

# Effect of Glossing on Chinese Senior High School EFL Learners' Vocabulary Acquisition

Yequi Zhu

Department of English, School of Foreign Studies, Nanjing University, China

**Abstract**—This study examined the effects of four types of glosses on Chinese EFL learners' vocabulary acquisition. The participants were 175 senior high school students, who were divided into five groups according to different glossing conditions: L1 glosses, L2 glosses, multiple-choice L1 glosses, multiple-choice L2 glosses, and no glosses. The participants were asked to read an English material that contained ten target pseudo-words. They then took an immediate vocabulary post-test, which consisted of two recognition tasks and two recall tasks. Two weeks later, they took an unexpected delayed vocabulary post-test with the same tasks. The findings revealed that glossing had a significantly positive effect on the learners' vocabulary learning. Single L1 glosses were the most effective among the four types of glosses, while multiple-choice L2 glosses were the least effective. It was also found that the four types of glosses produced better effects on vocabulary acquisition in the recognition tasks than in the recall tasks and that all the gloss groups had better short-term word retention than long-term one, except for the single L2 gloss group, which displayed the opposite effect.

**Index Terms**—gloss, vocabulary acquisition, single glosses, multiple-choice glosses

## I. INTRODUCTION

Glosses have been widely used to assist second-language readers in understanding texts. Typically provided in the margin or at the bottom of text pages, glosses offer language learners the definitions or explanations of unfamiliar words within a text and supposedly facilitate their form-meaning mapping (Nation, 2013). Textual glossing encompasses various types, including single glosses and multiple-choice glosses. Each type has its unique benefits and potential limitations, and their effectiveness can vary depending on the learning context or task type. While single glosses that provide simple definitions or L1 equivalents are common, some scholars (Hulstijn, 1992; Laufer & Hulstijn, 2001) argue that this type of glossing may deter learners from inferring word meanings from context, which is a strategy considered essential for vocabulary acquisition. They believe that meanings inferred from context are better retained than those provided by glosses, as the effortful process of inferring enhances word retention. In response to this argument, multiple-choice glosses have been proposed as an alternative to single glosses (Hulstijn, 1992). However, the findings concerning the effectiveness of different types of glossing remain inconclusive (Yanagisawa et al., 2020). More empirical studies are needed to determine their effects on L2 vocabulary acquisition.

Regarding the population studied in glossing research, Zhang and Ma (2021) pointed out that the effect of glossing among English language learners at the middle school level had received little scholarly attention. This is particularly true in the Chinese context. While glossing is commonly used to facilitate reading comprehension and vocabulary acquisition in Chinese high school English education, few empirical studies have examined its effectiveness and how it influences high school students' word learning.

To gain a deeper understanding of glossing effectiveness for high school students, the study aims to provide a comprehensive comparison of gloss effects by examining the influencing factors such as gloss language (L1 and L2), gloss type (single and multiple-choice glosses), retention phase (immediate and delayed vocabulary post-tests) and task type (recognition and recall tasks). The results of this study would not only contribute to understanding how different types of glossing influence word learning in relation to different influencing factors but also provide insights into optimal gloss designs for lower-proficiency EFL learners.

## II. LITERATURE REVIEW

The relative effectiveness of L1 glosses and L2 glosses has been extensively explored in glossing research, but previous studies comparing their effects have yielded mixed results. Laufer and Shmueli (1997) found that L1 glosses are more effective than L2 glosses in both short-term and long-term word retention. Jacobs et al. (1994) compared the effectiveness of L1 and L2 glosses and found that both L1 and L2 glosses were more effective than the non-gloss condition, but no significant difference was revealed between L1 and L2 glosses. Similar results were obtained by Ko (2012), who investigated the effects of L1 and L2 glosses on L2 word learning by examining 90 Korean English learners' performance on vocabulary tests. It was found that there existed no significant difference between the two types of glosses. Choi (2016) recruited 180 Korean English learners to test the effects of L1 and L2 glosses and found

no significant difference between them in short-term retention of target words, but the L1 group outperformed the L2 group in long-term retention of words that occurred four times in the reading material. Similarly, Yoshii (2006) found no difference between L1 and L2 glosses in terms of incidental vocabulary learning, but the L1 text-only group had better long-term retention. The review of these previous studies reveals that despite extensive research on the effectiveness of L1 and L2 glosses, the findings remain inconclusive and that in-depth explorations are still needed.

