The Effects of Multimedia Annotations on L2 Vocabulary Immediate Recall and Reading Comprehension: A Comparative Study of Text-Picture and Audio-Picture Annotations under Incidental and Intentional Learning Conditions

By

Zhaohui Chen

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy
Department of Secondary Education
College of Education
and
Department of World Language Education
College of Arts and Sciences
University of South Florida

Major Professor: Tony Erben, Ph.D.
Jeffra Flaitz, Ph.D.
William Kealy, Ph.D.
John Ferron, Ph.D.

Date of Approval:
September 22, 2006

Keywords: cognitive, modality, attention, look-up, noticing, task type, ESL, dual-coding

© Copyright 2006, Zhaohui Chen
Dedication

In loving memory of my father

此文献给我亲爱的父亲陈锁法
Acknowledgements

I will always be grateful to my major professor, Dr. Tony Erben, for inspiring me and encouraging me to take up this project. Thank you for the many hours you devoted to discussing my progress, reading and responding to my numerous drafts, and giving me your insightful suggestions. Your support and guidance have made my academic journey at the University of South Florida an enjoyable and valuable one.

I was fortunate to have a committee of scholars who helped me to become a better researcher. Thank you to my other committee members: Dr. Jeffra Flaitz for her belief in me, for her understanding and encouragement; Dr. John Ferron for sharing his wisdom and providing me with invaluable feedback on statistical analysis; Dr. Bill Kealy for his sharp wit and expertise in multimedia learning.

I am thankful to the other faculty members in Second Language Acquisition and Instructional Technology program, particularly Dr. Carine Feyten, Dr. Joyce Nutta, Dr. Wei Zhu, Dr. Linda Evans and Dr. Jim White. I have appreciated and benefited from their instruction and guidance both academically and professionally.

I am also grateful to a number of colleagues and friends who shared their insights and gave me encouragements, especially when I was finishing my dissertation from a distance. Special thanks to Dr. Phil Smith for being generous, considerate and understanding; thanks to Aline Harrison, Rui Cheng, Annmarie Zoran and Iona Sarieva for their helpfulness, support and warm hugs; Thanks to Darunee Dujsik and Nader
Morkus for helping me with the date processing; thanks also to Martha Castaneda, Ruth Ban, Robert Summers, Michelle Macy, Jeannie Ducher, Farah Khorsandian-Sanchez and Li Jin for being supportive; and thanks to all the other SLAIT students and colleagues who have helped me one way or another.

My deepest appreciation also goes to the directors and instructors at the three English Language Institutes who have opened their classrooms to me and helped generously with the project. Among them are Christa Hansen, Cindy Kieffer, Jodi Webber, Ronda Sheffield, Lee Pettigrew, Norman Johnson. Sandy Greenstreet. I would also like to thank the students from the three English Language Institutes for participating in the project. Without them, this project would not have been possible.

Finally I would like to thank my family members in China for their support. My heartfelt appreciation is to my wife, Yifang Zhang, for her unconditional support and love.
# Table of Contents

List of Tables vi

List of Figures vii

List of Acronyms ix

Abstract x

Chapter I: Introduction 1
  Statement of the Problem 4
  Theoretical Framework 4
  Purpose of the Study 6
  Research Questions 6
  Hypotheses 7
  Significance of the Study 8
  Definitions of Terms 9
  Limitations and Delimitations 11
  Organization of the study 13

Chapter II: Review of the Literature 14
  L2 Vocabulary Learning 15
    Incidental and Intentional Learning 16
    Incidental Vocabulary Learning 18
  Annotation 20
    Traditional Annotation 21
    Multimedia Annotation 24
  Text Annotation 25
    Text Annotation in Printed Reading Texts 25
      Text Annotation and L2 Vocabulary Learning 25
      Text Annotation and L2 Reading Comprehension 29
    Text Annotation in Multimedia Texts 31
      Text Annotation and L2 Vocabulary Learning 31
      Text Annotation and L2 Reading Comprehension 33
    Summary 34
  Picture Annotation 35
    Annotation and L2 Vocabulary Learning 36
    Annotation and L2 Reading Comprehension 42
    Summary 43
  Video Annotation 44
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annotation and L2 Vocabulary Learning</td>
<td>44</td>
</tr>
<tr>
<td>Annotation and L2 Reading Comprehension</td>
<td>50</td>
</tr>
<tr>
<td>Summary</td>
<td>51</td>
</tr>
<tr>
<td>Audio Annotation</td>
<td>52</td>
</tr>
<tr>
<td>Annotation Studies Summary</td>
<td>56</td>
</tr>
<tr>
<td>Cognitive Theory of Multimedia Learning</td>
<td>57</td>
</tr>
<tr>
<td>Generative Theory of Learning</td>
<td>58</td>
</tr>
<tr>
<td>Dual-coding Theory</td>
<td>59</td>
</tr>
<tr>
<td>Cognitive Theory of Multimedia Learning</td>
<td>60</td>
</tr>
<tr>
<td>Extension of Cognitive Theory of Multimedia Learning</td>
<td>62</td>
</tr>
<tr>
<td>Application of Cognitive Theory of Multimedia Learning to Multimedia Annotation</td>
<td>64</td>
</tr>
<tr>
<td>Chapter Summary</td>
<td>67</td>
</tr>
<tr>
<td>Chapter III: Methods and Procedures</td>
<td>70</td>
</tr>
<tr>
<td>Design of the Study</td>
<td>71</td>
</tr>
<tr>
<td>Variables</td>
<td>74</td>
</tr>
<tr>
<td>Dependent Variables</td>
<td>74</td>
</tr>
<tr>
<td>Independent Variables</td>
<td>74</td>
</tr>
<tr>
<td>Types of Annotation</td>
<td>74</td>
</tr>
<tr>
<td>Types of Learning</td>
<td>75</td>
</tr>
<tr>
<td>Other Independent Variables Controlled</td>
<td>75</td>
</tr>
<tr>
<td>Participants and Setting</td>
<td>76</td>
</tr>
<tr>
<td>Instruments</td>
<td>79</td>
</tr>
<tr>
<td>Questionnaire</td>
<td>79</td>
</tr>
<tr>
<td>Reading Material and Target Words</td>
<td>80</td>
</tr>
<tr>
<td>Program</td>
<td>81</td>
</tr>
<tr>
<td>Posttests</td>
<td>83</td>
</tr>
<tr>
<td>Vocabulary Knowledge Scale</td>
<td>83</td>
</tr>
<tr>
<td>Word Recognition Test</td>
<td>84</td>
</tr>
<tr>
<td>Multiple-Choice Reading Comprehension Tests</td>
<td>84</td>
</tr>
<tr>
<td>Written Recall</td>
<td>85</td>
</tr>
<tr>
<td>Validity Issue of the Posttests</td>
<td>85</td>
</tr>
<tr>
<td>Scoring</td>
<td>86</td>
</tr>
<tr>
<td>Vocabulary Knowledge Scale</td>
<td>86</td>
</tr>
<tr>
<td>Word Recognition Test</td>
<td>87</td>
</tr>
<tr>
<td>Multiple-Choice Reading Comprehension Test</td>
<td>87</td>
</tr>
<tr>
<td>Written Recall</td>
<td>87</td>
</tr>
<tr>
<td>Procedures</td>
<td>87</td>
</tr>
<tr>
<td>Data Analysis</td>
<td>89</td>
</tr>
<tr>
<td>Chapter IV: Results</td>
<td>91</td>
</tr>
<tr>
<td>Pilot Study</td>
<td>91</td>
</tr>
<tr>
<td>Participants</td>
<td>92</td>
</tr>
<tr>
<td>Procedures</td>
<td>92</td>
</tr>
<tr>
<td>Results</td>
<td>92</td>
</tr>
</tbody>
</table>
Discussion

Annotation Look-up Behavior Results

Posttest Results

Vocabulary Knowledge Scale Scores

Word Recognition Test Scores

Multiple-Choice Reading Comprehension Test Scores

Written Recall Scores

Results by Research Questions

Question 1: Effects of Annotation Type on Vocabulary Immediate Recall

Question 2: Effects of Annotation Type on Reading Comprehension

Question 3: Effects of Learning Condition on Vocabulary Immediate Recall

Question 4: Effects of Learning Condition on Reading Comprehension

Question 5: Interaction between Annotation and Learning on Vocabulary Immediate Recall

Question 6: Interaction between Annotation and Learning on Reading Comprehension

Summary of Findings

Chapter V: Discussion

Interpretations of Results

Effects of Annotation Type on L2 Vocabulary Immediate Recall

Effects of Annotation Type on L2 Reading Comprehension

Effects of Learning Condition on L2 Vocabulary Immediate Recall

Effects of Learning Condition on L2 Reading Comprehension

Interaction between Annotation Type and Learning Condition

Theoretical Implications

Pedagogical Implications

Directions for Future Research

Conclusions

References

Appendices

Appendix A: Questionnaire

Appendix B: Reading Text

Appendix C: Target Words

Appendix D: Pictures for Target Words

Appendix E: Screenshot of Reading Passage with Text-picture Annotation
List of Tables

Table 3-1  2 x 2 Factorial Design 71
Table 3-2  Test Scores and Matching Levels at AEI 77
Table 3-3  Countries and Languages Represented by the Participants 78
Table 3-4  Adapted VKS scoring categories and meaning of scores. 84
Table 4-1  Descriptive Statistics for Tests of the Pilot Study 93
Table 4-2  Means and Standard Deviation for the Tests of the Pilot Study by Annotation Type 94
Table 4-3  Means and Standard Deviation for the Tests of the Pilot Study by Learning Condition 94
Table 4-4  ANOVA results of the Pilot Study 94
Table 4-5  Descriptive Statistics for Look-up Behavior 97
Table 4-6  Means and Standard Deviations for Look-up Behavior 98
Table 4-7  ANOVA for Look-up Behavior 98
Table 4-8  Correlation between Look-up Behavior and Posttest Measures 99
Table 4-9  Descriptive Statistics for Vocabulary Knowledge Scale by Annotation and Learning 101
Table 4-10 Means and Standard Deviations for Vocabulary Knowledge Scale 101
Table 4-11 ANOVA for Vocabulary Knowledge Scale 102
Table 4-12 Descriptive Statistics for Word Recognition Test by Annotation and 104
List of Figures

Figure 2-1  Cognitive Theory of Multimedia Learning  63
Figure 2-2  Processing of Text Annotation  65
Figure 2-3  Processing of Audio Annotation  66
Figure 2-4  Processing of Picture Annotation  66
Figure 3-1  Detailed Research Design  72
Figure 4-1. Results of Pilot Study  96
Figure 4-2. Means for the Vocabulary Knowledge Scale by Annotation Types  102
Figure 4-3. Means for the Vocabulary Knowledge Scale by Learning Conditions  103
Figure 4-4. VKS Scores by Annotation and Learning Condition  103
Figure 4-5. Means for the Word Recognition Test by Annotation Types  105
Figure 4-6. Means for the Word Recognition Test by Learning Conditions  106
Figure 4-7. WRT scores by Annotation Type and Learning Condition  106
Figure 4-8. Means for the Reading Comprehension Test by Annotation Types  109
Figure 4-9. Means for the Reading Comprehension Test by Learning Conditions  109
Figure 4-10. RC Scores by Annotation Type and Learning Condition  110
Figure 4-11. Means for Written Recall by Annotation Type  113
Figure 4-12. Means for Written Recall by Learning Condition  113
Figure 4-13. Written Recall Scores by Annotation Type and Learning Condition  115
Figure 4-14. Means for Vocabulary Posttests by Annotation Type  117
Figure 4-15. Means for Reading Comprehension by Annotation Type 118
Figure 4-16. Means for Vocabulary Posttest by Learning Condition 121
Figure 4-17. Means for Reading Comprehension by Learning Condition 123
List of Acronyms

L1: First Language
L2: Second Language
L3: Third Language
CALL: Computer-Assisted Language Learning
EFL: English as a Foreign Language
ESL: English as a Second Language
NS: Native Speaker
NNS: Non-native Speaker
SLA: Second Language Acquisition
TL: Target Language
The Effects of Multimedia Annotations on L2 Vocabulary Immediate Recall and Reading Comprehension: A Comparative Study of Text-picture and Audio-picture Annotations under Incidental and Intentional Learning Conditions

Zhaohui Chen

ABSTRACT

This dissertation investigated the effects of multimedia annotation on L2 vocabulary learning and reading comprehension. The overarching objective of this study was to compare the effects of text-picture annotation and audio-picture annotation on L2 vocabulary immediate recall and reading comprehension. This study also sought to examine the different effects under incidental and intentional learning conditions.

The participants were 78 intermediate adult ESL learners from three universities in northwest U.S. The participants read an Internet-based English text. Twenty target words, annotated in either text-picture or audio-picture, were embedded in the reading text. The participants accessed the annotations by clicking on the highlighted target words. Two instruments were used for measuring vocabulary immediate recall: Vocabulary Knowledge Scale and Word Recognition Test. Two measurements were used to assess reading comprehension: Multiple-choice Reading Comprehension Questions and L1 Written Recall.

In term of annotation types, the results indicated that the audio-picture annotation group did significantly better than the text-picture group in L2 vocabulary immediate recall. However, there was no significantly different effect between the two annotations
on L2 reading comprehension. In terms of learning conditions, the intentional learning condition resulted in significantly better performance in L2 vocabulary immediate recall than the incidental learning condition. However, the incidental learning condition resulted in significantly better L2 reading comprehension than the intentional learning condition only in the Written Recall measure, but not in the multiple-choice Reading Comprehension Test. In terms of interaction between annotation type and learning condition, there was not interaction between annotation type and learning condition on L2 vocabulary immediate recall. The interaction between annotation type and learning condition on L2 reading comprehension was not significant in multiple-choice Reading Comprehension Text. However, the interaction was found to be significant in Written Recall: in the incidental learning condition, the difference between text-picture annotation and audio-picture annotation was not significant; in the intentional learning condition, participants in text-picture did significantly better than those in audio-picture on Written Recall.
Chapter I Introduction

Second language (L2) learners at all levels are faced with the difficulty of learning vocabulary. For example, Nation (1993, 2000) suggests that a native speaker of English knows about 20,000 word families. This poses a challenging task for English as a Second Language (ESL) learners. However, vocabulary learning has long been neglected within the field of Second Language Acquisition (SLA) (Coady, 1993; Coady & Huckin, 1997; Davis, 1989; Gass, 1987; Read, 2004; Zimmerman, 1997).

