

# Effects of Annotation Types Used at Different Point of Time during Reading on Vocabulary Learning\*

Xiaoling Liu

College of Foreign languages, Hunan University, Changsha, Hunan, China

Yang Gao

AECC Hunan Aviation Powerplant Research Institute, Zhuzhou, Hunan, China

**Abstract**—The study aims to investigate the effectiveness of two annotation types (text-only and text-picture) as well as the timing in annotation use during reading (pre-reading, while-reading and post-reading) on vocabulary learning of English major freshmen. Based on the analysis of the data collected from posttests and questionnaire, the following findings are obtained: 1) among three points of time in annotation use, the pre-reading annotation use results in the best vocabulary production and recognition, closely followed by the post-reading annotation use and then the while-reading annotation use; 2) when comparing the two annotation types, the text-picture annotations were more effective than the text-only annotations on improving vocabulary production and recognition; 3) there was no interaction between annotation types and points of time in annotation use. The findings indicate that the right use of annotations during reading will substantially promote vocabulary learning. They also shed some light on the research on annotation and vocabulary learning and provide implications on vocabulary instruction and teaching materials design.

**Index Terms**—timing in annotation use, annotation types, vocabulary learning

## I. INTRODUCTION

Vocabulary learning is a challenge that language learners at all levels have to face. Among recent second language vocabulary research, the distinction between incidental and intentional vocabulary learning has become one of the most influential debates (Nation, 2001; Hulstijn, 2001; Chen, 2013). Learning task, attention and context of learning have been thought to attribute to this set of distinction (Read, 2004). Studies have argued that both incidental and intentional vocabulary learning contribute to the vocabulary learning (Wesche and Paribakht, 1999; Nation, 2001; Hulstijn, 2001). Research arguing for the incidental vocabulary learning suggests that vocabulary learning occurs, in most cases, accidentally during reading or listening, but this kind of vocabulary learning is slow and incremental (Oxford, 1990, 2001). Reading is one of the main ways introduce new words to learners incidentally. Research investigating the effects of reading on vocabulary learning reveals that second language learners may incidentally gain certain degree of vocabulary knowledge in reading process (Rott, 1999, 2000).

Second language incidental vocabulary learning research in China is still at the initial stage, and the content is mostly the introduction of overseas research (Yue, 2008). The studies on second language incidental vocabulary learning and reading comprehension usually focused on the effects of different post-reading tasks, text difficulty, word exposure frequency and different glossing types (Duan and Yan, 2004, Lee et al, 2015, 2016). The number of studies in this field is small, and further research is badly needed.

Annotation has been considered as an assistant in second language reading material to facilitate reading comprehension with vocabulary learning being acquired incidentally (Watanabe, 1997; Ko, 1995; Laufer and Shmueli, 1997). As an instruction intervention which enhances the texts, annotation increases the exposure frequency of new words and draws learner's attention to the form and meaning of new words during reading. It is thought that by this way, both vocabulary learning and overall reading comprehension are enhanced. Research comparing the conditions with and without annotations has confirmed that annotations enhance incidental vocabulary learning (Hulstijn, et al., 1996; Hulstijn, 1992; Watanabe, 1997). The use of annotation supports the interactionist view of second language acquisition and depth of processing hypothesis (Hulstijn, 1992). Great attention has been focused on the comparison between annotation forms, language and types. (Watanabe, 1997; Ko, 1995; Laufer and Shmueli, 1997; Duan and Yan, 2004; Joyce, 1997; Yoshii, 2000, 2006; Chen, 2013). Most of the studies abroad come to mix findings. However, there is a big gap of the research statues between abroad and at home. Little research has been conducted to comparing the effectiveness of text-only and text-picture annotations in China (Wu and Xu, 2009). Moreover, few researchers have paid much attention on the points of time in annotation use related to the reading process.

\* Funded by Integrated English Course Construction and provincial project (HNJG-2020-0116).

The present study attempts to gain further insight into the effectiveness of different annotation types in enhancing incidental vocabulary learning, in the mean time, explore the effects of different points of time in annotation use on incidental vocabulary learning and reading comprehension of first-year English major students.

## II. LITERATURE REVIEW

### A. *Vocabulary Learning and Its Relation with Reading*

Research concerning vocabulary learning has received great recognition in the field of Second Language Acquisition and Foreign Language Learning (FLL) since 1990s, as researchers have argued for and proved the importance of vocabulary in SLA and FLL. Based on the relationship between vocabulary learning and other learning process, such as grammar learning, reading, listening and speaking, it has been proved that limited vocabulary knowledge will prevent language learners from developing reading ability and communicating effectively in target language and thus close the door to further learning (Richard, 1976; Nation, 2001; Schmidt, 2000). Researchers have pointed out that vocabulary knowledge is multi-faceted. There are two influential distinctions of vocabulary knowledge: size and depth, (Nation, 2001; Read, 2000), receptive knowledge and productive knowledge (Nation, 2001; Webb, 2005; Chen et al, 2019; Zhou, 2010). The research concerning vocabulary learning proves that the vocabulary size and depth are indicators of vocabulary knowledge and the test forms which could cover the assessment of both size and depth of vocabulary are conclusive (Nation, 2001).

