The Effects of MultimediaAnnotations on
the Vocabulary Retention of EFL Readers

Jui-chen Wu
Providence University
jcwu@pu.edu.tw

Hsi-chin Chu
National Taiwan Normal University
hcchu@ntnu.edu.tw

Abstract
The purpose of this study is to compare the effects of three modes of vocabulary annotation on immediate and delayed vocabulary retention using an on-line reading text. This study adopted a four-group pretest, immediate posttest, and delayed posttest experimental design. Twelve target words were presented in a 331-word reading text under four conditions in a computer program: (1) control with no annotation, (2) text annotation only, (3) picture annotation only, and (4) combined text-picture annotations. One hundred and sixty-nine high school students were randomly assigned to one of the four groups. All of the participants took a vocabulary recognition pretest in the first week. A week later, before taking a vocabulary recognition immediate posttest, they read a story on-line with words annotated in one of the four conditions. Two weeks later, they took a delayed posttest. ANCOVA analyses revealed that students in all three annotation groups attained significantly better gains of vocabulary on the immediate posttest and also on the delayed posttest than the control group, indicating that annotation in on-line text leads to vocabulary retention. However, a post hoc comparison of the three annotation groups demonstrated that there was no significant difference among the groups on both the immediate and delayed posttest, showing that the three modes of annotation yield similar effects on vocabulary gains through on-line reading.

Key Words: multimedia annotations, vocabulary retention
INTRODUCTION

The advent in computer technology has had a profound impact on language learning. Although the use of hardcopy textbooks and paper-and-pencil exercises are still the norm in traditional EFL classrooms, alternative texts presented through multimedia are provided either on disk or on the Internet to enable students to work independently. These texts generally come with textual, audio, or graphic annotations, which assist comprehension through multi-sensuous input. The most common type of annotation is vocabulary annotation in the reading texts. Such annotation enables students to acquire vocabulary through both top-down and bottom-up approaches. Students might guess the meaning through context and meanwhile check the definition provided in the annotation; conversely, they may look up the annotated definition and then map the meaning in the context. Using annotation on-line consumes little energy and time with respect to seeking the definition of an individual word and allows a smooth connection of textual meaning, thus increasing reading speed and study time (Leffa, 1992; Reinking & Rickman, 1990; Roby, 1999). Therefore, a computerized format for dictionaries and annotations is considered a more efficient method of learning vocabulary than a paper format; it is also preferred by students (Aust, Kelly, & Roby, 1993; Reinking & Rickman, 1990; Roby, 1999).

Despite the merits of vocabulary annotation in reading computerized texts, it is important to understand the effects of annotation in different presentation modes—text, picture, or text-picture combined—on vocabulary learning for practical concerns. Practitioners in the field need to be informed of the effectiveness of
each mode in order to optimize the function of the computer tool in assisting learning. Hence, the present study examines the effect of reading computerized text with three types of multimedia annotations on the vocabulary learning of EFL high school students.

**Theoretical Background**

There are two theories—dual coding theory and a generative theory of multimedia learning—that are related to the current investigation. Dual coding theory, proposed by Paivio (1986), is based on the assumption that information is stored in long-term memory in two systems: the verbal and nonverbal systems. The verbal system comprises printed words, speech sounds, and motor feedback from writing; the nonverbal system includes pictures, objects, and environmental sounds. The two systems are independent and one can be activated without the other. However, they may also be linked. This theory suggests that learning effects can be reinforced when a person is exposed to more than one code. Supported by the dual coding theory, multimedia presents (non)verbal codes to learners and thus enriches learning. As Naijar (1996) states, it is “…better when information is…processed through two channels than when the information is processed through one” (p.134). Readers are believed to benefit more from the interaction of multi-modal multimedia annotations.

Generative theory of multimedia learning (Mayer, 1997) elaborates on dual coding theory, assuming that L2 learners have two separate verbal systems, L1 and L2, plus a common imagery system. Dual coding theory views learning and cognition as active, dynamic processes. Meaningful learning is created via cognitive routes that use processes of selection, organization, and integration. It occurs
…when students select relevant visual and verbal information and store them in visual and verbal working memory respectively, where they are organized in coherent verbal and visual representations. If both types of information are present in working memory at the same time, connections can be made between corresponding visual and verbal representations in the stores. The information and the connections between them can then be integrated into a user’s mental model and thus can enhance learning (Jones & Plass, 2002, p.549).

It is therefore postulated that annotation in both picture and text modes might enhance vocabulary learning more than in the case where they are separately presented because their simultaneous presentation integrates the information from two routes and thus creates robust representation in memory.

**Previous Studies on Multimedia Annotation and Vocabulary Learning**

Several studies that have investigated different modes of annotation have paved the way for the current study.

Chun and Plass (1997) developed a multimedia program called *CyberBuch* and conducted a series of studies that compared the effect of three modes of multimedia annotations—text, text-picture, and text-video—on the recall of reading and vocabulary learning of American university students learning German as a second language (Chun & Plass, 1996a, 1996b, 1997). Students watched a video outlining the story and then read the annotated story before performing a summary recall and taking a vocabulary test. The results indicated that words annotated with the combination of text and picture and the combination of text and video proved more effective than words annotated with text only. While their studies did
reflect the effect of dual coding (Paivio, 1986), it is unknown whether individual types of annotation, namely, picture-only and video-only annotations, would yield the same effect.

