered** a nice girl at a bookstore, and she gave

me her telephone number, but every time I ring it, a **silly** girl is always talking at the other end, and it is not the girl I met last night.'

'Can I see the number you are calling?' Bill said.

The young man opened his notebook and **displayed** the number to Bill. It was 117.

Bill began to laugh, and the young man **inquired**, 'What's the matter? Why are you laughing?'

'Because that is the number you ring when you want to know the time,' Bill **replied**.

### **Glosses: (a sample)**

#### **English**

1. possessed (K1): had
2. vended (off-list): sold
3. numerous (K1): many
4. public telephone (K1): the phone everyone can use
5. request (K2): ask for
6. change (K1): small money

#### **Chinese**

1. possessed: 擁有
2. vended: 販賣
3. numerous: 許多的
4. public telephone: 公共電話
5. request: 索求
6. change: 零錢

\*The word frequency level (K1 = 1 ~ 1000, K2 = 1001 ~ 2000, or academic) was based on **Compleat Lexical Tutor**, <http://www.lextutor.ca/>.

**Appendix B**

*Sample of the Post-tests*

Definition-supply part: General instruction for each word

英文單字 English word	不認識請打 (Put a X, if you do not know the word)	若認識請寫出字義 (If you know the word, write its meaning in either Chinese or English.)
----------------------	---	--

	X	Chinese or English meaning		X	Chinese or English meaning
possess			vend		
numerous			dash		
request			encounter		
purchase			delighted		
inquire			outlander		
display			consider		

Cloze part: 請依括弧中的提示，拼寫出在文章中出現的單字(注意時態要正確)。  
(Spell the word in the correct form according to the context and the hints)

1. Bill p\_\_\_\_\_d (had/擁有) a small shop on a street corner in town.
2. He v\_\_\_\_\_d (sold/販賣) newspapers.
3. He also sold n\_\_\_\_\_s (many/許多的) other things in that shop.
4. When people wanted to use the phone, they often came into the shop to r\_\_\_\_\_t  
(ask for/索求) small money.
5. They will also p\_\_\_\_\_e (buy/購買) something from him.