Recent years have seen increased interest in L2 vocabulary research. One influential debate over the years is between incidental and intentional vocabulary learning (Gass, 1999; Hulstijn, 1992, 2001). The distinction between the two learning conditions has been attributed to learning task, learner attention and the pedagogical context of the learning (Read, 2004). Both approaches have been argued to contribute to the incremental learning of L2 vocabulary (Hulstijn, 2001; Paribakht & Wesche, 1997; Parry, 1997).

L2 research has argued for incidental vocabulary learning through reading (Coady, 1997; Krashen, 1989; Nation, 1990, 2001). This conforms to L2 learners’ reports that vocabulary learning happens, in most cases, accidentally during reading or listening. However, L2 incidental vocabulary learning tends to be incremental and slow (Hulstijn, 1992; Jacobs et al., 1994).

Annotation has been used as a standard feature in L2 reading materials to facilitate comprehension in which L2 vocabulary learning comes about as a by-product
(Davis, 1989; Holley & King, 1971; Hulstijn, 1992; Jacobs, Dufon & Fong, 1994; Joyce, 1997; Watanabe, 1997). As an instructional intervention, an annotation draws learner attention briefly away from reading, and focuses it temporarily on the form and meaning of the annotated word, thus enhancing vocabulary learning and overall reading comprehension. This echoes the interactionist view of SLA (Long, 1991, 1996) and the depth of processing hypothesis (Hulstijn, 1992).

The effects of text annotation on L2 vocabulary learning and reading comprehension have been examined by studies which produced mixed findings (Davis, 1989; Hulstijn, 1992; Hulstijn et al., 1996; Jacobs, 1994; Jacobs et al., 1994; Joyce, 1997; Ko, 1995; Rott, Williams, & Cameron, 2002; Watanabe, 1997). The advance of multimedia L2 learning and teaching has taken the investigation of annotation a step further (Lomicka, 1998). Different from the traditional marginal annotation, multimedia annotations can present vocabulary information in multiple modalities, such as audio (sound) and visual (text, picture and video) (Al-Seghayer, 2001; Chun & Plass, 1996a).

Studies have examined the effects of different types of multimedia annotations on incidental L2 vocabulary learning, in particular, the use of picture annotation and video annotation coupled with text annotation (Al-Seghayer, 2001; Chun & Plass, 1996a; Jones, 2003; Yoshii, 2000). These studies support dual-coding theory (Paivio, 1971, 1986, 1990) and confirm the cognitive theory of multimedia learning (Mayer, 1997, 2001) that maintains how meaningful learning engages learners in both verbal and visual cognitive processing systems. Dual annotation of text and picture or text and video are unanimously argued to be better than single annotations in facilitating incidental L2

Studies in audio annotation have mainly engaged the use of pronunciation of the target words, and their findings are inconsistent (Chun & Plass, 1996a; Svenconis and Kerst, 1995; Yeh & Wang, 2003). Audio is a different sensory modality from visual modality, such as printed text and pictures, because audio is processed by the auditory working memory while printed texts and picture are processed by the visual working memory; thus it should be treated separately as to its effect on learning. Studies of audio annotation should include not only the pronunciation, but also the definition or meaning of the target words.

Furthermore, research suggests that the addition of an audio element to dual annotations does not seem to have a definite effect on L2 vocabulary learning (Chun & Plass, 1996a; Svenconis and Kerst, 1995; Yeh & Wang, 2003). One possible explanation is that the information delivered simultaneously through different modalities (audio, verbal and visual) might overload the cognitive processing.

Only two studies have addressed the relationship between vocabulary annotation and reading comprehension in a multimedia setting (Chun & Plass, 1996b; Lomicka, 1998). Their studies confirmed the effects of annotations on reading comprehension, but the scarcity of research in this area needs further investigation. As an instructional intervention, multimedia annotation does briefly interrupt the reading process by drawing learner attention away from the intended reading goal; it is, therefore, necessary, when examining L2 vocabulary learning, to further investigate the effects of annotations on text comprehension.
Statement of the Problem

An overview of the studies on L2 vocabulary annotation, particularly multimedia annotation, suggests that there is little information about how different dual annotations, particularly text-picture and audio-picture annotations, affect L2 vocabulary learning and reading comprehension. Specifically, in ESL classrooms, there is little information on how these two types of dual annotation facilitate ESL students’ vocabulary learning through reading. This information is needed to understand the extent to which multimedia learning can be used in L2 reading instruction and the role of multimedia in L2 vocabulary learning. The scant information on audio annotation in multimedia L2 learning in comparison to other multimedia annotations warrants further investigation. In addition, incidental and intentional vocabulary learning in a multimedia setting has never been studied. The effects of multimedia dual annotations employing different modalities on L2 vocabulary learning and reading comprehension in incidental and intentional learning conditions remain unclear.

Theoretical Framework

As an interdisciplinary area of inquiry, SLA benefits from a broad range of theories and principles. Annotation research fits into the interactionist paradigm of SLA which argues that language input that learners receive need to be salient and be noticed to engage learners in negotiation of meaning (Long, 1996). The availability of annotation offers comprehensible input for unfamiliar words during the reading process and helps L2 learners to comprehend the text. Meanwhile, incidental learning of those unfamiliar words happens due to multiple inputs. For example, the learners notice the unfamiliar
words for the first time when seeing them in reading, then encounter them when checking the annotations, and finally seeing them when returning to reading (Watanabe, 1997). In addition, the use of multimedia annotation, which is only available upon mouse clicking, does not directly interfere with the reading process. In terms of cognitive explanations of SLA, annotation shifts the learner attention temporarily away from the main cognitive task of reading and focuses it briefly on the form and meaning of the unfamiliar words. This process is claimed to be facilitative of second language acquisition because deliberate attention to the target form helps the learners to notice the form and meaning. The noticing can give salience to an unfamiliar word so that it becomes more noticeable in future input, which, therefore, contributes to the learner's psychological readiness to acquire that feature (Long, 1996; Schmitt, 2000).

Because of the nature of this study and its focus on the use of multimedia annotations in ESL vocabulary learning and reading comprehension, the cognitive theoretical framework of multimedia learning (Mayer, 1997, 2001) was chosen as the theoretical framework. Mayer argues that, in multimedia learning, materials are normally presented in verbal (including written text and spoken text) and visual (including pictures, video, animation) forms; these presented materials enter the learners’ sensory memory through the visual channel (eyes) and the auditory channel (ears); and in the process, the learner is regarded as an active knowledge constructor who selects, organizes and integrating both visual/pictorial and auditory/verbal information such as printed texts, pictures, or spoken words (Mayer, 1997, 2001). Mayer’s theory describes the process of learning in a multimedia environment from different presentation mode and different
modalities, and extends itself into the realm of second language learning, and in particular multimedia annotations (Chun & Plass, 1996a).

**Purpose of the Study**

Motivated by previous studies on multimedia annotation and existing gaps in this literature, the overarching question addressed in this study was how different dual annotations affect L2 vocabulary learning and reading comprehension in both incidental and intentional settings. This study was designed to broaden our understanding of the application of multimedia leaning in a second language acquisition environment through the framework of cognitive theory of multimedia learning to second language vocabulary learning and reading comprehension. It examined the ways in which two different types of dual annotation, namely, text-picture and audio-picture, affected L2 vocabulary learning and reading comprehension. In addition, it addressed the effects of multimedia annotation on L2 students’ vocabulary learning in both incidental and intentional learning conditions.

**Research Questions**

This study was guided by the following research questions:

1. What are the different effects of text-picture and audio-picture annotations in facilitating L2 vocabulary immediate recall?

2. What are the different effects of text-picture and audio-picture annotations in facilitating L2 reading comprehension?
3. Which learning condition (incidental or intentional) results in better L2 vocabulary immediate recall?

4. Which learning condition (incidental or intentional) results in better L2 reading comprehension?

5. Is there an interaction between annotation type and learning condition on immediate L2 vocabulary immediate recall?

6. Is there an interaction between annotation type and learning condition on L2 reading comprehension?

**Hypotheses**

The following research hypotheses were investigated in order to answer the above six research questions:

1. **Hypothesis:** Students in the audio-picture annotation group will perform significantly better than students in the text-picture annotation group on the immediate vocabulary posttests.

2. **Hypothesis:** Students in the audio-picture annotation group will perform significantly better than students in the text-picture annotation group on the reading comprehension posttests.

3. **Hypothesis:** Students in the intentional learning condition will perform significantly better than students in the incidental learning condition on the immediate vocabulary posttests.
4. **Hypothesis:** Students in the incidental learning condition will perform significantly better than students in the intentional learning condition on reading comprehension posttests.

5. **Hypothesis:** There will be an interaction between annotation type and learning condition on L2 vocabulary immediate recall.

6. **Hypothesis:** There will be an interaction between annotation type and learning condition on L2 reading comprehension.

**Significance of the Study**

This dissertation was conducted for both theoretical and practical reasons. At the theoretical level, the study aimed at contributing to the growing body of multimedia L2 vocabulary studies. First, it provides the much-needed information on the nature of audio annotation on L2 vocabulary learning. Second, it sheds light on the use of different dual annotations for multimedia L2 learning. Third, it contributes to the extension of the cognitive theory of multimedia learning to second language learning.

This study provides practical information for CALL vocabulary material designers in choosing the right combination of modalities in facilitating L2 learning. It also informed language teachers and administrators who need to make solid decisions about multimedia programs to enhance L2 vocabulary learning and reading comprehension.
Definition of Terms

1. Annotation/Gloss – a brief definition or synonym or a short explanation of meaning given to an unknown L2 word during reading. Traditionally, an annotation is provided in the margin of the text or at the end of the text. In a multimedia setting, an annotation is linked to a hypertext. In this study, the words “annotation” and “gloss” are used interchangeably.

2. Audio annotation – a multimedia annotation that uses spoken text, preferably the voice of a native speaker of the L2. It can be the pronunciation of an unknown word, a short definition, or a sentence that contains the word.

3. English as a Foreign Language (EFL) – English being studied by non-native speakers as a foreign language in an environment where English is not spoken as the first language, as in China.

4. English as a Second Language (ESL) – English being studied by non-native speakers as a second language in an English-speaking country, as in the United States.

5. First language – The native language of a learner that has been acquired as a child.

6. Incidental learning – before engaging in the learning task, the learner is not told of the learning goal, but he/she is tested on that learning goal afterwards unexpectedly. For example, before reading a multimedia L2 text, an ESL student is told to read the text for comprehension, but is tested on specific pre-selected annotated words afterwards. In this case, the vocabulary learning is incidental learning.

7. Incidental vocabulary learning – Vocabulary learning that occurs incidentally during the process of reading or listening in a second language. In other words, the learners learn the word unintentionally when they are engaged in a reading or listening task.
8. Intentional learning – before engaging in the learning task, the learner is told of a subsequent posttest. For example, before reading a multimedia L2 text, an ESL student is told of a vocabulary test on those annotated words they will come across during reading. In this case, the vocabulary learning is intentional learning.

9. L2 (Second Language)- this term generally refers to a second language, including second language and foreign language. A second language refers to the language being studied as a second language in a country where it is used as the native language; a foreign language refers to a language being studied as a second language in a country where the language is not the native language. Second language and foreign language will be used interchangeably in the study.

10. Multimedia annotation – annotation presented through a multimedia program, in other words, when L2 learners encounter an unknown word when reading via the computer, they can click on the unknown word and access a provided annotation. This type of annotation employs multiple modalities, such as auditory (sound, spoken text) and visual (printed text, picture), and engage different presentation modes, such as verbal (printed text, spoken text) and nonverbal/visual (picture).

11. Multimedia – the presentation of learning material using both words and pictures via a computer (Mayer, 2001).

12. Multimedia learning – learning of new information from words and pictures via a computer.

14. Reading comprehension – after reading the text, students will be able to recall the main idea units from the text through answering open-ended comprehension questions.

15. Text annotation – an annotation in a written text. In a printed reading text, a text annotation is usually located in the margin, at the bottom, or at the end of the reading text; in a multimedia text, a text annotation is accessible upon clicking the unknown word. It usually appears in a pop-up window and does not block the text section where the word is embedded.

16. Picture annotation – an annotation that uses a picture to clarify the meaning or description of an unknown word. It is used in both printed reading text and multimedia reading text.

17. Video annotation – a multimedia annotation that uses digitized video or animation. It is used to clarify the meaning of an unknown word or describe the content of the word in a context.

18. Vocabulary immediate recall – the ability to recognize and recall the word meaning after one exposure of the word in reading context.

Limitations and Delimitations

This study focused on ESL vocabulary learning and reading comprehension; therefore, other aspects of ESL learning such as grammar, listening and speaking were excluded. The participants were post-secondary adult ESL learners; therefore, ESL learners at the elementary or secondary levels were excluded. The reading text was
written specifically in its content and length for the research study, catering to intermediate ESL students’ reading level.

Twenty target words in the text were highlighted and annotated. Instead of the traditional multiple factors involved in learning vocabulary, vocabulary learning in this study was operationalized as vocabulary immediate recall that indicates cognition of a word and recall of its meaning when presented after a one-time exposure during the web-based reading activity. Vocabulary immediate recall was only measured through immediate posttests; thus, long-term retention of word meaning was not measured.

The audio annotation in this study was comprised of a spoken text of the meaning for the annotated word by a native speaker. Therefore, the auditory modality in this case was limited to spoken text, excluding any other sounds such background music or sounds that are associated with words.