Textual glosses can be presented in a single mode (a short definition or explanation) or a multiple one (multiple choices) (Zhang & Ma, 2021). Rott and other researchers investigated the effects of multiple-choice glosses. They found that L1 multiple-choice glosses resulted in significantly deeper productive and receptive word gains (Rott et al., 2002), and compared with the no-gloss condition, glosses could induce learners' search for concrete meaning and form-meaning connections (Rott & William, 2003). However, comparisons between single glosses and multiple-choice glosses have produced mixed results. Some studies found no significant differences between the two types for incidental vocabulary learning (Hulstijn, 1992; Miyasako, 2002; Watanabe, 1997), while others indicated that multiple-choice glosses were more effective than single glosses (Nagata, 1999; Rott, 2005).

Mixed findings have also been reported concerning task type. Yoshii (2006) examined the effectiveness of L1 and L2 glosses in incidental vocabulary learning and found no significant differences between the two in the definition-supply task and the recognition task. Yanagisawa et al. (2020) observed that glosses affected vocabulary gains differently depending on the task type. In their study, the students scored much higher on the recognition test than on the recall test. Çakmak and Erçetin (2018) discovered that glosses facilitated vocabulary recognition and production but had no positive effect on text recall. In contrast, Chun and Plass (1996) found better incidental learning outcomes on recognition tests than on production tests. Zhang and Ma (2021), however, reported that textual glosses were more effective in vocabulary production. The inconsistency across the previous studies suggests that the effect of glosses likely interacts with task type in complex ways that are not fully understood. Further research systematically comparing the effect of glosses across different task types would help us understand the relationship between the gloss effect and task type and clarify the conditions under which glosses are most effective.

In addition, it is worth noting that most previous studies have primarily focused on investigating the effectiveness of gloss in intermediate or advanced English learners' vocabulary acquisition. While these studies have provided valuable insights into our understanding of the gloss effect, there is a notable gap in research concerning lower-level learners, particularly high school students (Zhang & Ma, 2021). High school students represent a significant population in terms of language development. They are at a crucial stage where they are expanding their vocabulary. More studies are needed to gain a deeper understanding of how glossing impacts the vocabulary acquisition of this particular group.

To address these gaps in the glossing literature, this study investigates how gloss language and gloss type jointly influence Chinese senior high school students' L2 vocabulary acquisition. By comparing students' performances in immediate and delayed vocabulary recall and recognition tasks, the effects of single and multiple-choice glosses in L1 and L2 will be analyzed.

### III. METHODOLOGY

#### A. Research Questions

The present study will address the following questions:

- 1) What is the relative effectiveness of different gloss types (single L1 glosses, single L2 glosses, multiple-choice L1 glosses, and multiple-choice L2 glosses) in Chinese senior high school students' incidental English word acquisition?
- 2) How do the effects of these glosses vary according to test timing (immediate or delayed vocabulary post-tests) and task types (recognition or recall tasks)?

#### B. Participants

The participants in the present study were selected from second-year students at a Chinese senior high school. As part of their regular academic practice, these students took an English final examination at the end of each semester. The purpose of this examination was to evaluate their proficiency across different language aspects, including grammar, vocabulary, reading, and writing. To ensure that all participants had a similar level of English proficiency, their average scores from the final English exams over the previous three semesters were calculated. Only students who obtained average scores ranging between 110 and 120 out of 150 were selected to participate in the study. A total of 175 students were chosen as the participants in the present study, who were then divided into five groups. These groups included the single L1 (SL1) group, the single L2 (SL2) group, the multiple-choice L1 (MCL1) group, the multiple-choice L2 (MCL2) group, and the control group that did not receive any form of glossing assistance.

#### C. Target Words

To prevent participants had prior knowledge of the target words and to ensure that they would not encounter the target words after the immediate vocabulary post-test, ten pseudo-words were created to replace ten words in the reading material. The replaced words accounted for approximately 3% of the total words in the text. The proportion of

the unchanged words (around 97%), therefore, fell within the range of familiar words Nation (2013) recommended for effective comprehension, i.e., at least 95% coverage of a written text.