The discussion regarding the distinction between receptive and productive vocabulary knowledge, the two types of vocabulary knowledge are widely estimated but not well defined (Melka, 1997). The controversial issue comes from the criteria for knowing a word receptively or productively. The distinction between receptive and productive knowledge displays in two aspects. One is the quality of the knowledge in mental lexicon, and the other is the ability that learners produce the knowledge. Thus, it is said that the receptive and productive dimensions are identified as “a bridging dimension between lexical competence and performance” (Zareva et al., 2005). Meanwhile, it is generally believed that words are known receptively first and then productively and vocabulary knowledge should be regarded as a continuum on which a word grows from receptive to productive status. Moreover, although the transfer from receptive to productive knowledge is not very clear, it is generally thought that receptive vocabulary knowledge is larger than productive vocabulary knowledge (Zhou, 2010). However, the concept of continuum does not necessarily mean that receptive knowledge predicts its production (Henriksen, 1999). For example, Henriksen (1999) suggests that only a limited number of words that are learnt receptively will ever become productive. Moreover, the concept of continuum also does not mean that vocabulary development is always linear. For example, Waring (1997) suggests that the linear view presupposes that one’s vocabulary must be complete before any aspect of production can proceed, which means it is impossible for learners to use a word if they do not understand all aspects of its meaning. Therefore, receptive knowledge and productive knowledge are interrelated, which can be seen as a continuum, but they are also independent, which do not have a linear relation. It is better to consider the receptive and productive knowledge both respectively and correlatively. Based on this consideration, this study tests the receptive and productive knowledge respectively and discusses the results from the prospects of each and both vocabulary knowledge.

With the recognition of the distinctions on vocabulary knowledge, there is a need to further these distinctions to the discussion of vocabulary learning which turns the conceptual distinctions to the practical ones. The well-recognized distinction in researches of vocabulary learning is intentional and incidental vocabulary learning (Nation, 2001). This study refines the definitions of intentional and incidental vocabulary learning in Chen’s (2013) study. Therefore, in this study, intentional vocabulary learning refers to the situation in which learners are informed the requirement on vocabulary learning before performing the task, while incidental vocabulary learning refers to the situation in which learners are NOT informed the requirement on vocabulary learning before performing the task.

Many studies connect vocabulary learning with reading (Nation, 2001; Shu, Anderson and Zhang, 1995; Pulido, 2004). Even though most of the researchers support that the vocabulary learning while reading is incidental, still others argue that some level of attention and cognitive process are needed for the “incidental” vocabulary learning to happen (Angelika Rieder, 2003; Zhao and Yang, 2010). The given attention which would directly affect the learning outcomes is determined by the context for the target vocabulary, attention types and task requirements. In other words, even though vocabulary learning is considered to be incidental, the task requirements, to large extent, determine the allocation of learner’s attention which will certainly affect the vocabulary learning outcomes. Thus many factors, such as the meaning and form of the word, context, type of attention and task requirements, should be carefully considered when designing an experiment or research.

Although reading has certain influence on vocabulary learning, negative evidences from several studies showing that incidental vocabulary learning through reading may lead to some problems (Grabe, 1997). Firstly, unknown words are ignored by learners because they confuse the unknown words with the words previous known or simple unawareness of the unknown words (Paribakht and Watanabe, 1999). Secondly, unknown words are ignored due to learners’ decision or unimportance of the unknown words to the understanding of the overall reading content. Thirdly, the meanings of unknown words are not inferable from the context due to the lack of vocabulary knowledge of surrounding words or contextual clues. Fourthly, wrong inferences of unknown words are made because it’s difficult to understand their

structure or meaning or when the contextual cues are insufficient. Fifthly, a single encounter of words can not guarantee vocabulary acquisition (Hulstijn et al. 1996).

With the disadvantages of incidental vocabulary learning mentioned above in mind, researchers suggest several ways to overcome those disadvantages and promote incidental vocabulary learning, including the use of dictionary, context-based vocabulary exercise, repetition, and annotation or gloss.

Studies on using dictionary while reading prove that the dictionary use will enhance the rate of vocabulary learning (Grabe and Stroller, 1997, 2001). However, there's unavoidable problem that look-up behavior in dictionary use certainly interrupt the reading process and devalue the effect of meaningful context. The introducing of context-based vocabulary exercises is frequently discussed by researchers and proved to be effective (Wesche and Paribakht, 2000). Discussion on the optimal number of exposures for retaining a word arrives no conclusive answer (Webb, 2007).

To facilitate incidental vocabulary learning, annotation and gloss were also employed. Nation (1990, 2001; Lee et al, 2015, 2016) argues for the beneficial effect of vocabulary annotation on both vocabulary learning and reading comprehension. He states that: first, annotation can provide knowledge of unknown words which in turn facilitate reading comprehension; second, annotation can prevent incorrect guessing; third, annotation can minimize reading interruption; fourth, annotation can promote individualized learning; fifth, annotation can draw attention to unknown words which in turn enhance vocabulary learning. He proposes that learners should be encouraged to use annotations or glosses to confirm their guessing of unknown words. Many studies also confirm the positive effects of annotation use on incidental vocabulary learning (Hulstijn, 1992; Watanabe, 1997; Ko, 1995; Laufer and Shmueli, 1997). Detailed review on annotation use will be provided in the following part.

### *B. Annotation and Its Relation with Vocabulary Learning*

Previously mentioned, annotation is an effective way to enhance vocabulary learning (Hulstijn, 1992; Watanabe, 1997; Ko, 1995; Laufer and Shmueli, 1997).

Empirical studies concerning different forms of annotations have been conducted, such as comparing the single annotation with multiple-choice annotations (Hulstijn, 1992; Watanabe, 1997; Miyasako, 2002; Duan and Yan, 2004). No conclusive result is reached on the comparison of the effectiveness of multiple-choice gloss and single gloss on vocabulary learning. However, differing from the studies abroad, unified conclusion is drawn that favor multiple-choice gloss over single gloss at home. As Hulstijn (1992) has concluded that multiple-choice glossing owns unavoidable disadvantage of causing high rate of incorrect inference, the present study will not adopt the method of multiple-choice glossing.