Yoshii and Flaitz (2002) investigated this issue. Following a paper-based probe of single annotation that facilitates L2 learning by Kost, Foss, and Lenzini (1999), Yoshii and Flaitz studied three multimedia annotations—text, picture, and text-picture—in university ESL students’ short- and long-term vocabulary learning through online reading. Proficiency level, that is, beginning or intermediate, and the available annotation types were compared on immediate and delayed posttests, which included picture recognition, word recognition, and a definition supply test. The results showed that the text-picture combination outperformed the others on the immediate and delayed posttests, but the differences were smaller on delayed posttests. No significant difference was found between proficiency levels in the immediate or delayed posttests. Retention scores decreased in the delayed posttests for all groups.

While both Chun and Plass (1997) and Yoshii and Flaitz (2002) used L2 (English) as the language for annotation, it is unknown whether using L1 instead of L2 would make a difference on the effects of annotation. In Taiwan, Chen (2002) investigated the effects of L1 and L2 text annotation on the reading comprehension and vocabulary retention of 85 college EFL students. After reading a short expository news text on a computer, students took a multiple-choice posttest with 15 vocabulary and 5 reading comprehension items. The findings indicated online annotation as beneficial for second language learning; students with L1 or L2 annotation in experimental groups outperformed those in the control group (without annotation) on the posttest, and the L1 annotation
group scored significantly higher than those in the control group in vocabulary retention.

In sum, the research on effectiveness of multimedia annotations through reading on L2 vocabulary acquisition seems positive. Yet, several issues remain largely untouched. The studies reviewed in this paper focus on collegians; it is unknown whether younger L2 learners would show a similar impact on vocabulary retention. In particular, previous studies examine the effects of annotation format by comparing textual features as well as audiovisual features on vocabulary learning (Chen, 2002; Chun & Plass, 1996a, 1996b, 1997; Yoshii & Flaitz, 2002). These studies, which were conducted in western countries, suggest that combined text-picture annotation results in enhanced vocabulary memorization. Although studies have been conducted to investigate the effects of hypertext on university EFL students’ comprehension and vocabulary acquisition in Taiwan (e.g., Chen, 2002), few have investigated the effects of multimedia annotation on Taiwanese EFL students, particularly senior high school students. It is still unclear whether text versus picture annotation would yield similar effects on students at this level. This study was thus conducted to explore the effects of three modes of vocabulary annotations in a computerized reading text—text, picture, and text-picture combined—on the vocabulary retention of senior high EFL students.

**RESEARCH QUESTIONS**

To gauge the effects of using multimedia annotations on vocabulary learning through on-line reading, this study aims to
answer the following questions:

1a. Will Taiwanese senior high school students using annotations in experimental groups (text, picture, or text-picture combination) outperform those who do not use annotations (control group) on the immediate vocabulary posttest?

1b. Will Taiwanese senior high school students using text annotation perform differently than readers using picture annotation on the immediate vocabulary posttest?

1c. Will Taiwanese senior high school students using annotations in the text-picture combination perform better on the immediate vocabulary posttest than students using text or picture annotations?

2a. Will Taiwanese senior high school students using multimedia annotations in experimental groups outperform the control group on the delayed vocabulary posttest?

2b. Will Taiwanese senior high school students using text annotation perform differently than readers using picture annotation on the delayed vocabulary posttest?

2c. Will Taiwanese senior high school students using multimedia annotations in text-picture combination perform better on the delayed vocabulary posttest than the students using text or picture annotations?

3. Is there a retention difference over time (between the immediate and delayed posttests) for distinct groups?
METHOD

Pilot Study to Determine Text and Target Words

To select text and target words suitable for the main study, a pilot study was conducted on a class of 37 first-year students from the same senior high school as in the main study during the first semester of the 2002–2003 academic year. Two steps were involved: selecting text and determining target words for annotation.

To select a suitable text, two texts with readability levels at 3.4 and 6.7 on the Flesch-Kincaid Grade Level, were selected for 12 students. These texts fell within the readability range of the textbook that the students were using at that time (2.6 to 8). The 12 students rated the difficulty levels on a five-point Likert scale, ranging from 1 (very difficult) to 5 (very easy). The texts scored means of 2.8 and 1.9. The easier text was chosen.

To determine the target words in the chosen text for annotation, three steps were taken. First, a classroom teacher, based on his or her experience with the students, circled words that might have been unfamiliar to the students, thus yielding eleven words. Then, 37 students rated the words using a familiarity checklist that contained these eleven words as well as nine distractors that were already learned. They rated the words on a four-level scale adapted from Scarcella and Zimmerman’s design (1998): (1) “I don’t know the word,” (2) “I have seen the word before but am not sure of the meaning,” (3) “I understand the word when I see or hear it in a sentence, but I don’t know how to use it in my own speech or writing,” (4) “I can use the word in a sentence” (p.36). In addition, students were required to write down the Chinese meaning of the English word if they answered (3) or (4) for any item. Words rated as
(3) or (4) by more than two students and defined correctly were excluded from the list. At this stage, three of the eleven words were excluded, leaving eight target words. Third, students were asked to read the text and circle the words that they did not understand. Apart from the eight remaining words, four were added to the list of unknown words because they were circled as unknown and were incorrectly translated by at least 30 students. These words were also cross-referenced with the word list in the students’ textbook in order to ensure that no such vocabulary had been explicitly taught.

In sum, the text chosen was a 331-word story with a readability score of 3.4 on the Flesch-Kincaid Grade Level and mid-difficulty level ($M = 2.8$). Twelve target words were unknown to most students (five nouns, four verbs, and three adjectives): *armour, bank, demon, oven, powder, flow, grind, hide, sail, clever, drown, and lying*. The text, a tale entitled *The King and the Turtle*, was selected from the book series *Childcraft: The How and Why Library* (see Appendix A).

**Participants**

One hundred and sixty-nine first-year students—77 males and 92 females—from a senior high school in central Taiwan participated in this study. Each participant was randomly assigned to one of four treatment groups: control group, annotated text group, annotated picture group, or combined text-picture annotation group.