The ten created pseudo-words followed the orthographic and morphological rules that govern the spelling and formation of words in the English language (Pulido, 2007). The selection of the replaced words was based on three criteria: (1) The words chosen for glossing were concrete nouns or verbs that were essential to conveying the main ideas of the reading material; (2) the words could only appear once in the text (to ensure controlled frequency); and (3) the context surrounding each target word allowed learners to infer its meaning, which was checked and confirmed by three Chinese English learners of a similar English proficiency level. Moreover, the pseudo-words' inflections were consistent with those of the replaced words in the reading material. For instance, the pseudo-noun would be pluralized when the original word was in its plural form, and the pseudo-verb would be used in the past tense if the original word called for it.

#### *D. Instruments*

##### *(a). The Reading Comprehension Test*

The reading comprehension test consisted of a reading material, glosses for ten target words, and five multiple-choice reading comprehension questions. The material, which was chosen by two teachers with over 20 years of teaching experience, was 305 words long and selected based on the teachers' assessment of the participants' English proficiency level. In the reading material, except for the ten target words, all the words were among the 3000 most frequent words in the Corpus of Contemporary American English (COCA). The readability of the reading material was validated by three students of a similar English proficiency level, confirming that the material was understandable to the participants.

The reading material and the comprehension questions were the same for all the groups. The only difference between the tests administered to each gloss group was the type of gloss used. The four experimental groups (i.e., the SL1 group, the SL2 group, the MCL1 group, and the MCL2 group) each read the material with one type of gloss, and the control group received the material without any glosses. The purpose of the reading comprehension test was to see whether the participants had read and understood the material.

##### *(b). The Vocabulary Post-Test*

The vocabulary post-test in this study consisted of two recall tasks and two recognition tasks. The first recall task required the participants to provide the Chinese definitions or English synonyms of the target words, and the second one to make a sentence with each target pseudo-word. In the recognition tasks, the participants needed to match the Chinese definition with the target words and then choose the correct Chinese meaning of each target word from two options. The vocabulary post-test was administered twice, immediately after the reading comprehension test and two weeks after the first post-test. The immediate and delayed post-tests were used to assess the participants' retention of the target words. The delayed post-test was identical to the immediate post-test, except that the order of the test items was different.

#### *E. The Pilot Study*

A pilot study was conducted to ensure the suitability of the reading material and to determine the appropriate time limits for the tests the participants were about to take. As a year-two English teacher at the senior high school suggested, the participants were given 15 minutes to read the material and complete the subsequent reading comprehension test. Another 15 minutes was given to complete the immediate vocabulary post-test. Following the pilot study, an interview was conducted to gather the students' opinions regarding the reading material and the post-test. The interview was recorded with the consent of the participants and lasted approximately 25 minutes. Based on the findings of the pilot study, it was determined that the selected reading material was suitable for the students, and 15 minutes was an appropriate time limit for each task.

#### *F. Data Collection*

The participants were given 15 minutes to read the material and complete a reading comprehension test. This was followed by a 15-minute immediate vocabulary post-test. Two weeks later, the participants were asked to complete an unexpected delayed vocabulary post-test. Before the experiment began, the researcher gave a brief oral introduction to the participants regarding the reading task. The gloss groups were informed that they could refer to glosses during the reading process if they encountered any unknown words. However, they were not told that they would be given glosses of different types.

As previously mentioned, both the immediate and delayed vocabulary post-tests included recall and recognition tasks. The recall tasks were conducted before the recognition tasks in both the immediate and delayed vocabulary post-tests to prevent the recognition test items from affecting the validity of the recall test.

#### *G. Data Analysis*

Once the data were collected, we checked the participants' answers and gave one point for each correct answer. The vocabulary post-test scores were entered into SPSS v29 for quantitative analysis. Descriptive statistics of mean scores and standard deviations were generated for a general comparison. A mixed ANOVA was performed to compare the mean scores for the five gloss conditions in relation to task type (the recall task and the recognition task) and test timing

(the immediate post-test and the delayed post-test) and to find out the variables' main effect and interaction effect. One-way ANOVA tests were run to examine the differences among the five groups in terms of short-term and long-term retention in the recall tasks and the recognition tasks. Post-hoc analysis was used to examine differences between each pair of the groups.