Besides the studies on forms of annotations, there are other studies focusing on language of annotations. These studies mainly concentrate on comparing annotation of L1 and L2 (Jacob et al., 1994; Ko, 1995; Laufer and Shmueli, 1997; Miyasako, 2002; Xu, 2010). A relatively conclusive result can be drawn which favors L1 gloss over L2 gloss both abroad and at home. Therefore, the present study will use L1 glossing so as to make good use of the positive effect of annotation.

Empirical studies comparing text-picture annotations and text-only annotations have been conducted with relatively conclusive result, that is, text-picture annotations lead to better results on vocabulary learning than text annotations (Boers, et al., 2017; Yoshii and Flaitz, 2002; Yoshii, 2006; Wu and Xu, 2009). However, when testing the receptive and productive aspects of vocabulary increase caused by annotations, Yoshii's (2006) study casts some doubt on the superior effect of text-picture annotations over text annotations. Yoshii's (2006) study compares the effectiveness of text-picture glosses and text-only glosses on incidental vocabulary learning. Japanese university students participate in the study. The results show that the significant superior effect of text-picture glosses over text-only glosses only exists in the increase on productive vocabulary knowledge but not on receptive vocabulary knowledge. Yoshii (2006) explains that the inconsistency results between receptive and productive aspects is due to the nature of the posttests as definition-supply test relies on both textual and pictorial cues while options in recognition test are mostly associated with textual cues. Therefore, to fill the gap of the studies comparing text-picture annotation and text-only annotation in China, and clarify the doubt casted by Yoshii's (2006) study, the present study is conducted.

In summary, it's confirmed that annotation helps incidental vocabulary learning. They prove the effectiveness of L1 annotation, multiple-choice annotation, and text-picture annotation over other annotation types. However, few studies consider the comparison among the different points of time in annotation use related to the reading process. Therefore, in this study, the comparison among the different points of time in annotation use related to the reading process is conducted, and vocabulary learning under such conditions addressed. This study chooses English-major freshmen as its subjects. The goal of the study is to try to find a best way to enhance vocabulary learning with least compromises on reading comprehension.

## III. RESEARCH DESIGN

### *A. Research Questions*

To test the effectiveness of annotation, multiple-choice annotation, and text-picture annotation over other annotation types and the points of time in annotation use on vocabulary learning, the following specific research questions are addressed in this study:

1. Which point of time in annotation use during reading (pre-reading, while-reading, or post-reading) results in best vocabulary production and recognition?
2. Which type of annotation (text-only or text-picture) result in better vocabulary production and recognition?
3. Is there any interaction effect between annotation types and points of time in annotation use on vocabulary learning? If any, what is it like?

### B. Participants

The participants are 105 English major freshmen from a certain university in China. Their average age is 18 years old. They are all Chinese and have been learning English since their primary school years, with no exposure to the target language abroad before.

The participants volunteer to participate the experiment and are assigned into three groups, namely pre-reading annotation use group (PreA), while-reading annotation use group (WA), and post-reading annotation group (PostA). Three groups receive the experiment treatment at the same time conducted by the researcher and other two teachers. The homogeneity among groups is tested based on the total scores of their placement English examination in the university and the scores of reading section in the exam, age, and years of learning English.

### C. Research Materials

Research materials used in the present study include training material, reading material, target words, questionnaires and assessment material.

The training materials are meant for letting participants get familiar with annotation use and test format. The reading part and word list of the English textbook which subjects commonly use are used as training material for annotation use. Unit Two in *An Integrated English Course 3*, which includes 14 short paragraphs, 714 words, are chosen as the reading material for training. 14 known words are annotated, half of which (7 words) with only L1 definitions and the other half of which (7 words) with both L1 definitions and corresponding pictures. Two experienced teachers are consulted on the choice of target words for the training material.

As for the reading material used in this study, the regular teaching material *An Integrated English Course 3* is used as reference for choosing reading material on content, length, sentence complexity and readability. Sentence complexity is calculated based on the average number of words in every sentence. The readability is measured by the Flesch-Kincaid Readability Test, which contains two subtests --- Flesch Reading Ease test and Flesch-Kincaid Grade Level. The following table gives the data of reference material and designed material.

TABLE 1  
RESEARCH MATERIALS DESCRIPTION

	Length	Sentence complexity	Readability	
			Flesch Reading Ease test	Flesch-Kincaid Grade Level
Reference material	714 words	17 words/sentence	85.3	5.1
Designed material	700 words	16 words/sentence	85.9	4.8

The subjective ideas toward the designed reading material with respect to its length, difficulty and understandability were collected by a questionnaire conducted by 15 students who do not participate in the experiment and who are of the same level with experiment participants. The results support the objective results given above.

14 words were chosen as target words. According to Nation (2001), the proportion of unknown words should be between 3% and 5% for best effects of annotation and reading comprehension. 14 target words constitute 4.4% of the total words of the reading material, which is of appropriate proportion that satisfies both reading comprehension and annotation effect initiation. Target words' categories are limited to noun and verb with 10 nouns and 4 verbs. They are selected from word lists required by TEM 4 (Test for English Majors Level 4). The 15 students who give their subjective ideas toward reading material were also required to pick out unknown words in the reading material. The 14 target words chosen were the most frequently picked out words by these students. The rest of low frequently picked out unknown words were replaced by easier synonyms from the English wordlist for English-major freshmen students. Half of the target words (7 words) are annotated with text only, in which 5 words are nouns and 2 words are verbs, and the other half of the target words (7 words) are annotated with text and picture, in which there are also 5 nouns and 2 verbs. The target words are bolded in the reading material.