Both external and internal validity limited the findings of this study. One possible threat to internal validity was selection of participants, participants in different experimental groups might not be functionally equivalent in respect to their knowledge of the target words.

With regard to external validity, possible threats included the following: (1) Population validity: intermediate ESL students were randomly chosen from English language institutes at universities from northwest U. S. Therefore, the findings cannot be generalized to ESL students at the beginning and advanced levels. (2) Ecological validity: participants are limited to a certain number of ESL learners from a specific geographic area in the U. S.
To alleviate internal validity threats, participants were randomly assigned to four experimental groups. In terms of external validity threats, the participants in this study were a unique group of students, but they were representative of the larger population of intermediate ESL students.

Organization of the study

The present study is organized in five chapters. Chapter I introduces the study and outlines the significance of the study. The main reasons for conducting study are stated and research questions are presented along with hypotheses. Operational definitions of the terms used in this study are also provided.

Chapter II elaborates on the L2 vocabulary learning and L2 reading comprehension literature. This is followed by a detailed review of literature on annotation and multimedia learning.

Chapter III discusses the design of this study. It provides the readers the overview of the procedures, including participants, instruments, data collection and the nature of data analysis.

Chapter IV presents the results of data analysis and answers each the six research questions accordingly.

Chapter V discusses the findings based on the research questions. It also presents theoretical implications and pedagogical recommendations in the field of second language acquisition. Directions for future research are also provided.
Chapter II Review of the Literature

Traditionally, the teaching and learning of vocabulary has been neglected in second language acquisition (SLA) research (Coady, 1993; Davis, 1989; Gass, 1988; Zimmerman, 1997). The argument in SLA to focus on form in communicative classroom activities tends to put the acquisition of grammar in the spotlight (Doughty & Williams, 1998; Long & Robinson, 1998). Therefore, in favor of syntax, vocabulary has generally been given a secondary place in the language curriculum (Nation, 1990, 2001). However, many L2 learners share the difficulties of learning vocabulary, and are often frustrated and discouraged by the unfamiliar words contained in reading texts. It seems that when L2 learners are engaged in reading texts or listening in the target language, the first and foremost challenge they are faced with is vocabulary, as argued by many second language professionals (Folse, 2005; Grabe & Stoller, 1997; Hulstijn, 2001; Laufer & Hulstijn, 2001; Nation, 1990, 2001; Read, 2004).

This chapter contains three parts. The first part presents an overview of L2 vocabulary learning in the context of reading, focusing on the distinction between intentional and incidental learning. This is followed by an introduction of annotation as an instructional intervention to facilitate incidental vocabulary learning. The second part reviews studies on annotation in both printed texts and multimedia texts, including text annotation, picture annotation, video annotation and audio annotation. The third part discusses the theoretical underpinnings of multimedia annotation, namely, the cognitive theory of multimedia learning and its extension in second language vocabulary learning.
The application of the cognitive theory of multimedia learning in multimedia L2 vocabulary annotations is also explained.

*L2 Vocabulary Learning*

Researchers have argued for the importance of vocabulary in SLA because limited vocabulary knowledge prevents L2 learners from communicating effectively in the target language (TL) (Davis, 1989; Gass, 1999; Stein, 1993; Wesche & Paribakht, 1999). As a result, vocabulary research has received increased recognition in SLA since the 1990s, especially after the appearance of Gass’s (1987) thematic collection on L2 lexical issues published in SSLA. Paribakht and Wesche (1999) stated that “since that time, a large body of L2 research and theory has developed around these and other topics, and it has become possible to deal comprehensively with single core issues in L2 lexical acquisition from multiple perspectives” (p. 175). This trend corresponds to the long-overdue emphasis on vocabulary acquisition.

The reciprocal relationships between vocabulary knowledge and reading comprehension have been demonstrated for both first language (L1) learning (Anderson & Freebody, 1981; Stanovich, 1986, 2000; Sternberg, 1987) and second language (L2) learning (Bossers, 1992, cited in Pulido, 2004; Haynes & Baker, 1993; Laufer, 1992). It is reasonable to assume that better readers tend to have a larger number of sight words, and those with a larger vocabulary tend to be better readers. An educated native speaker of English, suggested by some researchers (Laufer, 1992; Nation, 1993, 2001), knows about 20,000 word families. This poses a challenging task for ESL learners. The readability of an L2 text, to a large extent, depends arguably more on vocabulary than on
other factors such as sentence structure, syntax and rhetoric style (Bergman, 1978). Therefore, it is necessary for L2 learners to learn and enlarge the size of their vocabulary to be able to comprehend L2 texts.

L2 studies have attempted to establish a connection between vocabulary knowledge and reading comprehension (Laufer & Sims, 1985; Coady, 1993; Grabe & Stoller, 1997). Laufer and Sim (1985) entertained the possibility that L2 learners apply more vocabulary knowledge than that of the subject matter or syntax in understanding the reading text. In other words, a certain number of known words are necessary for L2 learners to approach an authentic L2 text comfortably. Laufer (1992) claimed a baseline of 3000 words as the threshold vocabulary size required for academic English reading. Furthermore, Nation (1990, 2001) argued that in order to comprehend an L2 text with relative ease, readers should be familiar with 95-98% of the words in the text at any level.

**Incidental and Intentional Learning**

The distinction between incidental and intentional vocabulary learning takes its origin from experimental psychology (Hulstijn, 2001). In the psychological literature, incidental learning and intentional learning are distinguished with different pre-learning instructions. Incidental learning refers to the situation in which the learners are not informed of their responsibility for certain information but are evaluated on that information later. On the other hand, in intentional learning, the learners are informed of their responsibility for certain information before doing the task.

In L2 vocabulary learning, there has been debate between incidental and intentional learning over the years (Hulstijn, 1992, 2001; Hulstijn, Hollander and
Greidanus, 1996). Despite the discrepancy among researchers on operationalizing the two terms, it is generally agreed that intentional vocabulary learning, in which student attention is directly engaged and focused on vocabulary, offers a greater chance for vocabulary learning (Hulstijn, 1992, 2001; Ko, 1995; Nation, 1991, 2001). In incidental vocabulary learning, students are engaged in other language learning activities such as reading or listening, and vocabulary learning occurs incidentally and does not involve a conscious effort to learn words (Schmitt, 2000). It is, therefore, defined as “a by-product, not the target, of the main cognitive activity, reading” (Huckin & Coady, 1999, p. 182).

Read (2004) attributed the distinction between incidental and intentional vocabulary learning to learner attention and pedagogical contexts of the learning. Factors such as the context in which the words are used, the task demands and others, as pointed out by Huckin & Coady (1999), are important to understand incidental vocabulary learning. In addition, tasks beyond reading itself, such as initiating look-up behavior, have been a recent twist added to the incidental vocabulary learning domain (Laufer & Hill, 2000).

It seems that more demanding tasks result in more incidental vocabulary learning through reading. In this respect, Laufer and Hulstijn (2001) proposed an important concept of "task-induced involvement" that is three-fold: the need to learn, the ability to search for information or form of the word, and the evaluation of the available information for the particular word under inquiry (Read, 2004, p. 148).

However, the incidental-intentional dichotomy has been challenged in recent years by a number of researchers (Paribakht & Wesche, 1997; Parry, 1991, 1993, 1997;
Zimmermand, 1994). Hulstijn (2001) suggested that it “is the quality and frequency of the information processing activities (i.e., elaborations on aspects of a word’s form and meaning, plus rehearsal) that determine retention of new information” (p. 275). It was also emphasized by Hulstijn (1992, 2001) and Hulstijn et al. (1996) that both approaches should coexist in L2 teaching and learning. It is a consensus among L2 researchers that both intentional instruction and incidental learning are necessary for second language learners, and they are two complementary activities (Hulstijn, 2001; Laufer & Hulstijn, 2001; Nation, 2001; Schmitt, 2000; Read, 2004).

**Incidental Vocabulary Learning**

It is a commonly held belief in both L1 and L2 research that vocabulary is mostly learned incidentally through reading (Coady, 1997; Krashen, 1989; Nagy, Herman & Anderson, 1985; Nation, 1990, 2001; Shu, Anderson & Zhang, 1995). Most L2 learners as well seem to agree that vocabulary learning happens in most cases accidentally during reading or listening.

After reviewing 144 studies on incidental vocabulary learning, Krashen (1989) concluded that incidental vocabulary learning affords better results than intentional vocabulary learning. In a review of incidental vocabulary learning, Huckin and Coady (1999) maintained that:

Most scholars seem to agree that, except for the first few thousand most common words, vocabulary learning predominantly occurs through extensive reading, with the learner guessing at the meaning of unknown words. This secondary type of learning is
called “incidental” learning because it is a by-product, not the target, of the main cognitive activity, reading.

They suggested three advantages of incidental vocabulary learning over explicit vocabulary instruction. (1) In comparison to the isolated word lists sometimes used in explicit instruction, incidental learning presents the known words in context (Oxford & Crookall, 1990). (2) Pedagogically speaking, incidental learning, serving as a double-edged sword, promotes simultaneously both vocabulary acquisition and reading comprehension. (3) By triggering selective attention, incidental learning enables more individualized learning of vocabulary (Schmitt, 2000).

However, the number of new words learned incidentally is relatively small compared to the number of words that can be learned intentionally (Hulstijn, 1992). Even with the use of a dictionary and the inferring strategy, incidental vocabulary learning tends to be incremental and slow (Hulstijn, 1992; Jacobs et al., 1994). In their study, Hulstijn, Hillander and Greidanus (1996) summarized several reasons why L2 learners failed to learn the unknown words incidentally during reading. (1) Sometimes L2 learners failed to notice the new words, or they assumed the familiarity with the new words. (2) Sometimes L2 learners noticed the new words, but ignored them. (3) The learners did not focus their attention on the unknown words. (4) The inferred meanings from context by the L2 learners were incorrect. (5) The low frequency of most unknown words prevents effective acquisition.

One way to facilitate incidental vocabulary learning is through the use of annotation or gloss. When consulting the dictionary for word meanings, students need to search for and choose the most appropriate meaning out of several possible definitions.
Unlike dictionary look-up, annotation offers the meanings of words based on the context and presents the meanings in the margin which is easily accessible to students.

**Annotation**

As defined by the Merriam-Webster Online Dictionary, gloss refers to “a brief explanation (as in the margin or between the lines of a text) of a difficult or obscure word or expression.” Annotation is defined as “a note added by way of comment or explanation.” Gloss and annotation, as a vocabulary learning aid, have been used interchangeably in L2 research and pedagogy.

The origin of gloss dates back to the Middle Ages, when Scripture verses were circled with notes written in smaller fonts (Davis, 1989). According to lexicographer Werner Hüllen (1989, cited in Roby, 1999), early glosses were usually learner-generated. Medieval students produced these interlinear or marginal scribbling as they struggled with Latin text. Glosses generated by teachers or material developers came much later. Widdowson (1978) divided these professionally generated glosses into two types: priming glossaries that serve as an advance organizer and are usually provided preceding the reading text, and prompting glossaries to which readers turn for consultation during the reading process.

Roby (1999) offers a comprehensive taxonomy of glosses that divides glosses using six criteria: (1) gloss authorship that divides glosses into those generated by the learners or by professionals such as instructors or material developers, (2) gloss presentation that distinguishes priming gloss and prompting gloss, (3) gloss functions that include procedural functions (for example, metacognitive, highlighting, and clarifying)
and declarative ones (such as encyclopedic and linguistic), (4) gloss focus that
emphasizes either textual or extratextual information, (5) gloss language in L1, L2 or L3,
and (6) verbal, visual or audio gloss forms. He defined glosses as any “attempts to supply
what is perceived to be deficient in a reader's procedural or declarative knowledge”
(Roby, 1999, p. 96).

In this study, annotation is defined as a brief definition or a short explanation of
meaning given to an unknown L2 word during reading. The annotations will be provided
in a multimedia setting where the students can access the annotations by clicking the
highlighted words in the web-based reading text. Two types of multimedia annotation
are available: text-picture and audio-picture.

*Traditional Annotation*

A vocabulary annotation is defined as a brief definition or synonym (Nation,
1990, 2001) or an explanation of the meaning of an unfamiliar word (Pak, 1986, cited in
Jacobs, 1994). Annotations are supplied to help L2 learners’ vocabulary learning and
enhance reading comprehension. Arrows, brackets, underlining, bolding, or highlighting
in the text are common ways to indicate these annotated words (Stewart & Cross, 1991).
Vocabulary annotations are typically located in the margins either on the side or at the
bottom of the page, or at the end of the reading text.

Foreign language/second language reading materials have long been using
annotation as a standard feature (Davis, 1989; Holley & King, 1971; Jacobs, 1994).
Holley & King (1971) posited two advantages of annotation in facilitating vocabulary
learning: (1) Annotations help L2 learners avoid incorrect guessing from the context, and
(2) the annotation-checking behavior engages rehearsal of the targeted vocabulary. For example, seeing the unknown word again in the annotation will reinforce the first exposure of the word in the text, making it possible for the word to integrate into the learner’s reading vocabulary. Watanabe (1997) explained the second advantage in terms of input frequency; that is, L2 learners experience multiple inputs of the targeted words. Take a particular unknown word, for example, L2 learners read the text and meet the unknown word (first input), they check the annotation for the meaning of the word (second input), and they return to the reading text trying to fit the meaning into the context (third input).

Nation (1990, 2001) argued for vocabulary annotation being facilitative to both vocabulary learning and reading comprehension: (1) the knowledge of unknown words provided by annotations facilitates reading comprehension; thus, authentic L2 texts can be used in an unsimplified manner; (2) annotations prevent incorrect guessing from context and provide context-specific meanings for unknown words during reading; therefore, L2 learners do not need to depend on teachers or dictionaries for definitions; (3) L2 learners can choose to access the annotations when needed in order to minimize interruption to the reading process; (4) annotation promotes individualized learning in that L2 learners are offered the freedom to choose needed annotations for word meanings; (5) annotations draw learners’ attention to the target words and might encourage and reinforce learning.