#### IV. RESULTS AND DISCUSSION

##### A. Results

A 2×2×5 mixed ANOVA was conducted to analyze the immediate and delayed vocabulary test scores. The results (Table 1) showed significant main effects of task types ( $F(1, 170)=583.684, p<.001$ ) and of test timing ( $F(1, 170)=207.504, p<.001$ ). The analysis also showed significant interaction effects of task types and gloss conditions ( $F(4, 170)=31.988, p<.001$ ), of test timing and gloss conditions ( $F(4, 170)=13.121, p<.001$ ) and of task types, test timing and gloss conditions ( $F(4, 170)=30.104, p<.001$ ). No significant interaction effect was observed between task types and test timing ( $F(1, 170)=2.629, p=.107$ ). The overall effect sizes were high except for the interaction effect of test timing and task types.

TABLE 1  
MANOVA OF 2 (TASK TYPE) × 2 (TEST TIMING) × 5 (GLOSS CONDITION) DESIGN

	<i>F</i>	Sig.	Partial Squared	Eta
<b>Task type</b>	583.684	<.001	.774	
<b>Test timing</b>	207.504	<.001	.550	
<b>Task type × Gloss condition</b>	31.988	<.001	.429	
<b>Test timing × Gloss condition</b>	13.121	<.001	.236	
<b>Task type × Test timing</b>	2.629	.107	.015	
<b>Task type × Test timing × Gloss</b>	30.104	<.001	.415	

TABLE 2  
DIFFERENCES WITHIN GLOSS CONDITIONS BY TEST TIMING IN THE RECALL/RECOGNITION TASK

	<b>Task type</b>			
	<b>The recall task</b>		<b>The recognition task</b>	
	<i>F</i>	Sig.	<i>F</i>	Sig.
<b>L1 gloss</b>	8.24	.005	104.71	.000
<b>L2 gloss</b>	7.24	.008	8.63	.004
<b>MCL1 gloss</b>	48.96	.000	11.00	.001
<b>MCL2 gloss</b>	48.96	.000	22.44	.000
<b>No gloss</b>	129.80	.000	.57	.449

TABLE 3  
DIFFERENCES WITHIN GLOSS CONDITIONS BY TASK TYPE ON THE IMMEDIATE/DELAYED VOCABULARY POST-TEST

	<b>Time of the vocabulary post-tests</b>			
	<b>Immediate</b>		<b>Delayed</b>	
	<i>F</i>	Sig.	<i>F</i>	Sig.
<b>L1 gloss</b>	7.31	.008	240.20	.008
<b>L2 gloss</b>	93.19	.000	674.46	.000
<b>MCL1 gloss</b>	32.66	.000	18.74	.000
<b>MCL2 gloss</b>	32.66	.000	41.25	.000
<b>No gloss</b>	90.23	.000	.31	.577

As task types, test timing and gloss conditions had a significant interaction effect, simple effect tests were run to examine 1) the differences between the immediate post-test and delayed post-test within each task type in all the five gloss conditions and 2) the differences between the recognition tasks and the recall tasks within the testing time in all the five gloss conditions. The results showed that within each gloss condition, there existed significant differences between the immediate and delayed post-test scores for the recall and recognition tasks except for the recognition tasks in the no-gloss condition (Table 2). Regarding the effect of task types, significant differences between the task types were observed within the immediate vocabulary post-test and also the delayed vocabulary post-test in all five gloss conditions except the delayed post-test in the no-gloss condition (Table 3).