Three questionnaires are included in this study. The first questionnaire is for collecting subjective ideas of students in validating reading material and target words. It consisted of 5 open-ended questions for personal information including student's ID number, gender, age, the number of years learning English, and the placement English test score, and 2 five-point Likert's scale questions and 1 open-ended question for material design related information. The second questionnaire is for collecting subjects' demographic information and their usual ways of vocabulary learning. 5 open-ended questions are given for collecting personal information including student's ID number, gender, age, the number of years learning English, and the placement English test score, and 2 multiple-choice questions are for vocabulary learning tendency information. The third one is for collecting subjects' opinions and attitudes toward the

experiment. 5 close-ended questions address the annotation modes and 4 close-ended questions address points of time in annotation use.

The assessment materials used in this study are vocabulary posttests which includes immediate posttests and delayed posttests. Vocabulary posttests take the form of word definition supply test and word recognition test to test both the receptive and productive vocabulary knowledge. For each target word, there is a corresponding item in both word definition supply test and word recognition test, thus there are 14 items respectively. In word definition supply test, words are given isolated from context. Only the meanings of the target words are asked to give as the annotations only provide the meanings of the target words. In word recognition test, multiple-choice form is adopted with one right choice and three distracters. In order to test the retention of vocabulary, delayed vocabulary posttests are set in the same test format with different item order from that of immediate posttests. The Cronbach's alpha from SPSS reliability analysis shows that the reliability was 0.828 for word definition supply test and 0.562 for word recognition test based on the pilot study.

#### D. Procedures

The experiment is conducted in the autumn semester of 2019, and takes one 30-minute session for the treatment and immediate posttests and one 10-minute session for the delayed posttests. The experiment is carried out as shown in the following order.

Firstly, the pre-experiment questionnaire is handed out to collect demographic information, English learning experience, and their usual way of vocabulary learning.

Secondly, three groups of students receive the treatment at the same time carried out by the researcher and two other instructors respectively. The three groups are pre-reading annotation use group (PreA), while-reading annotation use group (WA), and post-reading annotation group (PostA). Instruction in L1 is given without telling the experiment subjects about vocabulary tests afterwards. In PreA, annotations are given to the subjects first. The experiment subjects are asked to read the annotations for the following reading. Then annotations are collected, and reading material is handed out to the experiment subjects. In WA, annotations and reading material are given to the experiment subjects at the same time. In PostA, reading material is given to the experiment subjects first. The experiment subjects are asked to guess the meaning of unknown words while they read. After finishing reading, annotations are handed out to the experiment subjects to check their guessing about the unknown words. After the treatment, both the reading material and annotations are collected in all three groups.

Thirdly, the immediate posttests are conducted in all three groups. The immediate vocabulary posttests involve both the receptive and productive vocabulary testing. The order of the immediate posttests is word definition supply test first, and word recognition test second, addressing productive vocabulary knowledge and receptive vocabulary knowledge respectively.

Fourthly, the post-experiment questionnaire is handed out to collect subjects' opinions towards the experiment with respect to the annotation modes and the points of time in annotation use.

Fifthly, the delayed posttests are conducted in all three groups two weeks later. The subjects unexpectedly receive the delayed posttests. The delayed posttests include the vocabulary posttests, and the post-experiment questionnaire. The order of the delayed posttests is word definition supply test first, and word recognition test second. The content of the delayed vocabulary posttests is the same with the immediate vocabulary posttests with item order rearranged.

The following figure gives the participants assignment and procedure of the experiment.

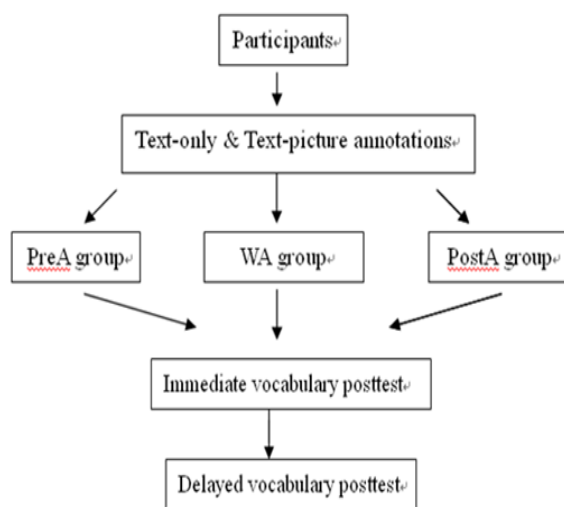


Figure 1 Group Participant Assignment And Research Procedures

## IV. RESULT AND DISCUSSION

In this part, the results of the study are presented according to the research questions.

A. *Effects of Timing in Annotation Use on Vocabulary Learning*

*Research question one: Which point of time in annotation use during reading (pre-reading, while-reading, or post-reading) results in best vocabulary production and recognition?*

The research question can be answered from three aspects --- vocabulary production, recognition and changes of vocabulary retention over time.

In terms of vocabulary production, Table 2 presents the descriptive statistics and the post hoc results of both immediate word definition supply test (IWDST) and delayed word definition supply test (DWDST).