Annotations seem to serve several purposes. First, they assist vocabulary learning (Holley & King, 1971; Hulstijn, 1992; Jacobs, Dufon & Fong, 1994; Ko, 1995; Watanabe, 1997). Second, they enhance reading comprehension (Davis, 1989; Ko, 1995;
Third, they draw learner attention to the annotated words in both forms and meaning, according to Schmidt’s (1995) noticing hypothesis. Furthermore, L2 learners prefer vocabulary annotations (Jacobs, Dufon & Fong, 1994) to dictionary use. Research has, therefore, focused on the effects of different types of annotation, and the effects of annotation on L2 vocabulary learning and reading comprehension.

The use of annotation as an instructional intervention draws learner attention briefly away from the main cognitive task of reading and toward the form and meaning of unknown words. This fits into the focus-on-form approach proposed by Long (1991). The approach is claimed to be facilitative to L2 acquisition. Focus-on-form “overtly draws students’ attention to linguistic elements as they arise incidentally in lessons whose overriding focus is on meaning or communication” (p. 45-46). Though generally associated with grammatical acquisition, focus-on-form is suitable to explain the facilitative use of annotation in L2 vocabulary learning.

According to Schmitt’s (2000) noticing hypothesis, the provision of annotations, by giving salience to the target forms, makes the annotated words more noticeable and contributes to the learners’ psychological readiness to acquire the target forms. Noticing involves the intake both of meaning and form of the annotated words, which helps the learners to progress from initial recognition to internalization of the target forms. When checking the annotations provided in an L2 reading text, learners pay deliberate attention to particular lexical forms and they notice the gap between their own performance in the target language and the performance of proficient users of the language.
Multimedia Annotation

The investigation of annotation has been taken a step further with the advance of multimedia application in second language teaching and learning (Lomicka, 1998). Davis (1989) predicted the promise of multimedia annotations because “hypertext is invisible and thus unobtrusive in the reading process, yet the user is able to obtain as much or as little detail about a given concept as is desired” (p. 42). Jacobs (1994) confirmed that computerized annotation was an effective means in L2 vocabulary learning and was preferred by L2 learners.

Different from traditional annotation, multimedia annotations are not restricted to textual information. Instead, multiple modalities can be used to present vocabulary information, such as auditory mode (sound) and visual mode (text, picture and video) (Al-Seghayer, 2001; Chun & Plass, 1996a). Words with annotations in the reading texts are usually indicated by hyperlinks. Upon clicking the hyperlink on the word, the learners can access various forms of annotation in a pop-up window or in a designated area of the screen such as the left side, right side or the bottom of the screen. Usually, if a pop-up window is used, it does not cover the text section where the annotated word appears, in order not to interfere with the reading process.

Overall, there are four types of annotations: text annotation, picture annotation, audio annotation, and video annotation. A text annotation offers a textual definition or meaning of the unknown word. A picture annotation usually employs a static picture to describe the meaning of the unknown word. An audio annotation often uses the voice of a native speaker to read the word, or to read a sample sentence containing the target word,
or to read the meaning of the target word. A video annotation employs a video clip to depict the meaning of the target word.

In the following section, annotation studies will be reviewed in four subsections: text annotation, picture annotation, video annotation, and audio annotation. Each of the subsections will cover both printed text and computerized text. In fact, annotations involving pictures, video and audio are mainly located in multimedia environments.

Text Annotation

In printed reading materials, text annotations are often located in the margin, at the bottom, or at the end of the reading text. In multimedia texts, when students click on an annotated word, they can view the meaning or definition of the word in a pop-up window or in a certain location of the computer screen. In this section, discussion will be first focused on text annotation in printed reading texts, followed by a review of text annotation in multimedia texts.

Text Annotation in Printed Reading Texts

Studies addressing text annotations will be reviewed in two parts: text annotation and L2 vocabulary learning, and text annotation and L2 reading comprehension.

Text annotation and L2 vocabulary learning. Hulstijn, Hollander, and Greidanus (1996) examined incidental vocabulary learning for second language learners. Their study identified the use of marginal text annotation as an effective method. Other studies confirmed that text annotation in printed reading text could enhance second language
learners’ retention of vocabulary (Hulstijn, 1992; Jacobs, Dufon & Fong, 1994; Ko, 1995; Watanabe, 1997).

In Hulstijn, Hollander, and Greidanus’ (1996) study, three different types of instructional scenarios on a French short story by Guy de Maupassant for advanced L2 learners were compared: (1) marginal annotation (annotation in L1 Dutch), (2) dictionary access (bilingual dictionary), and (3) text only. The issue investigated was whether marginal annotation or dictionary use would lead to better vocabulary retention. The posttests of students’ receptive and productive knowledge of the 16 target words showed significantly better performance by the marginal-annotation group than the dictionary group and text group.

Different types of text annotation have been described such as single and multiple-choice annotations, as well as annotations in L1 and L2 (Hulstijn, 1992; Jacobs, Dufon & Fong, 1994; Kang, 1995; Watanabe, 1997). In Hulstijn’s (1992) study, experiments I, III and V were conducted with adult Dutch as a second language learners. Two sets of conditions were compared: gloss to no gloss and multiple-choice glosses to single-glosses. The multiple-choice gloss for each target word was comprised of a correct synonym and three distractors, while the single-gloss simply offered a synonym. In experiment I, L1 gloss led to better vocabulary retention than when the meaning had to be inferred from the context. Experiment III proved that the multiple-choice gloss group achieved substantially higher vocabulary scores than the single-gloss group on the posttest. However, Experiment V did not find any significant difference between single gloss and multiple-choice gloss. Therefore, Hulstijn (1992) pointed out that because of the high rate of incorrect inference in the open multiple-choice approach, it did not
necessarily guarantee better learning of L2 vocabulary in comparison to the closed single-glossing format.

In line with comparing single-gloss and multiple-gloss, Watanabe (1997) conducted his research with 231 Japanese university students learning English as a second language. A between-subjects design was used to examine the effects of four conditions in learning 16 unknown words in an English (L2) article: (1) appositive (an explanation in L2 of each difficult word inserted immediately after the word), (2) single marginal glosses (in L2), (3) multiple-choice marginal glosses (in which two choices are presented in L2 for each difficult word, and one of them is the correct meaning), and (4) text only with no cue. Both single and multiple-choice gloss groups were found to perform significantly better on the vocabulary posttest than the appositive and text-only groups on both the immediate and delayed posttest. Though the finding of the single-gloss group scoring higher in the posttests than the multiple-choice gloss group somewhat contradicted the results of Hulstijn (1992), it is confirmed that adding marginal gloss to a reading text could be an effective method to enhance vocabulary learning during reading.

Dufon and Hong’s (1994) study on L2 Spanish reading used three formats: (1) no gloss, (2) L1 gloss and (3) L2 gloss. Their results showed that students who had access to glosses outperformed students without glosses on the immediate vocabulary translation posttest. However, the effectiveness of gloss was not found in the delayed posttest four weeks later. In terms of this, Jacobs et al. warned that although gloss is preferred over no gloss, gloss only has a potentially positive effect on vocabulary acquisition with sufficient L2 competence. In addition, certain proficiency level was necessary to make
effective use of L2 gloss. Therefore, the positive relation between gloss and vocabulary learning was established, at least for immediate retention if not for long-term retention. With regard to the comparison between L1 gloss and L2 gloss, no difference was found between the languages (L1 and L2) used for glossing. A questionnaire was used to examine students’ preference, and the result revealed that students preferred L2 glosses to L1 glosses if the L2 glosses were comprehensible.

In order to further investigate the possible difference between L1 and L2 glosses, Ko (1995) used a similar design as Jacobs, Dufon and Hong (1994) with 189 Korean college students learning English as a foreign language (EFL). Students took a vocabulary pretest and were asked to read an 854-word English text. In contrast to the finding of Dufon and Hong (1994), the multiple-choice vocabulary posttest immediately administered after the reading showed significant difference between L1 gloss and L2 gloss. In other words, students with access to L1 gloss outperformed significantly better than those with access to L2 gloss. The effect was found significant in the delayed posttest one week later.

The effectiveness of L2 gloss over L1 gloss in vocabulary retention was also challenged by Laufer and Shmueli’s (1997) study. Hebrew-speaking high school EFL students (N=128) were asked to read an English text in which 10 target words were glossed in Hebrew and 10 other target words were glossed in English. Multiple-choice assessment was used in the immediate posttest and delayed posttest five weeks later. Both tests showed that L1 gloss resulted in more vocabulary retention than L2 gloss. This conflicts with the finding by Jacobs et al. (1994), but the students’ level of the second
language in Laufer & Shmueli (1997) might be used as an explanation. Certain proficiency of the second language was necessary to fully utilize glosses in L2.

Text annotation and L2 reading comprehension. Several text annotation studies in printed reading materials have examined the effect of annotation on L2 reading comprehension, but with mixed findings (Davis, 1989; Ko, 1995; Jacobs, 1994; Jacobs et al., 1994; Joyce, 1997; Rott, Williams, & Cameron, 2002). Davis (1989) tried to investigate the facilitating effect of marginal glosses on foreign language reading. The study used a between-subjects design among three groups: (1) reading text with no aids, (2) reading text with pre-reading questions and vocabulary definitions, and (3) reading text with glossed vocabulary. The participants in the study were 71 intermediate-level U.S. college students learning French as a second language. Out of the 936-word reading text, 28 words and expressions were glossed. After the experiment, students in all the three groups were asked to recall the reading text in their L1 (English). Significant difference was found between the no-aid group and the other two groups with either pre-reading aids or glossing. The results supported the idea that use of glosses facilitated text recall or reading comprehension for intermediate foreign language learners.

Jacobs (1994) confirmed the positive effect of marginal glosses on reading comprehension. He attempted to study the potential interactive effects of text variables, learners variables and situational variables for 116 U.S. college students in their third-semester Spanish reading class. The presence or absence of vocabulary glosses in the reading text was designated as the text variable. Learner L2 proficiency was chosen as one of the learner variables, and time on task was analyzed as one situational variable.
Two reading conditions were presented: (1) unglossed L2 (Spanish) text and (2) L2 text with marginal gloss in L1 (English). The glosseed group was found to recall 30% more idea units than the unglossed group, while no interaction was found significant between any of the learners and situational variables and glossing. However, one interaction that did approach significant level was the one between glossing and time on task for one of the unit idea recall measures.

In Rott, Williams and Cameron’s (2002) experiment, 67 English-speaking college students taking 4th-semester German were divided into four reading conditions: (1) text with multiple-choice glosses; (2) text reconstruction in which students were asked to reconstruct the text after reading; (3) text reconstruction and multiple-choice glosses; and (4) text-only control group. L1 recall administered after the experiment proved that gloss groups yielded significantly better reading comprehension than the control group. The combined text reconstruction and gloss group had significantly better comprehension than the control group, which further confirmed the positive effect of gloss on reading comprehension.

However, arguments have been made against the effect of glossing on reading comprehension (Ko, 1995; Jacobs et al., 1994; Joyce, 1997). Ko’s (1995) study looked at the effects of glossing on both L2 vocabulary learning and reading comprehension. Although the effect of glossing on vocabulary learning was supported significantly, its effect on reading comprehension was not positive. L2 gloss was found to be the same as no-gloss in facilitating reading comprehension, and L2 gloss was found to even worsen reading comprehension.
In her study of comparing glossed and unglossed L2 text on reading comprehension, Joyce (1997) used three levels, first-semester, second-semester and third semester, of students learning French as a second language at one U. S. university. Students from each proficiency level were divided into glossed and unglossed groups. The results questioned the effect of gloss on L2 reading comprehension because gloss was found to provide no significant help for those who studied text with glosses in their comprehension as measured by L1 after-reading recall.

Text Annotation in Multimedia Texts

As technology has been used more widely in second language teaching and learning, one technique used in glossing is known as hypertext (Bunnell, cited in Davis, 1989). Davis (1989) noted that glossing through hypertext offers two features: (1) multimedia glosses are usually invisible and, therefore, do not interfere with the reading process; and (2) by clicking on any hyperlinked word, readers have the freedom to choose the amount of information needed on a particular word in the text (cited in Al-Seghayer, 2001). In text annotation, there have been a few studies that dealt with computerized materials (Lyman-Hayer et al., 1993; Koren, 1998; Nagata, 1999).

Text annotation and L2 vocabulary learning. Lyman-Hayer et al. (1993) examined the impact of multimedia programs on vocabulary acquisition, and analyzed the difference between text annotations in conventional printed text and text annotation in multimedia text. The delayed vocabulary test administered one week later confirmed that the students who worked in the multimedia text annotation group performed better than
those who worked in the traditional text annotation group. Therefore, they concluded that multimedia text annotation enhances students’ ability in vocabulary retention.

Replicating Watanabe’s (1997) study, Nagata (1999) investigated the effectiveness of computer-assisted interactive glosses. American college students (N=26) taking second-semester Japanese were asked to read a text on the computer in two versions, text with single gloss and text with multiple-choice gloss. The reading text and glossing appeared on the computer screen simultaneously. She found that multiple-choice gloss was significantly more effective than single gloss to help students in recalling the target vocabulary and grammatical items during the immediate posttest. Nagata explained the conflicting finding from Watanabe (1997) by arguing that in multiple-choice glossing, students made more efforts to interpret and paid more attention to the annotated words.

However, the delayed posttest which was conducted one month after the treatment proved no difference between the two glossing methods on the vocabulary test, while the difference on grammar test was still statistically significant, favoring the multiple-choice group. Compared to the immediate posttest, a dramatic decrease of retention was found in both vocabulary and grammar items, which confirmed that the one-time experiment had immediate effects on vocabulary learning, but might not be effective for long-term retention.