Table 4 reports the means and standard deviations of the participants' word gains in each gloss condition. In the recall tasks of the immediate post-test, the SL1 group got the highest mean score (8.11), followed by the MCL1 group (Mean=5.17), the SL2 group (Mean=2.51) and the MCL2 group (Mean=1.77). The control group's mean score (0.23) was the lowest. In the recognition tasks of the immediate post-test, the same mean score sequence was observed, i.e., the SL1 group (9.11), the MCL1 group (Mean=7.29), the SL2 group (Mean=6.09) and the MCL2 group (Mean=3.89) and the control group (Mean=3.74). In the recall tasks of the delayed post-test, the mean score of the SL1 group (5.03) was the highest, which was followed by those of the MCL1 group (4.17), the SL2 group (1.63), the MCL2 group (0.34). The control group's mean score was 0. In the recognition tasks of the delayed post-test, the SL1 group again got the

highest mean score (8.20), which was followed by the SL2 group (Mean=6.94), the MCL1 group (Mean=5.06), the MCL2 group (Mean=1.66) and the control group (Mean=0.11).

TABLE 4  
MEANS AND STANDARD DEVIATIONS OF WORD GAINS IN THE IMMEDIATE AND DELAYED POST-TESTS

	SL1*		SL2		MCL1		MCL2		No gloss	
	recall**	recog	recall	recog	recall	recog	recall	recog	recall	recog
<b>Immediate test</b>	8.11 (1.78)	9.11 (1.08)	2.51 (1.84)	6.09 (1.65)	5.17 (2.20)	7.29 (1.67)	1.77 (1.29)	3.89 (1.69)	0.23 (0.49)	3.74 (1.52)
<b>Delayed test</b>	5.03 (0.92)	8.20 (1.18)	1.63 (1.06)	6.94 (1.33)	4.17 (1.22)	5.06 (1.11)	0.34 (0.59)	1.66 (1.21)	0 (0)	0.11 (0.32)

Note.  $N = 35$ .

\* SL1 = the single L1 gloss, SL2 = the single L2 gloss, MCL1 = the multiple-choice L1 gloss, MCL2 = the multiple-choice L2 gloss

\*\* recall = the recall task, recog = the recognition task

A one-way ANOVA was performed to examine the differences among the five groups regarding the immediate word gains in the recall and recognition tasks. The results show that the five groups in the recall and the recognition tasks differed significantly from each other in their immediate vocabulary post-test (recall:  $F(4, 170) = 127.568, p = .000$ ; recognition:  $F(4, 170) = 77.097, p = .000$ ). The mean scores on the recognition tests were all higher than those on the recall tests. Given the significant differences, the participants generally did better in the recognition tasks than in the recall test in terms of their immediate word gains. Post hoc analyses (Table 5) revealed significant differences among the groups in terms of short-term retention. Specifically, in the recall tasks, all the groups exhibited significant differences between each other, except for the SL2 and MCL2 groups ( $p = .429$ ). In the recognition tasks, significant differences were observed among all the groups except for the MCL2 and the control groups ( $p = .699$ ). A one-way ANOVA was also run to see the differences among the five groups in terms of long-term word retention in the recall and recognition tasks. Similar results were obtained in that the five groups were also significantly different from one other in their delayed vocabulary post-tests in both of the two tasks (recall:  $F(4, 170) = 234.533, p = .000$ ; recognition:  $F(4, 170) = 346.412, p = .000$ ). Post hoc analyses (Table 6) revealed that all the groups showed significant differences from each other on the delayed post-test in the recognition tasks and the recall tasks ( $p < .05$ ).

TABLE 5  
POST-HOC ANALYSIS RESULTS OF THE FIVE GLOSS CONDITIONS ON THE IMMEDIATE POST-TEST

	the recall task		the recognition task	
	Dif.	<i>P</i>	Dif.	<i>P</i>
<b>SL1 vs. SL2</b>	5.60	.000	3.03	.000
<b>SL1 vs. MCL1</b>	2.94	.000	1.83	.000
<b>SL1 vs. MCL2</b>	6.34	.000	5.23	.000
<b>SL1 vs. No gloss</b>	7.89	.000	5.37	.000
<b>SL2 vs. MCL1</b>	-2.66	.000	-1.20	.001
<b>SL2 vs. MCL2</b>	0.74	.429	2.20	.000
<b>SL2 vs. No gloss</b>	2.29	.000	2.34	.000
<b>MCL1 vs. MCL2</b>	3.40	.000	3.40	.000
<b>MCL1 vs. No gloss</b>	4.94	.000	3.54	.000
<b>MCL2 vs. No gloss</b>	1.54	.000	0.14	.699