TABLE 2  
DESCRIPTIVE STATISTICS, ANOVA AND POST HOC TEST RESULTS FOR IWDST AND DWDST IN TIMING OF ANNOTATION USE

IWDST						DWDST					
(I) PTAU	M	SD	(J) PTAU	Mean Difference (I-J)	Sig.	(I) PTAU	M	SD	(J) PTAU	Mean Difference (I-J)	Sig.
PreA	5.19	1.600	WA	1.59**	.000	Pre A	2.77	1.889	WA	1.07**	.000
			Post A	.70**	.012				Post A	.24	.363
WA	3.60	1.748	Pre A	-1.59**	.000	WA	1.70	1.255	Pre A	-1.07**	.000
			Post A	-.89**	.002				Post A	-.83**	.002
Post A	4.49	1.622	Pre A	-.70**	.012	Post A	2.53	1.548	Pre A	-.24	.363
			WA	.89**	.002				WA	.83**	.002

PTAU= point of time in annotation use

\*\*p<.05

Based on the mean scores from both IWDST and DWDST, PreA group (M=5.19, SD=1.600; M=2.77, SD=1.889) outscored the PostA group (M=4.49, SD=1.622; M=2.53, SD=1.548) which in turn outscored WA group (M=3.60, SD=1.748; M=1.70, SD=1.255). ANOVA was performed in order to verify the significance of the differences. The results of ANOVA indicate that the differences among three points of time in annotation use were statistically significant in both IWDST and DWDST (Sig.=.000 <.05; Sig.=.000 <.05).

The significance of differences shown by ANOVA guaranteed the necessity of performing the post hoc test so as to find out where the differences lie. Therefore, the post hoc test was performed for both IWDST and DWDST. The results of post hoc test show that the differences between PreA group and WA group (Sig.=.000 <.05; Sig.=.000 <.05) and PostA group and WA group (Sig.=.002 <.05; Sig.=.002 <.05) reached statistical significance in IWDST and DWDST. However, the differences between PreA group and PostA group only reached statistical significance (Sig.=.012 <.05) in IWDST but not DWDST (Sig.=.363 >.05).

In terms of vocabulary recognition, Table 3 shows the descriptive statistics and the post hoc results of both immediate word recognition test (IWRT) and delayed recognition supply test (DWRT).

TABLE 3  
DESCRIPTIVE STATISTICS, ANOVA AND POST HOC TEST RESULTS FOR IWRT AND DWRT IN TIMING OF ANNOTATION USE

IWRT						DWRT					
(I) PTAU	M	SD	(J) PTAU	Mean Difference (I-J)	Sig.	(I) PTAU	M	SD	(J) PTAU	Mean Difference (I-J)	Sig.
Pre A	6.03	.978	WA	.47**	.005	Pre A	4.56	1.839	WA	.71**	.017
			Post A	.11	.495				Post A	.04	.885
WA	5.56	1.163	Pre A	-.47**	.005	WA	3.84	1.799	Pre A	-.71**	.017
			Post A	-.36**	.034				Post A	-.67**	.025
Post A	5.91	.989	Pre A	-.11	.495	Post A	4.51	1.767	Pre A	-.04	.885
			WA	.36**	.034				WA	.67**	.025

PTAU= point of time in annotation use

\*\*p<.05

It can be seen that the scores were relatively close among three groups with PreA group (M=6.03, SD=.978; M=4.56, SD=1.839) slightly outscoring PostA group (M=5.91, SD=.989; M=4.51, SD=1.767) which in turn outscored WA group (M=5.56, SD=1.163; M=3.84, SD=1.799) in both IWRT and DWRT. Moreover, ANOVA was performed to examine the significance of differences. The results reveal that the differences among three points of time in annotation use were statistically significant in both IWRT and DWRT. Therefore, the post hoc test was carried out to further examine the differences between groups in both IWRT and DWRT. The results of post hoc test shows that the differences between PreA group and WA group (Sig.=.005 <.05; Sig.=.017 <.05) and between PostA group and WA group (Sig.=.034 <.05; Sig.=.025 <.05) reached statistical significance in both IWRT and DWRT. However, the differences between PreA group and PostA group did not reach statistical significance in both IWRT and DWRT (Sig.=.495 >.05; Sig.=.885 >.05).

With respect to the changes of vocabulary retention over time, it is presented by the retention rate which is the delayed vocabulary posttests scores divided by the immediate posttests scores. Table 4 gives the retention rate of vocabulary posttests by points of time in annotation use.

TABLE 4  
THE RETENTION RATE OF VOCABULARY POSTTESTS IN TIMING OF ANNOTATION USE

Points of Time	n	WDST			WRT		
		M(I)	M(D)	RR	M(I)	M(D)	RR
PreA	35	5.19	2.77	53.37%	6.03	4.56	75.62%
WA	35	3.60	1.70	47.22%	5.56	3.84	69.06%
PostA	35	4.49	2.53	56.35%	5.91	4.51	76.31%

M(I)=means from immediate posttest, M(D)=means from delayed posttest, RR=retention rate

It indicates that vocabulary recognition had higher retention rate than vocabulary production. Moreover, in both WDST and WRT, PostA group (56.35%, 76.31%) achieved the highest retention rate which was followed by PreA group (53.37%, 75.62%) and WA group (47.22%, 69.06%).

As previously mentioned, the distinction of incidental and intentional vocabulary learning will lead to different vocabulary learning outcome. It was expected that intentional vocabulary learning would lead to better results than incidental vocabulary learning, which means the pre-reading annotation use condition would result in better performance than the while-reading annotation use condition and the post-reading annotation use condition. The results partially supported the above expectation with constant significantly better effect of pre-reading annotation use than while-reading annotation use. At the first glance, this order of the three conditions supported the expectation. However, the significant differences between the pre-reading annotation use condition and the post-reading annotation use condition were only found in immediate production posttest but not in other three posttests (immediate recognition posttest, delayed production posttest, and delayed recognition posttest). These non-significant differences between the pre-reading annotation use condition and the post-reading annotation use condition coincide with the results obtained by Ko (1995). According to Ko's (1995) explanation, the amount of attention or effort paid to annotations determines the vocabulary learning outcomes. In other words, if the same or nearly the same amount of attention or efforts is given to annotations, there will be no surprise to find close vocabulary learning outcomes no matter the vocabulary learning process is intentional or incidental. These results also coincide with Yoshii (2006)'s study on the inconsistency results between receptive and productive aspects of vocabulary learning. The productive tests are more likely to show significant differences than the receptive tests because of the differences on the depth of vocabulary knowledge they test.