Different from other studies using between-subjects design, Koren (1998) used a within-subjects design [within-subjects was also used in Hulstijn’s (1993) experiment III] to examine the possible advantage of hypertext clue over conventional glossing. First-year university EFL students (N=129), most of them speaking Hebrew as their first
language, read an English text. In the text, seven words were highlighted and on clicking contextual or structural clues were provided, and another nine words were given a traditional marginal gloss in the L2. On the first posttest administered three days later, students remembered the inferred words four times better than they remembered the glossed words; on the delayed posttest three months later, in comparison to the first posttest, retention of the glossed words was nearly identical (8%), but retention of the inferred words went down by nearly half from 33% to 17%. The results confirmed that text glosses were effective on incidental vocabulary learning, which confirmed findings from other studies, but contrary to other studies, the effect of glosses was retainable in this study.

*Text annotation and L2 reading comprehension.* A few studies in text annotation using multimedia text looked into the possible link between text annotation and L2 reading comprehension. With the major purpose of comparing annotation in printed text and multimedia text, Lyman-Hayer *et al.* (1993) administered a written recall protocol immediately after the experiment to examine whether the annotation influenced L2 reading comprehension. They found positive effects of text annotation on students’ reading comprehension.

In Hulstijn’s (1993) study, 44 Dutch-speaking high-school students were randomly divided into two task groups. The summary group was asked to summarize in Dutch after reading and the question group was asked to answer 10 comprehension questions. Both groups read the same 772-word English text, with 109 words glossed, where the glosses could be retrieved on the computer screen with a click of the enter key.
Although different individual look-up behaviors were found, no difference in reading comprehension was found, nor was an effect of the reading task variable on gloss use.

**Summary**

In summary of text annotation studies, different L1 and L2 are used, ranging from English, Japanese, French, and Hebrew to Dutch, and different reading materials were chosen by the studies as well. One aspect shared by most the text annotation studies is that one proficiency level (either elementary or intermediate or advanced) was used in comparing different annotations. All studies used between-subjects design except Koren (1998) and Hulstijn’s (1992) experiment III that used within-subjects design. Different lengths of time were used for the delayed vocabulary posttest, ranging from one week to three months.

In both printed text and multimedia text, the effect of text annotation on vocabulary learning and reading comprehension were confirmed, though challenged by some mixed conclusions. Jacobs (1994) pointed out that “no single variable alone can guarantee a particular effect (context and interaction between variables). In the case of glossing, the text difficulty, readers’ proficiency, number and choice of items glossed, and many other factors determine whether glossing will improve comprehension in a particular situation” (p. 129). As echoed by Nagata (1999), “We should keep in mind that the results may vary, depending on the type and difficulty of reading materials, the learners’ proficiency level, the kind of test procedure and measurement, and so forth” (p. 476).
Visual aids have long been assumed to be beneficial to second language learning. Tuttle (1975) argued that “foreign language students can benefit from many types of visual material… the still or flat picture can prove to be a rich resource in the foreign language classroom” (p. 9, cited in Omaggio, 1979). The use of imagery representation of foreign words by actual objects or imagery was also claimed by Kellogg and Howe (1971) to be facilitative to children’s vocabulary acquisition in a foreign language. Subsequently, a number of researchers have explored the effect of visual stimuli on L2 vocabulary learning and reading comprehension.

Kellop and Howe’s (1971) study compared written words and pictures as cues for oral acquisition of Spanish vocabulary by children. The pictures yielded faster learning of new words than the written stimuli. And the effect was retained in the long term as indicated by greater recall of words shown in pictures. Terrel (1986, cited in Kost et al, 1999) proposed that combining the form and visual representation of unknown L2 vocabulary helped learners to acquire concrete ideas and references. In reviewing the techniques used in learning L2 vocabulary, Oxford and Crookall (1990) acknowledged the effectiveness of visual imagery and maintained that, “[M]ost learners are capable of associating new information to concepts in memory by means of meaningful visual images, and that visual images make learning more efficient” (p. 17) and “the pictorial-verbal combination involves many parts of the brain, thus providing greater cognitive power” (p. 17).

Omaggio’s (1979) empirical study provided pictorial contexts to French-as-a-second-language students as advance organizers. It was hypothesized that the provision
of additional visual context would facilitate reading comprehension. The results showed that students with access to a pictorial context did significantly better on the recognition test and recall than those with access only to text. This supported the positive effect of pictures on reading comprehension.

In annotation studies, picture annotation has been used to clarify the meaning of those unknown words second language learners encounter in reading. According to dual coding theory, the way learners comprehend pictures differs greatly from that of comprehending textual information (Paivio, 1971). In other words, text is processed by the verbal cognitive subsystem, while a picture is processed by the non-verbal cognitive subsystem. Research has compared L2 vocabulary learning from text annotation, picture annotation, and a combination of text and picture annotation

Annotation and L2 Vocabulary Learning

Kost, Foss & Lenzini (1999) and Yoshii (2000) both compared three annotations: text only, picture only, and text and picture, respectively in printed text and multimedia text. Their findings suggest the superiority of combining text and picture over single annotation types such as text or picture, establishing the effectiveness of dual presentation types of vocabulary annotations (text + pictures). In terms of the difference between text annotation and picture annotation, the results were interpreted in relation to the task type in the posttests. Other studies confirmed the findings (Kang, 1995; Jones, 2003; Yeh & Wang, 2003)

Kost, Foss and Lenzini (1999) conducted their study with 56 U. S. university students learning second-semester German. A 272-word reading passage was adopted
with consideration of learners’ proficiency level. Sixteen words were chosen as the target words based on a vocabulary pretest, and were annotated in three different ways: text (English translation), picture, and a combination of text (English translation) and picture. The students were randomly assigned to the three annotation groups, and had access to one of the three annotations when reading. Three assessment tasks, namely production task, picture recognition task, and word recognition task, were used to evaluate students’ short-term vocabulary learning and retention two weeks later. It was hypothesized that students with access to both text and picture annotations would outperform those with either text-only or picture-only annotations in both immediate and delayed posttests, and the performance of single text or single picture annotations would depend on task types. Due to factors such as the difficulty level and frequency of target word exposure, the production task in which students were asked to provide L1 translation of target words did not show any difference among the three annotation groups and was therefore excluded from the analysis. The analysis of the word recognition and picture recognition tasks showed that the text-only group outperformed the picture-only group in the word recognition test, while the picture-only group achieved higher scores than the text-only group in the picture recognition test. The researchers argued that the relative superiority of text annotation or picture annotation depended on the task type. In the picture recognition task, the combination annotation differed significantly on the picture recognition task from the text annotation, but not from the picture annotation. In the word recognition task, the combination annotation differed significantly from both text only and picture only annotations in the immediate posttest, but not from the text annotation in the delayed posttest. In summary, the combined text-picture annotation
produced the best results in both immediate posttests and delayed posttests. The authors explained that the combined annotation provided two different types of information that increased retrieval options from both verbal and visual cognitive systems, which reflected Paivio’s (1971, 1986, 1990) dual coding theory.

Yoshii (2000, also Yoshii & Flaitz, 2002) replicated the above-mentioned study in a multimedia setting. ESL students (N=151) at the beginning and intermediate proficiency levels participated in the study. A pretest was used to verify the participants’ unfamiliarity with the 14 target words that were annotated in three conditions: text only, picture only, and a combination of text and picture. The text annotation was given in the L2 since participants represented 18 different first languages. The participants read an English short story embedding the 14 target words on the Internet for the purpose of reading comprehension. They were randomly assigned to one of three annotation conditions. Three unexpected vocabulary posttests were administered both immediately and two weeks after the treatment: a definition supply test, a picture recognition test, and a word recognition test. A 3x2 ANOVA was performed to examine the scores in immediate posttests and delayed posttests. In terms of single annotation effect, the claims made by Kost et al (1999), i.e. that it was dependent on the posttest task type, was not fully supported by Yoshii (2000). The picture-only group did consistently better than the text-only group in all the posttests, and the difference was statistically significant in the delayed picture recognition test. This might explain the positive effect of visual annotation as suggested by Yoshii (2002), “the results of the study encourage the use of pictures as alternative or as accompaniments to textual cues” (p. 46). In comparison to the single annotations, the combination group demonstrated better performance on both
immediate and delayed posttests. For the immediate posttests, the combination group performed significantly better than the text annotation group in the picture recognition task. The combination group also scored significantly higher than the picture annotation group in terms of definition supply task. For the delayed posttests, significant difference was indicated in the picture recognition between the combination group and the text-only group, but no significant difference was found in the word recognition test or the definition supply test. Overall, the combination of text and picture proved to be the most effective annotation condition for both vocabulary learning and retention.

The positive effect of picture annotation on L2 vocabulary learning, particularly in combination with text annotation, was also reflected in Kang’s (1995) study. Involving 98 elementary EFL Korean students, Kang (1995) compared three computer-based instructional approaches: L1 translation, L1 translation and picture, L1 translation and picture plus an example sentence. The target words in the first and second approaches were given in pair-wise definition-based form without any context, while the third approach (context-based) embedded the target words in contextual sentences accompanied by explanatory pictures. Vocabulary definition recall and knowledge transfer were used as immediate posttests and delayed retention tests. On the recall measure, a statistically significant difference was found among the groups, specifically, the context-embedded words with picture group scored the highest, followed by the text and picture group, with the text only group being the lowest. The third group also showed significantly better performance than the other groups on the retention tests, suggesting that a contextualized approach using both text and picture was more effective in promoting long-term recall of vocabulary definitions. On the knowledge transfer
measure, students were asked to use the target words communicatively in computer-generated simulations resembling real world situations. The groups with access to text and picture performed much better than the group with access only to text, confirming the use of multiple annotation combining text and picture. Interestingly, the study used a listening comprehension test to measure students’ ability of recognizing newly learned vocabulary in a spoken context. The results showed that the groups with access to pictures outperformed those without.

Yeh & Wang’s (2003) study investigated the effectiveness of different vocabulary annotations on vocabulary learning for EFL college students in Taiwan. Freshmen (N=82) from science programs participated in the study. A computer courseware was designed to introduce Thanksgiving to 82 freshmen enrolled in the freshmen English class. Three annotation types were presented for unknown vocabulary: text annotation (Chinese translation and English explanation), text and picture annotation, text and picture and sound. For the sound component, a native speaker read the word, spelled the word and read the sentence in which the target word was planted. Three tasks were used in the posttest: word association questions, multiple-choice questions on word meanings, and a cloze test. Analyses of the participants’ performance on the posttests indicated that the text and picture annotation was the most effective for vocabulary learning, and the different effectiveness of the three annotation types was further confirmed by an ANOVA analysis to be statistically significant, suggesting generally consistent findings with previous studies (Kang, 1995; Kost et al, 1999; Yoshii, 2000). However, the addition of sound to the combination of text and picture proved not to be effective, and the result was even worse than the text-only annotation as indicated by the group mean scores: highest
mean score of 23.41071 from group 2 (text and picture), followed by mean score of 22.44444 from group 1 (text-only), and lowest mean score of 20.77777 from group 3 (text and picture plus sound). Furthermore, Fisher’s LSD pair comparison indicated that the text-picture annotation group significantly outperformed the text-picture-audio annotation group.

Instead of using written reading material in the above-mentioned studies, Jones (2003) used an aural text to investigate the effects of multimedia annotations in a pretest-posttest design. The participants were 171 English-speaking students in a second semester French class. Given 14 minutes to listen to a passage, the participants had access to one of four annotation conditions: (1) no annotation; (2) text annotation; (3) picture annotation; and (4) text and picture annotation. After the treatment, students took an immediate multiple-choice vocabulary posttest and a delayed vocabulary post-test four weeks later. On both the immediate and delayed posttest, when either text or picture (verbal or visual) annotation was available, students’ performance on the vocabulary test was better than when no annotation was available, which confirmed the findings of the effectiveness of using annotations to facilitate second language vocabulary learning from other studies (Davis, 1989; Hulstijn, Hollander & Greidanus, 1996; Jacob, 1994). No significant difference was found between the text-only annotation and picture-only annotation. This echoed mixed findings from Kost et al. (1999) and Yoshii (2000). Furthermore, the results showed the highest performance in the immediate posttest by students from the text and picture annotation group (mean score was 19.75), statistically different from all the other groups, although this difference was not found to be significant between the combination group and the picture-only group in the delayed
posttest. Qualitative techniques were also employed in this study to profile students’ perception to the annotation types. Themes emerging from the interview data supported the effectiveness of multiple annotations in learning second language vocabulary.

In summary, the use of dual or multiple annotations is supported to be helpful for second language vocabulary acquisition, especially when both verbal and visual (text and picture) are presented in the annotation. The supportive nature of multimodal annotation is also supported by student preference (Kang, 1995; Jones, 2003; Yeh & Wang 2003). Echoing Kellogg & Howe’s (1971) statement that multiple annotation types utilizing visual imageries, rather than single mode annotations, can facilitate the learning and retrieving process of second language vocabulary, Plass et al (1998) remarked that “the organization and interpretation of two different forms of mental representations enhance retrieval performance by providing multiple retrieval cues” (p. 34).

**Annotation and L2 Reading Comprehension**

The pilot study of Lomicka (1998) examined the effects of multimedia annotations on L2 reading comprehension. In her study, 12 college students enrolled in a second-semester French course, and read a text under three multimedia annotation conditions: full glossing, limited glossing and no glossing. The limited glossing gave both definitions in French and translations in English of the glossed word; in addition to that, images, references, questions, and pronunciation concerning the glossed word were added to the full glossing. Online think-aloud was employed as a reading comprehension measurement, and the results of the study indicated that full multimedia glossing promoted a deeper level of text comprehension. However, their use of the glosses was
oriented toward the goal of translation and paraphrasing in order to achieve a minimal level of comprehension

**Summary**

Based on the assumption that visual aids such as pictures can facilitate second language vocabulary learning, studies have been carried out to compare picture annotation to text annotation (Kost et al., 1999; Jones, 2003; Yoshii, 2000). The findings are mixed, with text annotation being better in word recognition tasks, and picture annotation better in picture recognition tasks in Kost et al (1999), but Yoshii (2000) found consistently better performance by the picture-only group than the text-only group in both word-recognition and picture-recognition tasks, and Jones (2003) found no difference between text annotation and picture annotation.