To ensure that the four condition groups were similar in terms of vocabulary knowledge, every student took a Nation’s Vocabulary 1000 Level Test (Voc1000) at the outset (Nation, 1990), a week before the experimental study. In total, 153 students completed the test; the grand mean was 66.28 ($SD = 10.05$). The mean scores for each group are as follows: Control: $M = 68.44$ ($SD = 8.69$); Text: $M =$
63.68 ($SD = 10.73$); Picture: $M = 68.29$ ($SD = 11.57$); and Combination: $M = 64.83$ ($SD = 8.75$). According to Nation’s vocabulary level guidelines, the students’ vocabulary ranks at the 1000-word frequency level. A one-way analysis of variance (ANOVA) of the groups categorized as independent variables and the Voc1000 scores as dependent variables indicated that the four groups were equivalent in terms of performance on the Voc1000 test, $F (3, 149) = 2.26, p > .05$.

**Instrumentation**

Three types of instruments were used during data collection: a story with three types of on-line annotations, three forms of a vocabulary recognition test, and a reading recall prompt. The former was a computerized text while the latter two had a pen-and-paper format.

**Hypertext design with multimedia annotations.** A reading text containing 12 target words determined in the pilot study was created as hypertext using Visual Basic 6.0 Service Pack 5. The story, *The King and the Turtle*, was presented on five pages on-screen, with the words highlighted for all the four groups and annotated for the three experimental groups in three modes: text, picture, or text-picture combination. Any experimental group student could click on highlighted words and read annotations in one of the three modes. All the highlighted words come with an audio component that serves a multimedia function (Abraham, 2001; Al-Seghayer, 2001; Chun & Plass, 1996a). The recording of a young female native English speaker reading aloud the words was used as audio for all of the highlighted words. Hence, in the control condition, students could click the target words and listen to them read aloud. However, in the experimental condition, for any word that was clicked, annotation
information would appear in a smaller window below the reading passages on the same screen. In addition to audio information, the text information includes the Chinese translation, part of speech, and tense; the picture information included a photo or picture denoting the word.

During the experiment, students had unlimited access to the annotation available for their group. In sum, the annotation information for the control group was audio only; for the text group, audio and text; for the picture group, audio and picture; and for the combination group, audio, text, and picture. Figure 1 presents the actual screen that appears when a student in the combination group clicked on the target word *sailed* for comprehension. Annotation information for the word appears in the lower window and includes the audio, text, and picture modes (shown as a, b, and c respectively).

![Screen Sample of Selected Multimedia Annotation in Text-picture Combination Group](image)

**Figure 1**  
Screen Sample of Selected Multimedia Annotation in Text-picture Combination Group (a, audio; b, text; c, picture annotation)
Navigation controls—forward or backward to next or previous page—were also created on the corners of the screen, and a module was designed to ensure ease of use so that the learners could work independently.

**Vocabulary recognition test.** To assess participants’ vocabulary knowledge before the treatment and immediate and long-term retention after the treatment, three 20-item, multiple-choice recognition tests were designed for three rounds. The 20 items contained questions on the 12 target words and 8 distractors (see Appendix B). The order of questions and choices in each test was rearranged; the distractors also differed in each test.

A question contained five answer choices; one correct Chinese translation answer choice; three Chinese distractor choices; and a fifth option, “I don’t know.” The criteria used for choosing the three Chinese distractor words were as follows: one of the other 19 words tested, phonetically similar words, and semantically similar words. The fifth option, “I don’t know,” was included to reduce the probability of guessing correct answers. One point was assigned for each correct answer.

Although all the target words were confirmed as unknown in the pilot study, two words, an adjective *lying* and a noun *bank*, had to be excluded from final analysis because during the three-month period between the pilot study and the experimental treatment, these words were learned from their assigned outside reading. Thus, ten target words remained that were usable for an analysis of the results, changing the highest score of each test to 10 instead of 12.

**Reading recall.** This task, administered after reading and before the vocabulary test, was performed for two purposes: (1) to direct students’ attention to comprehension instead of vocabulary
memorization and (2) to serve as a disruption in the possibility that students kept vocabulary annotation in their working memory immediately after reading. Reading recalls were not analyzed because they functioned as a distracting task.

**Design**

To assess the effects of multimedia annotation on L2 word acquisition through reading, this study used a pretest—immediate posttest—delayed posttest experimental design. Each participant was assigned to one of the three experimental groups—text, picture, and text-picture combination—and one control group.

The pretest vocabulary scores served as a baseline for control. L2 learners’ scores on immediate and delayed posttests of vocabulary were examined. Students in the text group (hereafter Text group) could only look up the 12 target words annotated with the L1 translation. Students in the picture group (hereafter Picture group) could only look up the 12 target words provided with picture annotations. Students in the text-picture combination group (hereafter Combination group) could only look up the 12 target words provided with text and picture annotations. Control group students (hereafter Control group) could not look up any text or picture information from the 12 target words. Students could click on a given annotation as many times as they wished.

**Data Collection Procedure**

The data for the study was collected in three sessions over four weeks.

*Session one.* In the first week, students took a 10-minute vocabulary recognition pretest during normal class time.
Session two. During the second week, session two was conducted in a computer room. In the first 50-minute class period, students performed a practice session for 10 minutes, read a story on a computer for 20 minutes, and completed a written recall in the last 20 minutes. Subsequently, in the second period, students completed an immediate vocabulary recognition posttest.

In the practice session, all the students logged into the program to practice one annotated paragraph, *The Blind Men and the Elephant*, from their textbook reading. They were encouraged to click the highlighted words with the mouse and learn the meaning by reading annotations appearing on the lower part of the screen.