In the present study, the participants in the pre-reading annotation use condition certainly had to pay much attention to annotations as they were requested to memorize the words in the annotations, and those in the post-reading annotation use condition also need to give attention to annotations when they were guessing the meanings of unknown words and later checking annotations even though this process was finished without any intention. In addition, this result also can be explained from other two perspectives: the perspective of cognitive load and the perspective of involvement load hypotheses. Firstly, the memorization of the words in the annotations in the pre-reading annotation use condition undoubtedly increased the germane cognitive load which fosters the learning process. Likewise, the guessing of unknown words and checking annotations deepened the processing of the words in the post-reading annotation use condition, which also increased the germane cognitive load. Secondly, from the perspective of involvement load hypotheses, the participants in the pre-reading annotation use condition might have higher need as memorizing the words was one target they need to achieve, while those in the post-reading annotation use condition performed deeper evaluation as they had to guess the meanings of unknown words to comprehend the reading material. Moreover, based on the information collected by Question 9 in Questionnaire 2, the participants preferred the post-reading annotation use condition (43.8%) to the pre-reading annotation use condition (31.4%) on benefiting vocabulary learning. Therefore, it is safe to say that the non-significant differences between pre-reading annotation use condition and the post-reading annotation use condition are perfectly understandable.

Based on what has been discussed above, it can be easily found that the order of the effects of the three conditions on vocabulary learning is the pre-reading annotation use condition the first, followed by the post-reading annotation use condition and then the while-reading annotation use condition. This order is just opposite to that of the reading comprehension. This confirms the theory about limited cognitive capacity, as the more attention paid to the reading comprehension means less attention left for the target words, or vice versa.

One interesting finding is worth mentioning. When considering the changes of vocabulary retention over time, the post-reading annotation use condition reveals the highest retention rate (56.35% for word definition supply test, 76.31% for word recognition test). The change of superiority between the post-reading annotation use condition and the pre-reading annotation use on retention rate indicates that rote memorization of vocabulary would provide a short-term boost to vocabulary learning with a relative fast decay, but guessing or inferring of vocabulary during reading has a sustainable or long-term positive effect on vocabulary learning.

### B. Effects of Annotation Types on Vocabulary Learning

*Research question two: Which type of annotation (text-only or text-picture) result in better vocabulary production and recognition?*

This research question also will be answered from the following three aspects: vocabulary production, recognition and changes of vocabulary retention over time.

Firstly, regarding to vocabulary production, Table 5 shows the descriptive statistics and t test results of both immediate word definition supply test (IWDST) and delayed word definition supply test (DWDST).

TABLE 5  
DESCRIPTIVE AND INFERENTIAL STATISTIC RESULTS FOR WDST IN ANNOTATION TYPES

Annotation	n	IWDST			DWDST		
		M	SD	Sig.	M	SD	Sig.
Text-only	105	4.10	1.934	.004**	2.08	1.555	.019**
Text-picture	105	4.75	1.537		2.59	1.697	

\*\*p<.05

It shows descriptively that the text-picture annotation (M=4.75, M=2.59) did better than the text-only annotation (M=4.10, M=2.08) in helping participants' vocabulary production, and the differences between two annotation types reached statistical significance in both IWDST and DWDST (Sig.=.004 <.05; Sig.=.019 <.05).

Secondly, in terms of vocabulary recognition, Table 6 shows the descriptive statistics and t test results of both immediate word recognition test (IWRT) and delayed recognition supply test (DWRT).

TABLE 6  
DESCRIPTIVE AND INFERENTIAL STATISTIC RESULTS FOR WRT IN ANNOTATION TYPES

Annotation	n	IWRT			DWRT		
		M	SD	Sig.	M	SD	Sig.
Text-only	105	5.48	1.241	.000**	3.85	1.935	.000**
Text-picture	105	6.19	.681		4.76	1.584	

\*\*p<.05

It can be seen that the text-picture annotation (M=6.19, M=4.76) did better than the text-only annotation (M=5.48, M=3.85) in helping participants' vocabulary recognition, and the differences between two annotation types reached statistical significance in both IWRT and DWRT (Sig.=.000 <.05; Sig.=.000 <.05).

Thirdly, changes of vocabulary retention over time are presented by the retention rate. Table 7 gives the retention rate of vocabulary posttests by annotation types.

TABLE 7  
THE RETENTION RATE OF VOCABULARY POSTTESTS IN ANNOTATION TYPES

Annotation	N	WDST			WRT		
		M(I)	M(D)	RR	M(I)	M(D)	RR
Text-only	105	4.10	2.08	50.73%	5.48	3.85	70.26%
Text-picture	105	4.75	2.59	54.53%	6.19	4.76	76.90%

M(I)=means from immediate posttest, M(D)=means from delayed posttest, RR=retention rate

It indicates that vocabulary recognition had higher retention rate than vocabulary production. Moreover, both WDST and WRT, the text-picture annotation (54.53%, 76.90%) helped participants achieve much higher retention rate than the text-only annotation did (50.73%, 70.26%).