TABLE 6  
POST-HOC ANALYSIS RESULTS OF THE FIVE GLOSS CONDITIONS ON THE DELAYED POST-TEST

	the recall task		the recognition task	
	Dif.	<i>P</i>	Dif.	<i>P</i>
<b>SL1 vs. SL2</b>	3.40	.000	1.26	.001
<b>SL1 vs. MCL1</b>	0.86	.015	3.14	.000
<b>SL1 vs. MCL2</b>	4.69	.000	6.54	.000
<b>SL1 vs. No gloss</b>	5.03	.000	8.09	.000
<b>SL2 vs. MCL1</b>	-2.54	.000	1.89	.000
<b>SL2 vs. MCL2</b>	1.29	.000	5.29	.000
<b>SL2 vs. No gloss</b>	1.63	.000	6.83	.000
<b>MCL1 vs. MCL2</b>	3.83	.000	3.40	.000
<b>MCL1 vs. No gloss</b>	4.17	.000	4.94	.000
<b>MCL2 vs. No gloss</b>	0.34	.016	1.54	.000

## B. Discussion

### (a). Effect of Glossing in General

The present study found that glossing significantly affected Chinese senior high school students' incidental English vocabulary acquisition. Regardless of the type of glossing used, all gloss groups outperformed the control group (no glosses) in both the immediate and the delayed vocabulary post-tests and in both recall and recognition tasks, as evidenced by the significant interaction effects observed for task types and gloss conditions, test timing and gloss

conditions, as well as task types, test timing and gloss conditions. Glossing in reading materials plays an important role in enhancing the salience of words in the text, effectively capturing readers' attention and directing their focus toward unfamiliar words (Nation, 2013). This helps learners understand the meanings of words they encounter while reading, which can contribute to stronger form-meaning mappings and better word retention.

Among the four types of glossing, SL1 glosses are the most effective in both recognition and recall tasks for short-term and long-term retention, while MCL2 glosses are the least effective. This could be due to the low English proficiency level of the participants. The participants in this study were senior high school students whose English proficiency level was not high enough to understand L2 gloss definitions effectively or to make a correct choice from multiple-choice gloss options. Lower-proficiency learners tend to establish a direct one-to-one relationship between words in their native language (L1) and their second language (L2). In a sense, the L1 assumes a significant role in processing unfamiliar words, and single glosses in their native language facilitate their English word learning. Considering this cognitive reliance on the L1, providing lower-proficiency learners with single glosses in their native language can be an effective way to assist their English word learning.

The present study showed that MCL glossing, whether in Chinese or English, was less effective than SL1. Although multiple-choice glosses are supposed to increase learners' cognitive involvement in word learning and will consequently result in better word retention (Laufer & Hulstijn, 2001), lower-proficiency learners face challenges in making accurate inferences of word meanings due to their limited English proficiency, which results in incorrect form-meaning connections. This, however, does not mean multiple choice glossing is ineffective in English word learning. It proves to be significantly more effective than the no-gloss condition. Furthermore, MCL1 glosses were found generally more effective in word learning than SL2 except in the recognition tasks of the delayed post-test. Multiple-choice glosses entail relatively more cognitive involvement load in the acquisition of unfamiliar words (Hulstijn, 1992; Miyasako, 2002), as they encourage learners to search in the context for the meanings of the words and evaluate their inferences based on contextual clues. According to the Involvement Load Hypothesis (Laufer & Hulstijn, 2001), search and evaluation are the two cognitive factors that contribute to learners' successful word learning. This increased mental effort leads to better acquisition effects. Compared with the ones in L2, multiple-choice glosses presented in lower-proficiency learners' native language can effectively facilitate their word learning. However, it is undeniable that the choices made by lower-proficiency learners are sometimes incorrect, which explains why the effectiveness of multiple-choice glosses is significantly weaker than that of SL1 glosses.