After a 10-minute practice, all students began reading the target text on the computer for approximately 20 minutes. They were told that they needed to recall the story in Chinese after reading the text for comprehension. However, they were not informed that they needed to complete a vocabulary test after the recall test because the purpose of the study was to investigate the incidental vocabulary learning effect, and they were urged to look up the available annotations. After reading, the students turned off the computer program and wrote down the story recall. The recall data were not analyzed and are not reported in this paper.

Following a ten-minute break, during which students were instructed not to discuss the treatment with each other, an immediate vocabulary test was administered in the second period (lasting 10 minutes).

Session three. Two weeks later, students, without advance notification during their normal class time, completed a delayed vocabulary recognition posttest.
Analyses of Vocabulary Test

The vocabulary task was scored by giving points for students’ correct answers; one point was given for one correct answer for each of the 10 target words. Five out of the 169 participants who were unable to complete all three vocabulary tests were excluded from the analysis, leaving 164 scores. Two separate one-way analyses of covariance (ANCOVAs) were run on each immediate posttest and delayed posttest, with pretest scores set as the covariate, and the treatment group conditions (Control, Text, Picture, and Combination) set as the independent variable. Following ANCOVA, post hoc comparisons using the Scheffé test were compared to reveal the differences between each possible group pair. Then, paired t-tests were conducted to analyze any significant change in retention over time (between immediate and delayed posttests) among individual groups. The alpha level was set at .05 for the analyses in this study.

FINDINGS

The findings are presented in response to the three major research questions and their subsets.

Immediate Vocabulary Retention

This section reports the answers to research questions 1a, 1b, and 1c on immediate vocabulary retention.

Effects of annotations. As illustrated in Table 1, the ANCOVA results on the effects of all annotation modes on an immediate vocabulary posttest were significant \( F(3, 159)^a = 16.00, \)
which implies a difference in immediate retention among groups. The pretest, as the covariate, also contributed to test results \[ F (3, 159)^a = 5.03, p < .05 \], showing that it is necessary to control the variance in the pretest for the analyses of the immediate posttests.

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest</td>
<td>(1, 159)^a</td>
<td>5.03*</td>
</tr>
<tr>
<td>Group</td>
<td>(3, 159)^a</td>
<td>16.00***</td>
</tr>
</tbody>
</table>

Note. *p < .05, ***p < .001.

The adjusted mean scores of the immediate vocabulary posttest by groups (Table 2), with the pretest as the covariate, indicate the lowest student performance in the Control group (\( M = 3.77^a \)), higher scores for the experimental groups, Picture (\( M = 6.41^a \)) and Text (\( M = 6.54^a \)), and the highest score in the Combination group (\( M = 6.82^a \)).

<table>
<thead>
<tr>
<th>Group (( N = 164 ))</th>
<th>Control (( n = 44 ))</th>
<th>Text (( n = 41 ))</th>
<th>Picture (( n = 38 ))</th>
<th>Comb. (( n = 41 ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>3.77^a (.35)</td>
<td>6.54^a (.36)</td>
<td>6.41^a (.38)</td>
<td>6.82^a (.36)</td>
</tr>
</tbody>
</table>

Note. Standard errors are in parentheses. Comb. = text-picture combination. Mean scores are adjusted with pretest = 2.09.
Table 3 presents the results of the post hoc ANCOVA for the immediate vocabulary posttest (with adjusted mean from the covariate pretest = 2.09) to control for Type III error across four pairwise comparisons. The post hoc analysis of the immediate vocabulary posttest revealed the Text, Picture, and Combination groups performed significantly better than the Control group. This demonstrates that participants in experimental conditions (Text, Picture, and Combination) outperformed those in the control condition on immediate posttests ($p < .001$), indicating that students retain vocabulary more effectively after reading texts with some form of annotation.

Comparison of the effects of text versus picture annotation. Table 3 also shows no significant difference in word retention in the immediate posttest between the Text and Picture groups (mean difference $= .13$), suggesting that whether vocabulary is annotated with text or picture makes no difference in students’ short-term vocabulary retention.

Table 3
Pairwise Comparison of Immediate Posttest Means Among the Four Groups

<table>
<thead>
<tr>
<th>Group</th>
<th>Control ($n = 164$)</th>
<th>Text ($n = 164$)</th>
<th>Picture ($n = 164$)</th>
<th>Comb. ($n = 164$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>-2.77 a (.51)***</td>
<td>-2.64 a (.52)***</td>
<td>-3.05 a (.51)***</td>
<td></td>
</tr>
<tr>
<td>Text</td>
<td>2.77 a (.51)***</td>
<td>0.13 a (.52)</td>
<td>-0.28 a (.51)</td>
<td></td>
</tr>
<tr>
<td>Picture</td>
<td>2.64 a (.52)***</td>
<td>-0.13 a (.52)</td>
<td>-0.41 a (.52)</td>
<td></td>
</tr>
<tr>
<td>Comb.</td>
<td>3.05 a (.51)***</td>
<td>0.28 a (.51)</td>
<td>0.41 a (.52)</td>
<td></td>
</tr>
</tbody>
</table>

Note. Comb. = text-picture combination. Mean differences are adjusted for multiple comparisons. *** $p < .001$. 

---

Wu & Chu: Effects of Multimedia Annotations
Comparison of the effect of text-picture annotation with that of text annotation and picture annotation. With regard to the effect of text-picture versus text annotation, Table 3 also shows no significant difference in word retention between text-picture and text annotation (mean difference = –.28), indicating that whether text is annotated with text-picture or text makes no difference in students’ vocabulary retention. In addition, there is no significant difference in vocabulary retention between text-picture and picture annotation (mean difference = –.41), revealing that whether text is annotated with text-picture or picture only makes no difference in students’ vocabulary retention.