However, when combined, the dual annotation of text and picture was unanimously argued to be better than single annotations, especially better than text-only annotation (Kang, 1995; Kost et al., 1999; Jones, 2003; Yoshii, 2000; Yeh & Wang, 2003). In this respect, the dual-coding theory (Paivio, 1971, 1986, 1990) is supported empirically in the realm of second language vocabulary learning. The simultaneous activation of both verbal and visual cognitive systems offers learners dual channels to process the given information, and, therefore, have dual channels to retrieve information.

Furthermore, the addition of audio (the pronunciation of the target word) does not seem to enhance second language vocabulary; quite the contrary, it influences the learning result in a negative manner (Yeh & Wang, 2003). One possible explanation is that the information delivered through different modalities (audio, and visual) might
overload the cognitive processing; on the other hand, the learner preference of particular types of information presentation contributed to the finding. In Yeh & Wang’s (2003) study, the particular group of participants (Chinese EFL learners) contributed to the effect because Chinese EFL learners are usually more visual learners (Chen, 1998, cited in Yeh & Wang 2003). Further research to separate audio and visual in combination with text needs to be addressed.

In terms of reading comprehension, the effect of dual annotation of both text and picture is not clearly established due to the scarcity of studies. Further research is necessary in this respect.

**Video Annotation**

As streaming video becomes easier and more widely used in second language teaching and learning, annotation research has expanded its multimedia application of using static pictures to the use of animated videos. Studies have examined the effect of video annotation coupled with text in comparison to text-only, and to text coupled with pictures, specifically, the effect of different media types on second language vocabulary acquisition.

**Annotation and L2 Vocabulary Learning**

Chun and Plass (1996a) reported the results of three studies with a total of 160 second-year German students from three universities in California. The reading text was a 762-word German text compiled using the multimedia program CyberBuch, covering 11 screen pages. A total of 82 words were annotated in the program in three conditions:
(1) text (L1 definition) only, (2) text and picture, and (3) text and short video. Every annotated word was indicated by a unobtrusive symbol ('). When students clicked on the word, icons marking the available annotations appeared, and the chosen annotation appeared on the left of the computer screen. A within-subjects design was used for all three studies to eliminate the variable of individual differences. Study 1 and 2 used definition-supply (English equivalents) as the vocabulary posttest, and study 2 employed word recognition in the posttest.

The first focus of their studies was whether the availability of multiple annotations would result in better incidental vocabulary learning. In study 1, students (N=36) were unexpectedly asked to provide English equivalents to 15 target words (five each from the three annotation conditions) immediately after reading and two weeks later. The mean scores and percentages of correct answers were higher than what was expected for typical incidental vocabulary learning. The mean score of 3.89 (25.9%) on the immediate test and mean score of 3.97 (26.5%) on the delayed test were higher than the scores of their pretest (2.06/13.7%, and 1.94/12.9% respectively). The learning result was also better than the incidental vocabulary learning percentage of 5-15% reported by Coady (1993) and 5-21% by Knight (1994). Study 2 (N=103) used 36 target words in the posttest and found that 24.1% of them were retained in the immediate posttest. This number is better than both Coady’s (1993) 5-15% and Knight’s (1994) 5-21%. The results confirmed that using multiple annotations facilitated second language vocabulary learning. The word-recognition test used in study 3 offered much better retention rates of 77% (immediate posttest) and 77.1% (delayed posttest). The higher score on recognition was expected because recognition task tend to be easier than the definition-supply task.
The effects of different media types used for annotation was the second issue to be examined in Chun and Plass (1996a). In study 2, a significant difference was found in immediate retention scores for words between text-picture annotations and text-only annotations or text-video annotations, as shown by the ordering of mean scores and correct percentages from the highest to the lowest: text-picture (3.75/31.2%), text-video (2.76/23%) and text-only (2.15/17.9%). The result supported previous studies that showed pictures were effective in annotating words for second language learners (Kost et al, 1999; Yoshii, 2000; Kang, 1995). In particular, dual annotation that combined text and visual information (such as picture or video) is shown to be more effective than text-only annotation, which supported the dual-coding explanation that words annotated with both verbal and visual information will be more efficiently memorized.

Moreover, in dual annotations, text-picture seems to be more effective than text-video. The difference is explained by the authors in terms of viewing time of the two annotation types, text-picture annotation allowed longer viewing time and better mental representation of the presented information, and text-video annotation was comparatively shorter and offered less time for viewing and mental processing. To further demonstrate the effectiveness of text-picture annotation, students were asked to articulate their use of retrieval cues from different annotations. Their answers consistently favored picture annotation, though the difference was only marginally significant in comparison to text annotation, and was not significant in comparison to video annotation. The trend was claimed by the authors to correspond to the tendency for better word recall from text-picture annotations. No delayed posttest, however, was used to check the long-term retention of vocabulary learning. It is, therefore, uncertain whether the better performance
by the text-picture group in the immediate posttest would be retained by the time of the delayed posttest.

However, the same pattern of text-picture superiority was not found in study 3. Due to the small number of participants (N=21), no statistically significant difference was found among the three groups. Instead, very close mean scores and percentages of correct answers for target words were found among the three annotation conditions. The immediate posttest exhibited 75.1\% correction rate for text annotation, 76.2\% for text-picture, and 81.4\% for text-video; the delayed posttest indicated a similar range of 75\%, 81\% and 77.2\%. Though the result further supported the use of dual annotations combining text and visual, it seems inclusive to say one combination is favored over the other because at the immediate posttest, the text and video group outperformed the text and picture group, while the result switched in the delayed posttest. Upon a closer look, the better performance by the text-picture group in study 2 resembled that of the delayed posttest in study 3 favoring text and picture annotation, but since no delayed posttest was conducted in study 2, the interpretation is debatable. In addition, the difference reported in study 1 between the two dual annotation types was not statistically significant.

The comparison between picture and video as two annotation types was further explored in Al-Seghayer (2001). Similar within-subjects design was used here with 30 intermediate-level ESL students. Three annotation types were available: text annotation, text-picture annotation, and text-video annotation. The 30 target words, 10 in each of the annotation conditions, were chosen based on criteria of frequency, grammatical category, morphological category, and visual complexity. After reading the text via the multimedia program with access to different annotations, students took two vocabulary
posttests: recognition and production. The recognition test, in multiple-choice format, asked students to select from four alternatives the right English word corresponding to the provided information matching the modality in which the word was annotated. The production test required students to offer definitions in English for the 6 selected target words, 2 words from each of the three annotation conditions.

The results of the two vocabulary posttests presented the best performance by the text-video group. The mean and percentage of correct answers from highest to lowest are as follows: text-video (4.3/87%), text-picture (3.3/67%), and text-only (2.7/53%). The advantage of the text-video group was also shown by the distribution of frequency score of each annotation mode, more closely clustered for the text-video mode than the text-picture mode, with the text-only mode having the widest range. The most frequent mean score was 6 for text-video mode, 4 for text-picture mode, and 3-4 for text-only mode. Group-wise differences were further examined by the Friedman test, and the results indicated significantly better performance by the text-video group than both the text-only and text-picture groups, but no significant difference was found between the text-only group and text-picture group.

The effects of different annotation modes were also examined through qualitative data from questionnaires and interviews. The questionnaire data showed that students voted video the most helpful annotation mode (86.6%), followed by picture (70%) and text (10%). The same preference was reflected in the interview in which 90% of the participants agreed that video annotations presented the word meanings more clearly than picture or text annotations. These qualitative results corresponded to the results of vocabulary posttests, favoring text-video annotation.
On one hand, the result of text-video annotation leading to better vocabulary retention than text-only annotation is consistent with the dual-coding theory proposed by Paivio (1971, 1986, 1991). The information presentation in both verbal and visual forms helps learners to construct referential connections between the two mental representation processes and achieve better learning. However, the dual-coding effect was not strongly supported by the text-picture annotation in this particular study. On the other hand, the finding of text-video being more effective than text-picture was contradictory to Chun and Plass (1996a). The results obtained by Chun and Plass (1996a) demonstrated higher vocabulary recall scores by text-picture group than text-video group.

Al-Seghayer (2001) explained the different findings of his study from Chun & Plass (1996a) to issues such as the first language of the participants, language used in the annotations, types of video clips, target words, and vocabulary tests used in the two studies. In explaining the better effect by text-video, Al-Segheyer (2001) posited three theoretical explanations. First, he pointed out that the stimuli offered by video (dynamic visual mode) provided “a gestalt” (p. 224) that helped learners to build mental representations in learning the new vocabulary. This explanation agreed with Hanley, Herron & Cole (1995) who argued for video clips as more effective organizers than pictures in foreign language learning. Second, the video annotation stimulated learners’ curiosity and interest level that consequently led to more focused attention and better retrieval. The third reason was related to the redundancy hypothesis (Sherwood, Kinzer, Hasserlbring, & Bransford, 1987) that proposed text-video redundancy to facilitate information processing.
Annotation and L2 Reading Comprehension

Chun & Plass (1996b) is the only study that examined the effects of multimedia annotations, in particular those involving text-picture and text-video annotations, on L2 reading comprehension. Using the same data set from Chun & Plass (1996a), this study hypothesized that dual-coded annotation (verbal and visual) would lead to better reading comprehension than text-only annotation. Following the after-reading vocabulary test, participants were asked to write recall protocols in English (L1), summarizing the story they had just read in the multimedia program. The recall was consequently scored based on 12 main idea units or propositions of the reading text. One point was given to each mentioned proposition, and scores for each student and each proposition were tallied for analysis.

In the analysis, a correlation was sought between annotated words used by students in the recall protocol and the frequency with which these words were mentioned in the recall. The twelve propositions, therefore, were categorized into three groups: propositions that contained words annotated both verbally (text) and visually (picture or video), propositions that contained words annotated verbally (text), and propositions that contained words that were not annotated. The mean scores for the three categories were 2.32 (verbal and visual annotation), 1.53 (verbal annotation), and 1.95 (no annotation). A T-test showed statistically significant differences among the three categories, suggesting the greater recall in reading comprehension of words with both verbal and visual annotations during reading. In other words, words dually annotated are easier and better remembered than words annotated only verbally or no annotated at all. This supported the

Another aspect worth mentioning in Chun & Plass (1996) was the investigation of a possible relationship between vocabulary acquisition and reading comprehension. Specifically, it was hypothesized that better reading comprehension would be shown by students who looked up more annotated words as indicated by higher scores on the written recall protocol. Unfortunately, no correlation was revealed by the data. However, a moderate correlation was found between participants’ vocabulary knowledge and reading comprehension, as shown by the better reading comprehension of students who scored higher in the vocabulary posttest.

Summary

To summarize, the studies reviewed in this section utilized both picture annotation and video annotation coupled with text annotation. The results support the dual-coding theory (Paivio, 1971, 1986, 1990) that annotation provided in both verbal (text) and visual (picture or video) forms facilitated L2 learners’ vocabulary acquisition. It also reflects the cognitive theory of multimedia learning proposed by Mayer (1997, 2001, 2003) that claims that meaningful learning engages learners in both verbal and visual cognitive processing systems. However, which dual-annotation is more effective on vocabulary learning was reported differently by the two studies, with Chun & Plass (1996a) favoring text-picture, though the difference was not always significant, and Al-Seghayer (2001) supporting text-video. Different theoretical explanations have been provided for the evidence of the more effective dual-annotation mode.
In terms of reading comprehension, only Chun and Plass (1996a) addressed this problem. Their finding appears to support the facilitative role of dual-annotation in post-reading recall and reading comprehension. However, text-picture and text-video are put into one category of visual-verbal annotation, and no distinction was articulated between text-picture and text-video. As pointed out by Chun & Plass (1996a), “[T]he difference between pictures and videos as annotations require separate analyses and preclude combining them into a composite category of visual annotations” (p. 190). Moreover, the effect of dual annotation of text-picture is not yet clearly established due to the scarcity of studies (Lomicka, 1998); more research separating the two dual-coding modes is necessary to address any possible differential effects on L2 learners’ overall reading comprehension. Therefore, more empirical evidence is needed to clarify whether there are different effects between text-picture and text-video annotations on L2 reading comprehension.

Audio Annotation

It is worth noting that little research has been done in audio annotation. Audio annotation gives pronunciation, sample sentence, definition or meaning of a target word in spoken form. It has never been studied separately from other annotation modes, but mostly as an additive component. The only format in which audio annotation has been studied is the pronunciation of target words. Findings on audio annotation are rather mixed and uncertain. On one hand, Svenconis and Kerst (1995) suggested that the use of audio could significantly improve vocabulary learning, especially when coupled with a second technique such as semantic mapping. On the other hand, Chun and Plass (1996a)
challenged the effect of audio annotation. In addition, it seems that the addition of an audio component to other annotations is not effective (Yeh & Wang, 2003); instead it distracts learners’ attention. Table 2-4 summaries audio annotation studies.

Svenconis and Kerst (1995) investigated the effectiveness of semantic mapping techniques in L2 vocabulary learning in a hypertext environment. The participants (N=48) of the experiment were English-speaking high school students in grades 9 through 12 learning Spanish as a second language. The 72 target words were presented in two methods, word listing and semantic mapping. In the word listing method, the words were presented in alphabetical order on the computer screen, without any indication of the relationship among the words. In the semantic mapping method, the words were presented in semantic maps that visually explained the relationships among the words. In order to investigate whether the addition of sound would improve vocabulary learning, pronunciation of the target words was used as a factor. The participants were, therefore, divided into four experimental groups: (1) word listing without sound; (2) word listing with sound; (3) semantic mapping without sound; and (4) semantic mapping with sound. In the multiple-choice vocabulary posttest, no significant effect was found for the word presentation method, which suggested that semantic mapping does not necessarily lead to better vocabulary retention than the traditional word listing method. But the group of semantic mapping with sound produced the highest overall mean score of 20.38, higher than the other three groups (word listing without sound had a mean score of 19.82, word listing with sound 18.81, and semantic mapping without sound 17.98). An interaction was identified between vocabulary presentation method and sound factor: the combination of semantic mapping and sound seemed to be positively facilitating L2 vocabulary learning,
while the addition of sound to word listing negatively influenced vocabulary learning. The results suggested that audio annotation could be a significant factor in L2 vocabulary learning if coupled with a second variable, such as semantic mapping. In addition, results of the two delayed posttests demonstrated the increasing effect of sound over time.