Previous studies (Boers, et al., 2017; Yoshii and Flaitz, 2002; Chen, 2013) shows that words annotated with both verbal and visual modes of information would lead to effective vocabulary retention. The results of this study further support this conclusion. The retention rate reveals the changes of vocabulary retention over time, including both productive and receptive aspects. Comparing the results of vocabulary retention to the previous studies (Yoshii and Flaitz, 2002; Chen, 2013; Wu and Xu, 2009), the results confirmed the long-time effectiveness of the multimedia annotations, in this case, the text-picture annotations, in facilitating vocabulary learning. Moreover, the information gathering from two questionnaires also helps the understanding of the results. According to the results of questionnaire 1, the participants mostly received instruction of vocabulary by using pronunciation and image, and as for the favorite way of learning vocabulary, they showed more interests in multimedia assisted vocabulary learning. Question 1 to 5 in questionnaire 2 gives some insight of the participants' opinions towards comparing the two types of annotation used in the experiment. Comparing to the text-only annotations, the participants favored text-picture annotation more. The preference of text-picture annotations to text-only annotations can also contribute to the results.

### C. Interaction Effect between Types and Timing on Vocabulary Learning



*Research question 3: Is there any interaction effect between annotation types and points of time in annotation use on vocabulary learning? If any, what is it like?*

The results of the interaction between annotation types and points of time in annotation use in vocabulary learning based on four vocabulary posttests are presented in figure 2.

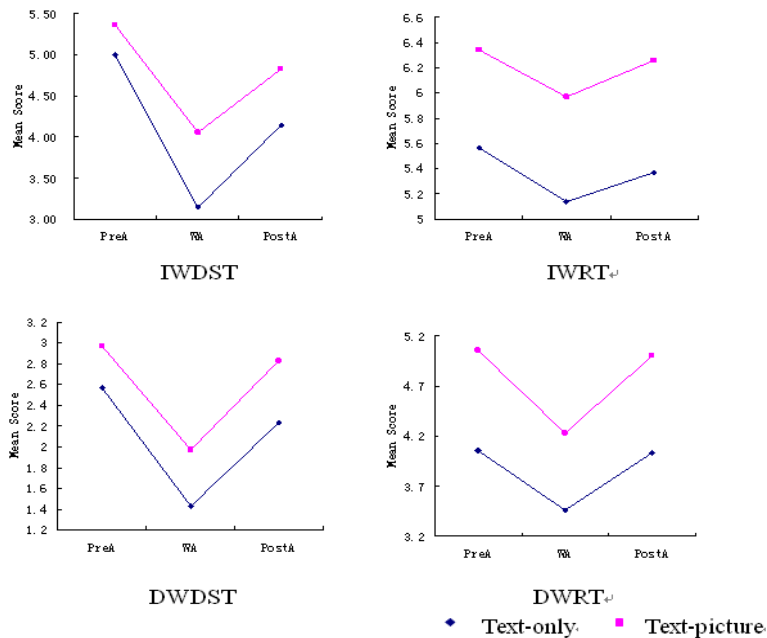


Figure 2 Mean Scores by Annotation Types and Timing in Annotation Use during Reading

Figure 2 exhibits the profile plots of the four vocabulary posttests based on the mean scores. It indicates that there was no interaction between annotation types and points of time in annotation use in all four vocabulary posttests.

The results mean that the superiority of text-picture annotations over text-only annotations in facilitating vocabulary retention does not depend on points of time in annotation use. It can be explained by the experiment design. Referring to the comparison of annotation types, within-subject design was adopted, which means each subject made use of two types of annotations. When comparing points of time in annotation use, between-subject design was used, which means subjects in different groups receive different treatment. Therefore, the differences between scores of these two independent variables would have different natures.

## V. CONCLUSION

The main purpose of this study is to compare the effectiveness of two annotation types (text-only and text-picture) and of three points of time in annotation use (pre-reading, while-reading and post-reading) on helping vocabulary learning of English major students. Moreover, the interaction between annotation types and points of time in annotation use is investigated. The main research findings are presented as follows:

Firstly, among three points of time in annotation use, the pre-reading annotation use results in the best vocabulary production and recognition, closely followed by the post-reading annotation use and then the while-reading annotation use.

Secondly, when comparing the two annotation types, the text-picture annotations were more effective than the text-only annotations on helping vocabulary production and recognition.

Finally, there was no interaction between annotation types and points of time in annotation use.

Based on the research results, some pedagogical implications can also be suggested.

Firstly, the finding of superior effect of text-picture annotation over text-only annotation on vocabulary learning could be taken into consideration when designing instruction materials.

Next it informs instructors to choose the right type of annotations to enhance vocabulary learning.

Moreover, the outperformance of pre-reading annotation use group indicates that if vocabulary learning is the ultimate goal of learning, using the annotations before reading could be the best way to achieve the goal. This also implies that drawing students' attention directly to the task, such as vocabulary memorization in this case, can have positive impact on their learning performance. If the vocabulary learning is not the ultimate goal, to use annotation after reading would also provide positive effect on vocabulary learning and in the mean time guarantee reading comprehension.