Delayed Vocabulary Retention
This section answers research questions 2a, 2b, and 2c on delayed vocabulary retention.

Effects of multimedia annotations. As illustrated in Table 4, the ANCOVA result for the effect of the four annotation modes on a delayed vocabulary posttest was significant \[ F(3, 159)^a = 6.85, p < .001 \], revealing a difference in vocabulary delayed retention among groups. The pretest as the covariate \[ F(3, 159)^a = 13.40, p < .000 \] also significantly contributed to the test results, showing a need to control the variance in the pretest for the four groups on the delayed posttests.

Table 4
ANCOVA Results for Delayed Vocabulary Posttest by Group

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest</td>
<td>(1, 159)^a</td>
<td>13.40***</td>
</tr>
<tr>
<td>Group</td>
<td>(3, 159)^a</td>
<td>6.85***</td>
</tr>
</tbody>
</table>

Note. ***p < .001
The adjusted mean scores of the delayed posttest on vocabulary by group (Table 5), with the pretest as the covariate, show a trend similar to that of the immediate posttest on vocabulary. Students in the Control group attained the lowest mean score ($M = 3.50^a$); those in the Text group and Picture group, higher scores ($M = 4.87^a$ and $M = 5.06^a$ respectively); and those in the Combination group, the highest score ($M = 5.27^a$).

Table 5
Adjusted Mean Scores of Delayed Vocabulary Posttest by Group

<table>
<thead>
<tr>
<th>Group (N = 164)</th>
<th>Control (n = 44)</th>
<th>Text (n = 41)</th>
<th>Picture (n = 38)</th>
<th>Comb. (n = 41)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M$ (SE)</td>
<td>3.50$^a$ (.30)</td>
<td>4.87$^a$ (.31)</td>
<td>5.06$^a$ (.33)</td>
<td>5.27$^a$ (.31)</td>
</tr>
</tbody>
</table>

Note. Standard errors are in parentheses. Comb. = text-picture combination. Mean scores are adjusted with pretest = 2.09.

Table 6 depicts the post hoc ANCOVA for the delayed vocabulary posttest (adjusted mean from covariate pretest = 2.09), which controls for the Type III error across the four pairwise comparisons. Post hoc analysis of vocabulary delayed posttest showed that the Text, Picture, and Combination groups performed significantly better than the Control group. The results showed that students in experimental conditions (text, picture, and combination) outperformed those in the control condition on the delayed posttests ($p < .001$). This indicates that students retain more vocabulary after reading texts with annotations than texts without.

Comparison of the effect of text versus picture annotation. Table 6 also shows that there is no significant difference in vocabulary retention in the delayed posttest between the reading of
text with text annotation and the reading of text with picture annotation \( (\text{mean difference} = -.19) \), revealing that whether the vocabulary was annotated with text or picture makes no difference in students’ long-term vocabulary retention.

**Table 6**

Pairwise Comparison of Delayed Posttest Means Among the Four Groups

<table>
<thead>
<tr>
<th>Group</th>
<th>Control ((n = 44))</th>
<th>Text ((n = 41))</th>
<th>Picture ((n = 38))</th>
<th>Comb. ((n = 41))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>-1.37* (.44)**</td>
<td>-1.56* (.44)***</td>
<td>-1.78* (.43)***</td>
<td></td>
</tr>
<tr>
<td>Text</td>
<td>1.37* (.44)**</td>
<td>-0.19* (.45)</td>
<td>-0.41* (.44)</td>
<td></td>
</tr>
<tr>
<td>Picture</td>
<td>1.56* (.44)***</td>
<td>0.19* (.45)</td>
<td>-0.22* (.45)</td>
<td></td>
</tr>
<tr>
<td>Comb.</td>
<td>1.78* (.43)***</td>
<td>0.41* (.44)</td>
<td>0.22* (.45)</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Comb. = text-picture combination. Mean differences are adjusted for multiple comparisons. **\( p < .01 \). ***\( p < .001 \).

**Comparison of the effect of text-picture annotation with the effect of text annotation and that of picture annotation.** With regards to the effects of text and picture annotations, Table 6 reveals no significant difference in long-term word retention between text-picture and text annotation \( (\text{mean difference} = .05) \), showing that whether text is annotated with text-picture or text has no effect on long-term retention. Further, no significant difference appeared in long-term retention between text-picture and picture annotation \( (\text{mean difference} = .22) \), indicating that whether text is annotated with text-picture combination or picture makes no difference in students’ vocabulary retention.

**Difference in Retention Between the Two Posttests Within**
Groups

This section answers research question 3 on the retention difference over time (between the immediate and delayed posttests) for individual groups.

Table 7 shows the retention differences within groups (the adjusted means of the delayed posttests minus those of the immediate posttests) as manifested in the decreasing scores on the delayed tests being larger for the annotation groups than for the control. Table 7 shows that there was retention loss in every group: Control–0.25, Text–1.68, Picture–1.37, and Text-picture–1.54. The Control group had the lowest retention loss; the Text group, the highest. Therefore, a t-test was run to test the difference in loss between the delayed and immediate posttests within each group. Figure 2 shows the mean difference tendency for all the groups on the two posttests. The Control group evidently differed from the other groups in amount of vocabulary lost over time.