*(b). Effect of Glossing in Relation to Test Timing*

Regarding short-term and long-term retention, this study revealed the significant main effect of test timing and the significant interaction effect of test timing and gloss conditions. All gloss groups experienced a significant decrease in mean scores from the immediate to the delayed post-test except the SL2 group. This suggested that some of the gained vocabulary knowledge had been lost over time, which is generally consistent with the findings of previous studies (Cheng & Good, 2009; Choi, 2016; Ko, 2012).

It was also found that different from the other glosses, SL2 glosses had a better effect on long-term word retention than on short-term word retention. This may be attributed to the combined impact of learners' English proficiency level and the appropriate cognitive load induced by this type of glossing. The Involvement Load Hypothesis (Laufer & Hulstijn, 2001) argues that higher involvement loads lead to better learning outcomes. However, word retention may not solely depend on how much mental effort learners invest in task completion, and this can be particularly true with low-proficiency learners. In terms of cognitive involvement, SL1 glosses require the lowest as learners' native language is used to provide direct and immediate access to the meaning of the words. The low cognitive involvement load may result in weaker memory traces that fade more quickly over time. In the case of MCL1 glosses, though learners must choose from the given options, the options are provided in the learners' native language, which allows for quick association of the words with their meanings, and facilitate immediate recognition and understanding of the words. In a sense, the mental effort involved in the word decoding process is still relatively low. MCL2 glosses pose the greatest challenge for learners, as evidenced by the lowest mean post-test scores. Learners must choose the correct option that defines the unfamiliar word, which can be particularly challenging when all the input is presented in a foreign language. MCL2 glosses are most demanding in terms of the cognitive loads they place on learners. However, this type of gloss does not optimize lower-proficiency learners' delayed word retention. The cognitive involvement load is not the primary influencing factor contributing to their word retention, as they often have difficulties correctly and efficiently inferring the meanings of unfamiliar words. In contrast, SL2 glosses require learners to put in moderate cognitive effort to understand the words and context in a foreign language, resulting in relatively deeper processing of the target words and stronger word retention that persist over time. SL2 glosses, with their moderate cognitive load for lower-proficiency learners, achieve the optimal balance of cognitive processing and decoding challenge. Compared with SL1 glosses, though this may lead to lower immediate recognition test scores due to the increased cognitive demands, the deeper processing helps enhance delayed word retention.

*(c). Effect of Glossing in Relation to Task Type*

The results showed that task types had a significant main effect on vocabulary acquisition and that the interaction effect of task types and gloss conditions was also significant. This suggests that gloss types influence learners'

vocabulary acquisition differently in recognition and recall tasks. The higher word gains observed on the recognition test indicate that glosses help students identify word meanings in context, and the lower gains on the recall test suggest that glosses may not adequately support learners in independently recalling and reproducing those meanings. Across all glossing conditions, senior high school English learners consistently perform better in the recognition tasks than in the recall tasks. This difference in performance can be attributed to the different natures of the two tasks. The recall task requires learners to independently retrieve word knowledge from memory without contextual cues. This makes recall more cognitively demanding than recognition, as it relies more deeply on knowledge formed through word processing. To successfully recall an unfamiliar word, learners must process it in a way that allows them to establish robust form-meaning connections. In contrast, the recognition task does not require this depth of processing since learners only need to identify the correct definition from provided options. This explains why learners generally perform better in recognition tasks than in recall tasks.

#### IV. CONCLUSION

The present study investigated the effect of four gloss types on Chinese senior high school EFL learners' vocabulary acquisition. The findings revealed that glossing exerted a positive effect on Chinese EFL learners' vocabulary acquisition in both the recognition tasks and the recall tasks, as well as for short-term and long-term word retention. Among the four types of gloss examined in this study, SL1 glosses were the most effective, and MCL2 glosses were the least effective. Regarding the interactive effect of task type and gloss type, it was found that the four types of glosses were consistently more effective in the recognition tasks than in the recall tasks, regardless of immediate or delayed effects. Moreover, all the gloss groups exhibited better short-term retention than long-term retention, while the single L2 gloss group showed the opposite.