Chun and Plass (1996a) challenged the positive effect of audio annotation. In their studies, an audio component was added to three different annotations types (text, text-picture, and text-video); that is, a German native speaker pronounced each target word. Of the three successive studies, participants from study 1 and 2 were asked to report their use of retrieval cues for vocabulary learning. Among the reported cues of text, picture, video and sound, sound was used the least as a retrieval cue, as shown by the percentage of correct answers on vocabulary test, 2.2% and 4.3% for sound cue in the immediate and delayed posttests for study 1, and 0.6% in the immediate posttest for study 2. The authors suggested that the audio component was not useful in learning vocabulary since it showed very limited importance as a retrieval cue.

In comparison to the use of word pronunciation in the above two studies, Yeh and Wang (2003) elaborated the audio annotation in which a native speaker read the word, spelled the word and read the sentence that embedded the word. Three types of vocabulary annotation were examined: text annotation, text-picture annotation, and text-picture-audio annotation. Participants (N=82) were Chinese EFL learners. Analyses of the participants’ performance on the posttests indicated that the text-picture annotation was the most effective for vocabulary learning, and the text-picture-audio annotation was the least effective. This showed that the addition of audio annotation to the text-picture annotation was not effective, and it was even worse than the text-only annotation as
indicated by the group mean scores: highest mean score of 23.41071 from text-picture annotation, followed by mean score of 22.44444 from text-only annotation, and lowest mean score of 20.77777 from text-picture-audio annotation. The difference between text-picture and text-picture-audio was tested by Fisher’s LSD pair comparison to be significant.

Yeh and Wang (2003) offered three possible reasons for the relative ineffectiveness of text-picture-audio annotation. The first reason was L1 processing mechanism transfer as claimed by Chen (1998, cited in Yeh & Wang, 2003) that Chinese EFL learners used more visual strategies than English native speakers and were therefore less skillful in using the provided audio information. This was confirmed by high preference of visual learning style over low auditory learning style by the participants in the questionnaire data. Thus, Chinese students did not effectively process the information provided by the audio annotation. The second reason resided in the fast speech rate of the audio annotation. Coupled with the visual learning style of Chinese students, the fast speech rate distracted the participants and exceeded their listening proficiency. And finally, the combination of text, picture and audio failed to give participants enough time to process the available information.

In summary, studies in audio annotation have mainly engaged the use of pronunciation of the target words, and their findings are inconsistent. Audio is a different sensory modality from visual modality such as printed text and pictures. It should be treated separately as to its effect on learning. In the studies reviewed above, the audio annotation was never studied by itself. Studies of audio annotation should include not only the pronunciation, but also the definition or meaning of the target words. As
suggested by Chun and Plass (1996b), the auditory component needs to be further addressed in terms of its effect on vocabulary learning.

*Annotation Studies Sumary*

Different annotations have been investigated in L2 vocabulary learning in both printed texts and multimedia texts. The effects of text annotation on L2 vocabulary learning and reading comprehension are rather mixed (Davis, 1989; Hulstijn, 1992; Hulstijn *et al.*, 1996; Jacobs, 1994; Jacobs *et al.*, 1994; Joyce, 1997; Ko, 1995; Rott, Williams, & Cameron, 2002; Watanabe, 1997).

Based on the assumption that visual aids such as pictures can facilitate second language vocabulary learning, studies have been carried out to compare picture annotation to text annotation (Kost *et al.*, 1999; Jones, 2003; Yoshii, 2000). The findings are, however, inconsistent, despite the fact that dual annotation of text and picture was unanimously argued to be better than single annotations, especially better than text-only annotation (Kang, 1995; Kost *et al.*, 1999; Jones, 2003; Yoshii, 2000; Yeh & Wang, 2003). The effect of dual annotation of text and picture on L2 reading comprehension is not clearly established due to the scarcity of studies.

Studies utilizing both picture annotation and video annotation coupled with text annotation further support the dual-coding theory (Paivio, 1971, 1986, 1991). Their results of L2 vocabulary learning also confirm the cognitive theory of multimedia learning (Mayer, 1997, 2001) that maintains that meaningful learning engages learners in both verbal and visual cognitive processing systems. However, different studies found advantage of different dual annotations, and offered different interpretations (Al-
Seghayer, 2001; Chun & Plass, 1996a). In addition, in both text-picture and text-video annotations, verbal (text) and nonverbal (picture or video) modes are used to present the information, but only visual sensory-motor is engaged (text, picture, video) in registering the information (Mayer, 1997, 2001)

On the other hand, audio is a different sensory modality from visual modality such as printed text and pictures, but studies in audio annotation have mainly engaged the use of pronunciation of the target words, and their findings are inconsistent. It should be treated separately as to its effect on learning. In addition, studies of audio annotation should include not only the pronunciation, but also the definition or meaning of the target words.

Furthermore, the addition of audio (pronunciation of the target word) does not seem to have a definite effect on L2 vocabulary learning (Chun & Plass, 1996a; Svenconis and Kerst, 1995; Yeh & Wang, 2003). One possible explanation is that the information delivered through different modalities (audio, verbal and visual) might overload cognitive processing; therefore, further research to separate audio and visual in combination with text needs to be addressed (Mayer, 2001). In comparison to text-picture annotation, text-audio annotation uses verbal mode to present information, but engages both visual and auditory sensory modality channels to register the information.

Cognitive Theory of Multimedia Learning

The development of information technologies in recent years has engendered an increasing application of instructional technology in education. Inspired by Wittrock’s (1974, 1990) generative theory and Paivio’s (1971, 1986, 1990) dual-coding theory,
Mayer (1997, 2001, 2003) proposed the cognitive theory of multimedia learning that extends the two theories and maintains a theoretical foundation for interpreting how people learn in multimedia environments. It has been applied and extended by Plass, Chun, Mayer and Leutner (1998) to multimedia second language learning.

*Generative Theory of Learning*

In 1974, Wittrock introduced the generative model of learning. His model posited that human learning was “a function of the abstract and distinctive, concrete associations which the learner generates between his prior experience, as it is stored in long-term memory, and the stimuli” (p. 89). In other words, this model emphasizes the active integration of new ideas with the learner's existing schemata, particularly by using four types of learning strategies: 1) recall, 2) integration, 3) organization, and 4) elaboration (Wittrock, 1974, 1990).

In recall, the learner uses techniques such as repetition, rehearsal, review or mnemonics to pull information out of long-term memory. Methods such as paraphrasing, summarizing, or questions-generalizing are applied to integrate new information with prior knowledge; thereafter, the learner utilizes techniques such as outlining, categorization, clustering or concept mapping to connect new information to prior knowledge in an organized way. Finally, given an established connection between the new information and the learner’s prior knowledge, the new information is synthesized and elaborated through mental images, writing, visual displays and similar methods. Individual strategies are used alone or in collaboration with others to achieve a learning goal.
Dual-Coding Theory

The dual coding theory established by Paivio (1971, 1986, 1990) offers an account of both verbal and non-verbal cognitive processes. Paivio (1986) stated, "Human cognition is unique in that it has become specialized for dealing simultaneously with language and with nonverbal objects and events. Moreover, the language system is peculiar in that it deals directly with linguistic input and output (in the form of speech or writing) while at the same time serving a symbolic function with respect to nonverbal objects, events, and behaviors. Any representational theory must accommodate this dual functionality" (p 53).

Dual-coding theory is deeply influenced by the imagery-based mnemonic technique that was rooted in the ancient method of loci (Sadoski & Paivio, 2001). The mnemonic technique applies verbal-imaginal dual coding in which the targeted words (verbal) are transformed into non-verbal images, and images are consequently translated back into verbal in recall activities. This technique is widely used in learning a second or foreign language.

One major assumption of dual-coding theory is that human mental process derives perceptions from various sensory modalities, and these modality-specific representations can be divided into two categories: verbal representations as those derived from speech or writing, and non-verbal representations as those derived from images (Sadoski & Paivio, 2001). In other words, the theory entertains two cognitive subsystems, one represents
and processes language input, and the other specializes in interpreting nonverbal objects/events such as imagery.

Dual-coding theory identifies three types of processing: (1) representational, the direct activation of verbal or non-verbal representations, (2) referential, the activation of the verbal system by the nonverbal system or vice-versa, and (3) associative processing, the activation of representations within the same verbal or nonverbal system. The two subsystems described by dual-coding theory function independently, parallely and interconnectedly (Paivio, 1986, 1990; Sadoski & Paivio, 2001).

**Cognitive Theory of Multimedia Learning**


The dual channels assumption argues that humans possess separate channels for processing visual/pictorial and auditory/verbal information (Mayer, 2001). In this respect, Mayer combines the presentation-mode approach and sensory-modality approach. On one hand, consistent with the distinction between verbal and nonverbal systems that respectively process verbal and nonverbal information (Paivio, 1986), Mayer (2001) conceptualizes the presentation-mode by distinguishing verbally presented information such as printed text or spoken words, and nonverbally presented information such as pictures or animations. On the other hand, Mayer (2001) adopts the sensory-modality approach (Baddeley, 1986, 1992, 1999) by differentiating the visual channel (eyes) and
the auditory channel (ears). The visual channel processes information such as pictures, video or printed texts, while the auditory channel processes information such as spoken words or background sounds.

In the second assumption, each channel is assumed to have limited cognitive capacity. In other words, when facing new information presented via multimedia, learners are limited in the amount of information that can be processed in either the visual or auditory channel at one time. For example, if the learner is presented a series of pictures on the computer screen, although all the pictures may be perceived by the visual modality, only a few images can be brought into and held in the working memory at any one time. The same applies to the auditory modality. For example, in language learning, when a learner listens to a dialog, only a portion of the dialogue can be selected to be processed in the learner’s working memory, with possible subsequent integration into long-term memory.

In active processing, Mayer borrowed from generative learning (Wittrock, 1974, 1990) the concept that meaningful learning engages three essential cognitive processes, namely, selecting relevant information from newly presented materials, organizing the selected information into coherent mental representations, and integrating organized information with prior knowledge (Mayer, 1997, 2001). In typical multimedia learning, at the first stage, when presented with a multimedia message, the learner pays attention to the presented materials, processes the information through the auditory and visual channels, and brings selected words and pictures from sensory memory into working memory. At the second stage, the selected words and pictures are processed and organized in the working memory into coherent verbal and nonverbal representations;
sometimes the verbal and nonverbal representations interact with each other referentially (Paivio, 1971, 1986, 1990). At the third stage, connections are established between the verbal and nonverbal models in relation to the learner’s existing knowledge.

Figure 2-1 presents a model of cognitive theory of multimedia learning (Mayer, 1997, 2001, 2003). Three memory stores are represented: sensory memory, working memory and long-term memory. In multimedia learning, materials are normally presented in verbal (including written text and spoken text) and visual (including pictures, video, animation) forms. These presented materials enter the learners’ sensory memory through the visual channel (eyes) and the auditory channel (ears). In other words, the eyes register visual images (including pictures, printed texts) and bring them into the visual sensory memory; the ears register auditory images (including spoken text or background sounds) and bring them into the auditory sensory memory. Due to limited cognitive capacity, only selected sounds and images enter the working memory where they are temporarily held, manipulated and constructed into verbal and visual representations. Finally, long-term memory is activated to integrate the newly selected and manipulated information. In this model, the learner is regarded as an active knowledge constructor who selects, organizes and integrating both visual/pictorial and auditory/verbal information such as printed texts, pictures, or spoken words (Mayer, 1997, 2001).

*Extension of Cognitive Theory of Multimedia Learning*

The use of actual objects or imagery techniques in foreign language vocabulary learning is well documented in research (Davis, 1989; Kellogg & Howe, 1971),
especially along with the advance of multimedia application in language teaching and learning. Plass et al (1998) argued for an extension of the cognitive theory of multimedia learning to second language learning. In multimedia language learning, both visual and auditory sensory modalities are engaged to process information that is consequently represented in both verbal and visual models. For example, when presenting a new word via a multimedia program, both the word in printed text and a picture describing the words can be presented on the computer screen; therefore the L2 learners can select relevant information (either text, picture, or both text and picture) to process in the working memory and integrate the newly selected information into the long-term memory, that is to memorize the meaning of the word.

![Figure 2-1. Cognitive Theory of Multimedia Learning (adapted from Mayer, 1997, 2001, 2003).](image)

Their extension addressed two aspects of multimedia second language learning: vocabulary learning and reading comprehension. In terms of vocabulary learning, the extension maintained that “second language learners possess two separate verbal systems and a common visual system” (Plass et al, 1998, p. 26). They further argued that second
language learners “learn new words when they can establish a direct connection between the words in their native language, the corresponding picture of an object or action, and its foreign equivalent” (Plass et al, 1998, p. 26).

The second focus of the extension is reading comprehension. Based on the arguments of Wittrock (1974, 1990), the reading process engages the referential connections between the mental representations of information or ideas presented in different modes such as the visual mode and verbal mode. In multimedia language learning, the presentation of information in multiple modes might facilitate L2 learners’ reading comprehension (Chun & Plass, 1996b; Ommagio, 1979).

Application of Cognitive Theory of Multimedia Learning to Multimedia Annotation

Mayer (2001) defines multimedia as the “presentation of material using both words and pictures” (p. 2). Words refer to materials that are presented in verbal forms, such as printed text or spoken text; pictures refer to materials that are presented in pictorial form such as static pictures, video, or animation. Mayer (2003) refined the definition by breaking it down into what, how and why: what refers to the instructional content, how refers to the presentation method that entails both the use of computers and different instructional methods such as using different modalities, and why refers to the purpose of multimedia learning being to promote learning and foster changes in learner’s knowledge and performance.