Table 7
Retention Difference Between the Posttests Within Groups

<table>
<thead>
<tr>
<th>Condition</th>
<th>Pair</th>
<th>Mean</th>
<th>SD</th>
<th>df</th>
<th>t</th>
<th>p (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control (n= 44)</td>
<td>Impost-Depost</td>
<td>.25</td>
<td>1.91</td>
<td>43</td>
<td>.87</td>
<td>.389</td>
</tr>
<tr>
<td>Text (n = 41)</td>
<td>Impost-Depost</td>
<td>1.68</td>
<td>1.78</td>
<td>40</td>
<td>6.05</td>
<td>.000***</td>
</tr>
<tr>
<td>Picture (n= 38)</td>
<td>Impost-Depost</td>
<td>1.37</td>
<td>1.65</td>
<td>37</td>
<td>5.11</td>
<td>.000***</td>
</tr>
<tr>
<td>Comb. (n= 41)</td>
<td>Impost-Depost</td>
<td>1.54</td>
<td>1.67</td>
<td>40</td>
<td>5.88</td>
<td>.000***</td>
</tr>
</tbody>
</table>

Note. Comb. = text-picture combination. Impost = immediate posttest; Depost = delayed posttest. *** p < .001.

The results of the t-test in Table 7 indicate that there was no
significant loss in vocabulary retention within the Control group ($t = .87, p > .05$) between immediate and delayed vocabulary recognition tests, while significant loss occurred from immediate to delayed scores under each experimental condition (Text: $t = 6.05, p < .001$; Picture: $t = 5.11, p < .001$; Combination: $t = 5.88, p < .001$). While the Control group students retained 92.8% of the lexical items they had recognized on the immediate posttest (3.77 vs. 3.50), the annotation groups had a greater vocabulary loss over two weeks. The Text group retained 74.5% (6.54 vs. 4.87); the Picture group, 78.9% (6.41 vs. 5.06); and the Combination group, 77.3% (6.82 vs. 5.27).

Figure 2
Mean Differences for Groups on Immediate and Delayed Posttests

In particular, Figure 2 also shows that there is a tendency toward
interaction between the Text and Picture groups, albeit an insignificant one. During the immediate posttest, the Text group retained more words than the Picture group. However, the Text group scored less than the Picture group on the delayed posttest. In other words, the Text group had higher memory loss of vocabulary than the Picture group.

**DISCUSSION**

**The Effects of Multimedia Annotation on Vocabulary Retention**

First, it was found that irrespective of which multimedia mode—text, picture, or text-picture—vocabulary is presented, there is an effect on immediate and delayed vocabulary retention. This finding is consistent with previous studies reported in the literature, in that annotation has a positive effect on vocabulary learning (Abraham, 2001; Laufer & Hill, 2000).

Two factors contribute to the finding that vocabulary retention was significantly better in the three experimental groups than the control group. The first factor is annotation. The annotated words in the text were salient for learners because the words were highlighted and verifiable with immediate help for meaning support, thereby reducing mistakes and incorrect guesses. This finding is supported by the claim of other studies that L2 vocabulary learning is enhanced by the use of annotations (Hulstijn, Hollander, & Greidanus, 1996; Jacobs, Dufon, & Fong, 1994; Kost et al., 1999).

The second factor is the effect of multimedia annotations. The higher retention rate in the experimental groups of the study reaffirmed the findings of previous studies that multimedia programs
help students engage in more active learning and facilitate higher scores on vocabulary posttests (Chun & Plass, 1996a). In addition, the indication that hyperlinks for target words do not interfere with readers’ reading flow concurs with previous studies (Davis, 1989; Davis & Lyman-Hager, 1997; Martínez-Lage, 1997; Nation, 2001). Finally, our study indirectly supported that click-then-appear annotated vocabulary information accommodates readers’ instant need for vocabulary assistance (Nation, 2001), increases reading speed (Leffa, 1992; Reinking & Rickman, 1990; Roby, 1990), eases anxiety, and increases readers’ preference (Martínez-Lage, 1997).

Effects of Text and Picture Annotations

This study found that there is no difference between the Text group and Picture group in short- and long-term vocabulary retention. This finding echoes Sutton’s (1999) study that investigated the effects of L1, L2 translation, and image annotation on vocabulary learning, which also found no significant difference between text and image groups in vocabulary learning. However, this study further expands on this finding, providing evidence of a subtle interaction between the two variables across time, albeit an insignificant one. The Text group retained vocabulary slightly better than the Picture group on the immediate posttest, albeit insignificantly, and subsequently regressed to retention less than the Picture group on the delayed posttest (Figure 2). This showed the different functions that text and picture have in our memory. Concepts portrayed through pictures are better remembered in the long run than words in text. This is in agreement with previous findings that pictures tend to be remembered better over time, whereas words are more easily forgotten (Chun & Plass, 1996a, 1997). This is referred to as the
“hypermnesia effect,” according to which “good static images are recalled better over time, whereas words and transient images tend to be forgotten” (Read, 2000, p.50). The findings of this study reflect such an effect in that pictured information assisted in vocabulary retention for a longer time, though no significant difference was found.

**Effects of Combined Text-picture Annotation**

The Combination group did not outperform either the Text or Picture groups in this study. This can be attributed to four factors: picture denotation difficulty, information redundancy, ceiling effect, and type of dependent measure.

First, it is likely that not all pictures used in this study accurately depicted the words and thus interfered with the storyline. Researchers proposed that exposure to language input must be made comprehensible in order to facilitate acquisition (Krashen, 1985; Long, 1985). It is probable that not all pictures in this study provided a facilitative “input device for language development” with the storyline (Hanson-Smith, 1999, p.190). It is probable that while students tried to comprehend the story, neutral images from the pictures might not facilitate—and perhaps even hampered—understanding as the images may have become redundant information.