The current findings have broadened our understanding of glossing, particularly concerning an understudied population, namely senior high school students. Furthermore, this study examined the interaction effect of task type and test timing, yielding valuable insights into how these factors influence the effectiveness of glossing. Pedagogically, the findings help determine how glosses should be created and used to maximize the benefits of glossing for high school students' word learning. This study revealed that SL1 glosses are the most effective for lower-proficiency learners in both immediate and delayed word retention. Providing glosses in students' native language within reading materials can thus facilitate their incidental vocabulary acquisition regardless of retention phases. Moreover, SL2 glosses reveal an increasing effect on long-term word retention, so they, too, can be employed to facilitate lower-proficiency learners' long-term retention. In essence, instructional L2 material developers should strategically use different glosses depending on the specific word learning objectives they intend to achieve.

#### REFERENCES

- [1] Çakmak, F., & Erçetin, G. (2018). Effects of gloss type on text recall and incidental vocabulary learning in mobile-assisted L2 listening. *ReCALL*, 30(1), 24–47.
- [2] Cheng, Y., & Good, R. (2009). L1 glosses: effects on EFL learners' reading comprehension and vocabulary retention. *Reading in a Foreign Language*, 21(2), 119–142.
- [3] Choi, S. (2016). Effects of L1 and L2 glosses on incidental vocabulary acquisition and lexical representations. *Learning and Individual Differences*, 45, 137–143.
- [4] Chun, D., & Plass, J. (1996). Effects of multimedia annotations on vocabulary acquisition. *Modern Language Journal*, 80(2), 183–212.
- [5] Hulstijn, J. (1992). Retention of inferred and given word meaning: Experiments in non-intention vocabulary learning. In P. Arnaud, & H. Bejoint (Eds.), *Vocabulary and Applied Linguistics* (pp. 113–125). Macmillan.
- [6] Jacobs, G., Dufon, P., & Hong, F. (1994). L1 and L2 vocabulary glosses in L2 reading passage: Their effects for increasing comprehension and vocabulary knowledge. *Journal of Research in Reading*, 17(1), 19–28.
- [7] Ko, M. H. (2012). Glossing and second language vocabulary learning. *TESOL Quarterly*, 46(1), 56–76.
- [8] 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.
- [9] Laufer, B. & Shmueli, K. (1997). Memorizing new words: Does teaching have anything to do with it? *RELC Journal*, 28(1), 89–108.
- [10] 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), 1–20.
- [11] Nagata, N. (1999). The effectiveness of computer-assisted interactive glosses. *Foreign Language Annals*, 32, 469–479.
- [12] Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge University Press.
- [13] Pulido, D. (2007). The effects of topic familiarity and passage sight vocabulary on L2 lexical inferencing and retention through reading. *Applied Linguistics*, 28(1), 66–86.
- [14] Rott, S. (2005). Processing glosses: a qualitative exploration of how form-meaning connections are established and strengthened. *Reading in a Foreign Language*, 17(2), 95–124.
- [15] Rott, S., & Williams, J. (2003). Making form-meaning connections while reading: A qualitative analysis of word processing. *Reading in a Foreign Language*, 15(1), 89–90.
- [16] Rott, S., Williams, J., & Cameron, R. (2002). The effect of multiple-choice L1 glosses and input-output cycles on lexical acquisition and retention. *Language Teaching Research*, 6(3), 183–222.

- [17] Watanabe, Y. (1997). Effects of single and multiple-choice glosses on incidental vocabulary learning. *JACET Bulletin*, 28, 177–191.
- [18] Yanagisawa, A., Webb, S., & Uchihara, T. (2020). How do different forms of glossing contribute to L2 vocabulary learning from reading? A Meta-regression analysis. *Studies in Second Language Acquisition*, 42(2), 411–438.
- [19] Yoshii, M. (2006). L1 and L2 glosses: Their effects on incidental vocabulary learning. *Language Learning & Technology*, 10(3), 85–101.
- [20] Zhang, C., & Ma, R. (2021). The effect of textual glosses on L2 vocabulary acquisition: A meta-analysis. *Language Teaching Research*, 0(0). <https://doi.org/10.1177/13621688211011511>

**Yeqiu Zhu** received her Ph.D. in English Language and Literature at Nanjing University, China. She is currently an associate professor at the School of Foreign Studies at Nanjing University in Jiangsu, China. Her research interests include second language acquisition, cognitive linguistics, second language pedagogy, and corpus linguistics.