## REFERENCES

- [1] Boers, F., Warren, P., He, L., Deconinck, J. (2017) Does adding pictures to glosses enhance vocabulary uptake from reading? *System*, 66, 113–129
- [2] Brown, R., Waring, R. and Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening and listening to stories. *Reading in a Foreign Language*, 20(2), 136-163.
- [3] Chen, I.-J., & Yen, J.-C. (2013). Hypertext annotation: Effects of presentation formats and learner proficiency on reading comprehension and vocabulary learning in foreign languages. *Computers & Education*, 63, 416–423.
- [4] Duan Shiping and Yan Chensong. (2004). Effect of Multiple-choice Annotation on Incidental Vocabulary Learning. *Foreign Language Teaching and Research*, 36(3), 213-218.
- [5] Ellis, R. (2002). Grammar teaching - practice or consciousness raising? In J. C.Richards and W.A.Renandya (Eds.), *Methodology in language teaching: an anthology of current practice*. Cambridge: Cambridge University Press. 167-172.
- [6] Grabe, W. & Stoller, F.L. (1997). Reading and vocabulary development in a second language: a case study. In: Coady, J. & Huckin, T. (Eds.), *Second Language Vocabulary Acquisition*. Cambridge: Cambridge University Press, 98-122.
- [7] Hulstijn, J.H (1992). Retention of inferred and given word meanings: Experiments in incidental vocabulary learning. In: P.J.L Arnaud & H. Bejoint (Eds.), *Vocabulary and Applied Linguistics*, London: Macmillan, 113-125.
- [8] Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction*. Cambridge: Cambridge University Press, 258-286.
- [9] Laufer, B., & Shmueli, K. (1997). Memorizing new words: Does teaching have anything to do with it? *RELC journal*.28 (1), 89-108.
- [10] Lee, H., & Lee, J. H. (2015). The effects of electronic glossing types on foreign language vocabulary learning: Different types of format and glossary information. *The Asia-Pacific Education Researcher*, 24(4), 591–601.
- [11] Lee, H., Lee, H. & Lee, J. H. (2016) Evaluation of Electronic and Paper Textual Glosses on Second Language Vocabulary Learning and Reading Comprehension. *The Asia-Pacific Education Researcher*, 25, 499–507
- [12] 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.
- [13] Nation, P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- [14] Paribakht, T. S., & Wesche, M. (1999). Reading and "incidental" L2 vocabulary acquisition: An introspective study of lexical inferencing. *Studies in Second Language Acquisition*, 21(2), 195-224.
- [15] Pulido, D. (2004). The relationship between text comprehension and second language incidental vocabulary acquisition: A matter of topic familiarity? *Language Learning*, 54(3), 469-523.
- [16] Read, J. (2004). Research in teaching vocabulary. *Annual Review of Applied Linguistics*, 24(1), 146-161.
- [17] Rieder, A. (2003). A cognitive view of incidental vocabulary acquisition: from text meaning to word meaning? *View*, 11(1&2), 53-71.
- [18] Rott, S. (2000). Relationships between the process of reading, word inferencing and incidental word acquisition, in assigning meaning to form. In Lee, J. & Valdman, A., editors, *Issues in Language Programme Direction*, Heinle and Heinle: Boston, 255-282.
- [19] Schmitt, N. (2000). *Vocabulary in Language Teaching*. Cambridge: Cambridge University Press.
- [20] Watanabe, Y. (1997). Input, intake, and retention: Effects of increased processing on incidental learning of foreign language vocabulary. *SSLA*, 19, 287-307.
- [21] Wesche, M., & Paribakht, T. S. (1999). Incidental L2 vocabulary acquisition through reading: An introspective study. In M. Wesche & T.S. Paribakht, (guest editors), *Studies in Second Language Acquisition*, 21 (2), 195-224.
- [22] Wesche, M. & Paribakht, T.S. (2000). Reading-Based Vocabulary Exercises in Second Language Vocabulary Acquisition. *The Modern Language Journal*, 84(2), 196-213.
- [23] Wu Wei and Xu Hong. (2009). Effects of Different Types of Annotations on Incidental Vocabulary Learning in a Multimedia Environment. *Journal of PLA University of Foreign Languages*, 32(1),44-49.
- [24] Xu, H. (2010). Review of effects of glosses on incidental vocabulary learning and reading comprehension. *Chinese Journal of Applied Linguistics*, 33(1), 56-73.
- [25] Xu, Xiaohui, (2010), the effects of glosses on incidental vocabulary acquisition in reading. *Journal of Language Teaching and Research*, 1(2), 117-120.
- [26] Yoshii, M. & Flaitz, J. (2002). Second language incidental vocabulary retention: The effect of text and picture annotation types. *CALICO Journal*, 20(1), 33-58.
- [27] Yoshii, M. (2006). L1 and L2 glosses: Their effects on incidental vocabulary learning. *Language Learning & Technology*, 10, 85-101.
- [28] Zareva, A., Schwanenflugel, P., & Nikolova Y. (2005). Relationship between lexical competence and language proficiency: Variable sensitivity. *Studies in Second Language Acquisition*, 27, 567–595.
- [29] Zhao Longwu and Yang Jie. (2010). Incidental Vocabulary Acquisition and the Corresponding Lexical Knowledge Base---On Incidental Vocabulary Acquisition in a Second Language (1). *Foreign Language Research*, 144-147.
- [30] Zhou, Songbo. (2010). Comparing receptive and productive academic vocabulary knowledge of Chinese EFL learners. *Asian Social Science*, 6(10), 14-19.

**Xiaoling Liu** was born in Ningxiang, China in 1964. She received her TESOL certificate from City University of New York and master's degree from Hunan University, China in 1987. She is currently a full professor in the College of Foreign languages, Hunan University, Changsha, China. Her research interests include Applied Linguistics, Stylistics, and Foreign language Teaching and Learning.

**Yang Gao** was born in Zhuzhou, China in 1986. She received her master's degree from Hunan University, China in 2012. She is currently an engineer and translator in AECC Hunan Aviation Powerplant Research Institute, Zhuzhou, China.