Second, combining text and picture annotations might produce redundant information. For the senior high school students in this study, either text or picture annotation separately might have been adequate for incidental vocabulary learning in the task. Parallel presentation of word meaning in two modes, text and picture, might yield no additive effect because it is redundant. An extra picture or text to denote meaning makes no difference in vocabulary retention.
Third, vocabulary tests consisting of a mere ten words may result in a ceiling effect in the test results, proving that they are not sensitive enough to accurately differentiate among treatment groups. The ceiling effect might be reduced if students are exposed to longer or multiple readings and tested on more words.

Another reason proposed by researchers is that picture support might benefit text comprehension rather than vocabulary retention because the latter mainly involves rote learning rather than cognitive processing (Plass, Chun, Mayer, & Leutner, 1998; Plass, Jones, & Leutner, 2002). This offers another possible explanation of why combined text-picture annotation was not significantly different from text or picture annotation in immediate or delayed vocabulary retention in this study. The dependent measure for the annotation effect, namely, test of vocabulary retention, might call for lower-level processing than test of reading comprehension, and thus it is not a valid measure.

Memory Loss Between Posttests

The three annotation groups yielded higher vocabulary scores on posttests and a higher rate of retention. This may be explained by a theory of “corroborative evidence” that postulates that there is a positive relationship between retention rates and depth of processing (Groot, 2001, p.77). For words to be learned and acquired for long-term retention, it takes active use of the words and intensive processing of new words. For experimental groups, it is predictable that words learned in a receptive way in such a short time, without repeated encounter and active use later on, would lead to a higher retention loss in the delayed posttest two weeks later.
IMPLICATIONS

Some implications relevant to multimedia design for language teaching and further studies are suggested here.

Implications for Multimedia Design

Computer technology is a novel way of teaching and learning but not the only way. How, when, and in what way it can provide the maximum benefit to students is an issue for teachers to consider.

Teachers must be aware of and incorporate different media into their teaching to improve classroom instruction. Meanwhile, it is important to note the design of information presentation in terms of diverse modes for different types of assistance (Chun & Plass, 1997). For example, audio assistance for listening, pictorial aids for objects, and use of video for a higher-level contextual support should be incorporated into different instructional contexts. Moreover, with innovative support from computer technology, the influences of the social, cultural, and educational contexts should also be taken into consideration (Warschauer, 1998).

Undoubtedly, further research is needed in this area to ensure well-designed and more productive utilization of multimedia programs in teaching specific aspects of language. We believe that well-designed programs meet students’ individual needs, stimulate interest, and increase motivation.

Implications for Future Research

In this study we examined how multimedia annotations could benefit students’ short- and long-term vocabulary retention through reading. Overall, senior high school students retained words more
effectively with access to multimedia annotations. However, this study still suffers from some limitations.

First, we only investigated a small group of students during a limited period of time with limited materials. The results might not be the same with subjects of different ages or those from different locations. Moreover, this study did not take into account the differences in individual learning styles; in contrast, other studies have acknowledged that different types of learning styles, visualizers or verbalizers, benefit from modes of media differently. Therefore, future studies should address these issues by choosing students from different schools or by incorporating a wider range of materials and selecting students of different ages for comparison.

Second, this study was not a longitudinal survey tracing the contextual learning of vocabulary over time (Huckin & Haynes, 1993). Limited in capacity, this study did not incorporate multiple measures of short- and long-term vocabulary tasks into reading comprehension tasks to probe the different facets of retention on students’ vocabulary learning and reading comprehension. Furthermore, a longer duration of treatment and an increased number of tested vocabulary words would provide additional information for researchers on the effectiveness of treatment differences as well as help avoid the ceiling effect. In order to attain a broader cognitive and psychological understanding of multimedia annotation effects, the administration of a series of tests over a longer period of time is needed to ensure that students benefit from the program.

Finally, previous research has suggested that additive qualitative information, like interview or think-aloud methods, might help triangulate findings on the cognitive process or insight from learners (Abraham, 2001; Yoshii & Flaitz, 2002). Further studies addressing
language learning via computer-assisted technology should consider measures for a more comprehensive understanding of multimedia effects on second language teaching.

Though the findings in this study did not yield conclusive answers to the research questions, the work has added and contributed to the growing research in the area of multimedia reading and vocabulary learning with annotations.

ACKNOWLEDGEMENTS

This article is adapted from the MA thesis of the first author, who was supervised by the second author. We would like to thank Mr. Lin-Wen Chao for his assistance in developing the multimedia reading program used in this study. We would also like to express our gratitude to the two anonymous reviewers and the editors for their comments and suggestions on the drafts of this article.

REFERENCES


Hanson-Smith (Eds.), *CALL environments: Research, practice, and critical issues*. TESOL Inc.


Reinking, D., & Rickman, S. S. (1990). The effects of


**ABOUT THE AUTHORS**

Jui-chen Wu is a part-time lecturer in the Department of English Language, Literature, and Linguistics at Providence University. She conducts research on computer assisted language learning.

Hsi-chin Janet Chu is an associate professor in the English Department at National Taiwan Normal University. Her teaching and research interests include EFL reading, TEFL methodology, and teacher education.
APPENDIX A

Reading Text “The King and the Turtle”

The King and the Turtle\(^1\) ~ a tale from India

Long ago, a king lived in a great palace with beautiful gardens. In the gardens was a lake. The king’s small sons loved to play in the lake. They swam about and sailed their boats on it.

One day the king told his sons that he had ordered some fish to be put in the lake. The princes were very excited and ran to the lake to see the new fish. But there, as well as the fish, was a large green turtle. The boys had never seen a turtle before and were very frightened. They thought it must be a demon.

They ran back to the palace crying,” Father, father! There is a terrible demon lying on the bank of the lake!”

The king ordered his servants to catch the demon and bring it to him. When the turtle was brought in, the princes cried and hid behind the door.

The king did not like to see his sons so frightened, so he ordered that the turtle should be killed.

“But the demon is wearing armour,” said the servants. “How shall we kill him?”

“Grind him to powder,” said one. “Put him in a hot oven,” said another.

One after the other, the men thought of terrible ways to kill the turtle. One old man, who was afraid of water, suddenly spoke.

“Throw him into the lake,” he said, “just where the water flows over the rocks and into the river. Then he will surely be drowned.”

When the turtle heard this, he cried out, “How could you do such a cruel and heartless thing to me? Your other plans were dreadful, but this is worse. Please, please do not throw me into the lake!”

The king decided that because the turtle was more afraid of the lake than anything else, that was where he should be thrown.

And the clever turtle laughed as he swam down the river to his home in the sea.
APPENDIX B

Sample of Vocabulary Recognition Test

班級：_______ 座號：_____ 姓名：____________ 受測日期：92._______
作答說明：請從下列各題選項中選出一個適當的單字意義，完全不了解請回答( 5 )。

1. (   ) turtle:  (1) 魚 (2) 龜 (3) 加速 (4) 轉動 (5) 不知道
2. (   ) palace:  (1) 王子 (2) 安置 (3) 地方 (4) 皇宮 (5) 不知道
3. (   ) powder:  (1) 力量 (2) 粉末 (3) 磅 (4) 磨成 (5) 不知道
4. (   ) oven:  (1) 爐;灶 (2) 鍋 (3) 外殼 (4) 結束 (5) 不知道
5. (   ) lying:  (1) 站立 (2) 航行 (3) 飛翔 (4) 躺著 (5) 不知道
6. (   ) sail:  (1) 售出 (2) 尾巴 (3) 航行 (4) 流動 (5) 不知道
7. (   ) order:  (1) 爐;灶 (2) 訂購 (3) 模仿 (4) 較老的 (5) 不知道
8. (   ) demon:  (1) 天使 (2) 怪物 (3) 惡魔 (4) 國王 (5) 不知道
9. (   ) cruel:  (1) 天然的 (2) 哭泣的 (3) 吁喝的 (4) 殘忍的 (5) 不知道
10. (   ) hide:  (1) 打中 (2) 躲藏 (3) 顯示 (4) 碾碎 (5) 不知道
11. (   ) grind:  (1) 同意 (2) 烹煮 (3) 磨成 (4) 躲藏 (5) 不知道
12. (   ) afraid:  (1) 害怕的 (2) 緊張的 (3) 殘忍的 (4) 殘忍的 (5) 不知道
### Wu & Chu: Effects of Multimedia Annotations

<table>
<thead>
<tr>
<th>13. (   ) swim:</th>
<th>15. (   ) armour:</th>
<th>17. (   ) bank:</th>
<th>19. (   ) garden:</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) 航行</td>
<td>(1) 手臂</td>
<td>(1) 海灘</td>
<td>(1) 花園</td>
</tr>
<tr>
<td>(2) 游泳</td>
<td>(2) 衣服</td>
<td>(2) 湖泊</td>
<td>(2) 車庫</td>
</tr>
<tr>
<td>(3) 漂浮</td>
<td>(3) 盔甲</td>
<td>(3) 堤岸</td>
<td>(3) 侍衛</td>
</tr>
<tr>
<td>(4) 甜甜的</td>
<td>(4) 惡魔</td>
<td>(4) 儲水池</td>
<td>(4) 對不起</td>
</tr>
<tr>
<td>(5) 不知道</td>
<td>(5) 不知道</td>
<td>(5) 不知道</td>
<td>(5) 不知道</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>14. (   ) clever:</th>
<th>16. (   ) flow:</th>
<th>18. (   ) catch:</th>
<th>20. (   ) drown:</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) 可怕的</td>
<td>(1) 漂浮</td>
<td>(1) 抓住</td>
<td>(1) 拖曳</td>
</tr>
<tr>
<td>(2) 清潔的</td>
<td>(2) 流動</td>
<td>(2) 相配</td>
<td>(2) 流動</td>
</tr>
<tr>
<td>(3) 殘忍的</td>
<td>(3) 漂浮</td>
<td>(3) 殺死</td>
<td>(3) 淹沒</td>
</tr>
<tr>
<td>(4) 聰明的</td>
<td>(4) 淹死</td>
<td>(4) 釋放</td>
<td>(4) 畫圖</td>
</tr>
<tr>
<td>(5) 不知道</td>
<td>(5) 不知道</td>
<td>(5) 不知道</td>
<td>(5) 不知道</td>
</tr>
</tbody>
</table>
多媒體註解對英語為外語讀者詞彙記憶的影響

摘要
本研究旨在比較三種單字註解模式運用在線上閱讀，對中學生英語字彙的即時記憶及延遲記憶的成效。研究方法採用前測、立即後測及延遲後測組間實驗控制方法。研究者將一篇經前置研究所選出 331 字故事選出 12 個生字加上四種線上多媒體註解模組 - 無註解、文字註解、圖片註解及文字加上圖片註解，169 位高中學生被隨機分為對照組一組及實驗組三組 - 文字組、圖像組及文字＋圖像組。第一週受測者進行字彙前測，一週後進行線上多媒體閱讀，並在閱讀後進行字彙即時後測，二週後接受字彙延遲後測。使用 ANCOVA 統計檢視顯示每組實驗組兩次字彙後測成績表現均優於對照組，證明多媒體線上註解有助於字彙記憶。然而再以 Post hoc 統計比較三組實驗組組間即時及延遲字彙記憶成績表現均未達顯著差異，顯示三種多媒體註解運用在線上閱讀對字彙記憶表現相近。

關鍵詞：多媒體註解 字彙記憶