Multimedia learning is designed in accordance with the cognitive process of the human mind. Dual-coding theory argues for two subsystems of information processing, one for verbal information and one for visual materials (Paivio, 1971, 1986, 1990).
Therefore, multimedia is taking advantage of the full capacity of human information processing systems. Mayer (2001) explains the advantages of multimedia learning both quantitatively and qualitatively: (1) presenting the same material in two channels is better than one channel by offering the information twice (redundancy principle); (2) words and picture complement each other in facilitating human mental representations of both visual and verbal information.

Multimedia annotation, as an instructional intervention in L2 vocabulary learning, conforms to the definition of multimedia and applies the cognitive theory of multimedia learning. This section will explain this point by breaking the multimedia annotation into three different modes: text, picture and audio.

*Figure 2-2. Processing of text annotation*

In Figure 2-2, the shaded boxes indicate the cognitive processing of text annotation. When the text annotation appears on the computer screen, L2 learners, using eyes the visual sensory modality, perceive the printed text, select from available visual information, and send it to working memory in the form of sensory images. The sensory
images of the printed words are converted to corresponding sounds that are further
organized into verbal models of representation.

The shaded boxes in Figure 2-3 describe the path for processing audio annotation.
Audio annotation is perceived by the ears and is held briefly in the L2 learner’s auditory
sensory memory. If the sounds are given active cognitive processing, some of them will
be selected to be included in the sound base of the working memory. Then the selected
sound fragments will be organized into coherent mental representations, and built into the
verbal model in collaboration with the prior knowledge.

Figure 2-3. Processing of audio annotation

Figure 2-4. Processing of picture annotation
As shown by Figure 2-4, the shaded boxes explain how picture annotations are processed. When the annotation is given in picture form, L2 learners perceive the pictures via visual sensory modality, select relevant visual information and hold it temporarily in the working memory. Additional processing will bring the visual information into pictorial model and organize it into mental representation. In the next step, active cognitive processing connects the new pictorial representation to existing knowledge that results in integrated learning outcome.

Chapter Summary

This review of the related literature has discussed the importance of vocabulary for L2 learning. Different annotations have been investigated in L2 vocabulary learning in both printed texts and multimedia texts. Supporting Paivio’s dual-coding theory (1971, 1986, 1991), dual annotation of text-and-picture was unanimously argued to be better than single annotations (Kang, 1995; Kost et al., 1999; Jones, 2003; Yoshii, 2000; Yeh & Wang, 2003). Studies utilizing both picture annotation and video annotation coupled with text annotation further supported dual-coding theory, and their results also confirmed the cognitive theory of multimedia learning (Al-Seghayer, 2001; Chun & Plass, 1996a; Mayer, 1997, 2001) that claims that meaningful learning engages learners in both verbal and visual cognitive processing systems. The effect of dual annotation of text and picture on L2 reading comprehension is not clearly established due to the scarcity of studies. However, different studies found an advantage of different dual annotations, and offered different interpretations (Al-Seghayer, 2001; Chun & Plass, 1996a).
Studies in audio annotation have mainly engaged the use of pronunciation of the target words and their findings are inconsistent (Chun & Plass, 1996a; Svenconis and Kerst, 1995; Yeh & Wang, 2003). As a different sensory modality from visual modality such as printed text, pictures and videos, audio should be treated separately as to its effect on learning. Studies of audio annotation, therefore, should include not only the pronunciation, but also the definition or meaning of the target words. Finally, more evidence is necessary to investigate audio and visual separately in combination with text (Mayer, 2001).

Studies investigating dual multimedia annotations have only examined the different effects between text-picture and text-video annotations. However, in both text-picture and text-video annotations, verbal (text) and nonverbal (picture or vide) modes are used to present the information, but only visual sensory-motor is engaged (text, picture, video) in registering the information (Mayer, 1997, 2001). In comparison, audio-picture annotation uses both verbal and nonverbal modes to present information, but engages both visual and auditory sensory modality channels to register the information. No study has ever examined the possible differential effect of audio-picture annotation in comparison to text-picture annotation.

Most annotation studies involve subjects from a single L1 background, except Yoshii (2000) and Al-Saghayer (2001); therefore, there is a need to further examine a participant population with multiple first languages. In this case, ESL students from a range of first language backgrounds would be a good sample. Various studies have found that certain L2 proficiency is necessary to effectively use annotations; therefore,
intermediate ESL students were chosen to participate in this study (Dufon & Hong, 1994; Jacobs, 1994; Laufer & Shmueli, 1997).

Only two studies have addressed the relationship between vocabulary annotation and reading comprehension in a multimedia setting (Chun & Plass, 1996b; Lomicka, 1998). As an instructional intervention, multimedia annotation does briefly interrupt the reading process by drawing learner attention away from the intended reading goal. It is, therefore, necessary, when examining L2 vocabulary learning, to further investigate the effects of annotations on text comprehension.

An overview of the studies on L2 vocabulary annotation, particularly multimedia annotation, suggests that there is little information about how different dual annotations, in particular text-picture and audio-picture annotations, affect L2 vocabulary learning and reading comprehension. Specifically, in ESL classrooms, there is little information on how these dual annotations facilitate ESL students’ vocabulary learning through reading. This information is needed to understand the extent to which multimedia learning can be used in L2 reading instruction and the role of multimedia in L2 vocabulary learning.

The effects of multimedia dual annotations employing different modalities on L2 vocabulary learning and reading comprehension are still an unresolved problem. In addition, there has been no study to compare incidental learning and intentional learning in a multimedia setting.
Chapter III Methods

The previous chapter reviewed the related literature on L2 vocabulary learning, annotation studies, and cognitive theory of multimedia learning. This study used a between-subjects design to compare the effects of dual multimedia annotation utilizing different sensory modalities (text-and-picture, audio-and-picture) on intermediate L2 students’ vocabulary immediate recall and reading comprehension in both incidental and intentional learning conditions. In particular, the following six research questions were addressed:

1. What are the different effects of text-picture and audio-picture annotations in facilitating L2 vocabulary immediate recall?

2. What are the different effects of text-picture and audio-picture annotations in facilitating L2 reading comprehension?

3. Which learning condition (incidental or intentional) results in better L2 vocabulary immediate recall?

4. Which learning condition (incidental or intentional) results in better L2 reading comprehension?

5. Is there an interaction between annotation type and learning condition on immediate L2 vocabulary immediate recall?

6. Is there an interaction between annotation type and learning condition on L2 reading comprehension?
The present chapter describes the design of the study, the participants and the setting, instruments, the data collection procedures and the methods of data analysis.

*Design of the Study*

This study investigated the effect of different types of multimedia annotations on L2 vocabulary immediate recall and reading comprehension. Students’ vocabulary immediate recall and reading comprehension will be compared in two annotation types (text-picture and audio-picture) and under two learning conditions (incidental and intentional.) In other words, the study was designed to investigate different learning conditions and instructional interventions (providing different multimedia annotations to unknown words) on vocabulary immediate recall and text reading comprehension.

This study adopted a between-subjects design immediate posttests. There were two dependent variables: L2 vocabulary immediate recall and reading comprehension. There were two independent variables: learning conditions (intentional and incidental) and annotations (text-picture, and audio-picture). In total, as shown by the 2 x 2 factorial design in Table 3-1, there were four treatments: intentional leaning in text-picture, intentional learning in audio-picture, incidental learning in text-picture, and incidental learning in audio-picture. A more detailed research design can be found in Figure 3-1.

**Table 3-1**

*2 x 2 Factorial Design*

<table>
<thead>
<tr>
<th>Annotation</th>
<th>Text-picture</th>
<th>Audio-picture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning</td>
<td>Intentional</td>
<td>Incidental</td>
</tr>
</tbody>
</table>
For the vocabulary posttest, only immediate posttests, Vocabulary Knowledge Scale (VKS) and Word Recognition Test (WRT), were used in this study. The delayed posttest was not used because the focus of the study was to investigate the effects of different annotations on vocabulary immediate recall after an initial exposure of the targeted words in a reading text. Hulstijn (2001) argues for this practice with two reasons: (1) after initial exposure to new information, people tend to forget the
information if there are no additional exposures or rehearsals; therefore, the fall in performance on delayed posttest is expected; (2) delayed posttest will be necessary if the research focus is on what happens with the new information after initial exposure under various conditions of rehearsal or multiple exposures.

The scores of the immediate vocabulary posttests were used to measure the participants’ vocabulary learning. L2 vocabulary learning involves multiple factors, such as phonology, morphology, syntax, and semantics (Laufer, 1990), and knowing a word includes multiple components, such as spelling, pronunciation, grammatical form, relative frequency, and collocations (Richards, 1976; Nation, 1990). The distinction between receptive and productive vocabulary generally refers to having some knowledge of a word and being able to use it in speech and writing (Melka, 1997; Read, 2000). This study adopted the distinction made between receptive knowledge and productive knowledge by Nation (1990) and Read (2000) that receptive knowledge refers to the ability to recognize and recall the meaning of a target word, while productive refers to the ability to use the target word in the learner’s own speech or writing.

Considering the fact that the participants in the study were not taught the features of the target words explicitly, and they only had limited exposure to the target words through reading, vocabulary immediate recall in this study falls into the category of receptive learning. Thus vocabulary immediate recall in the present study means the ability to recognize and recall the meaning of the target words after reading.

The following section of this chapter describes in more details the participants and the setting, instruments, the data collection procedures and the methods of data analysis.
Variables

Dependent Variables

The dependent variables in this study were students’ vocabulary immediate recall and reading comprehension. Vocabulary immediate recall was measured by two vocabulary posttests: VKS (Vocabulary Knowledge Scale) and WRT (Word Recognition Test). Reading comprehension was measured by multiple-choice Reading Comprehension questions and a Written Recall.

Independent Variables

Types of annotation. The primary independent variable was the type of multimedia annotation that was available to students during the reading process. Two different dual annotations were included to investigate whether they had different effects on L2 vocabulary immediate recall and reading comprehension. A text-picture annotation provides a textual definition/explanation of the unknown word and a picture that describes the unknown word. An audio-picture annotation offers a picture description of the unknown word along with an audio explanation of the word spoken by a native speaker.

All the pictures were taken from the World Wide Web. Each picture was chosen to express the essential meaning of the target word as understood in the reading context. A native speaker of English was asked to read and record the audio annotation.

The definitions and pictures were also given to experienced ESL instructors and non-participating ESL students for verifications. The definitions were examined and
revised to be compressible to ESL students and expressive of the contextual meanings of
the target words. The pictures were judged on the criteria of being graphically descriptive
of the word meanings. Necessary revisions and modifications were carried out.

Types of Learning. With respect to the learning conditions, Hulstijn’s (1992)
definition for types of learning was used. The type of learning was determined to be
either intentional or incidental based on the learners’ responsibility for the learning task,
which was specified in advance of the task. Therefore, participants who were told in
advance that they would have a reading comprehension test after reading the text were
treated as incidental vocabulary learning subjects. Those participants who were told in
advance that they would have vocabulary tests and reading comprehension tests after
reading were treated as intentional vocabulary learning subjects.

Two different learning conditions were included in order to test whether a
predetermined objective would influence students’ performance on comprehending the
text and retaining vocabulary. Particularly in terms of vocabulary immediate recall, the
intentional learning condition was assumed to draw more student attention to the target
words, which, in turn, was believed to ensure depth of processing and facilitate
vocabulary learning (Hulstijn, 1992).

Other Independent Variables Controlled

For the control of participants’ previous knowledge of the target words, the target
words were identified by non-participating intermediate ESL students and confirmed by
their instructors to be unfamiliar to typical intermediate ESL students.
Other relevant independent variables such as age, gender, and computer familiarity were controlled by random assignment of the participants into the four conditions.

*Participants and setting*

The participants of this study were 85 intermediate-level ESL students from three universities in northwest United States. They were enrolled in the Intensive English Program (IEP) at these universities. The individual institutes define the intermediate level using similar placement tests. These universities provide ESL classes at a variety of proficiency levels, ranging from beginning to advanced levels.

All the three IEPs offer similar curricula that recognize the need for both fluency and accuracy to prepare non-native speakers of English to meet their academic and professional goals. Instruction in IEPs strives to make use of authentic, meaningful content and to present form as a facilitator of communicative and social interaction, rather than an end. It is the IEP philosophy that authentic, meaningful content involves an integration of, rather than a separation of, the linguistic skill areas of listening, speaking, reading, writing, and grammar.

Their IEP curriculum offers classes at six proficiency levels: beginning (levels 1-2), intermediate (levels 3-4), and advanced (levels 5-6). The students were assigned into different proficiency levels based on their performance on the Michigan English Placement Test (EPT) which has 20 listening items, 30 grammar items, 30 vocabulary items and 20 reading items. The TPT was supplemented by an oral interview and a writing test. The scores for level classification are shown in Table 3-2.
**Table 3-2**

*Test scores and matching levels at one English Language Institute*

<table>
<thead>
<tr>
<th>Level</th>
<th>EPT Listening (20 items)</th>
<th>EPT Structure (80 items) Grammar, Vocabulary, Reading</th>
<th>Oral Interview (eg: Fred Test)</th>
<th>Writing test (Holistic)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0-4</td>
<td>0-19</td>
<td>0-35</td>
<td>1+, 1</td>
</tr>
<tr>
<td>2</td>
<td>5-7</td>
<td>20-30</td>
<td>36-48</td>
<td>2+, 2, 2-</td>
</tr>
<tr>
<td>3</td>
<td>8-10</td>
<td>31-41</td>
<td>49-60</td>
<td>3+, 3, 3-</td>
</tr>
<tr>
<td>4</td>
<td>11-14</td>
<td>42-52</td>
<td>61-70</td>
<td>4+, 4, 4-</td>
</tr>
<tr>
<td>5</td>
<td>15-17</td>
<td>53-65</td>
<td>71-80</td>
<td>5+, 5, 5-</td>
</tr>
<tr>
<td>6</td>
<td>18-20</td>
<td>66-80</td>
<td>80+</td>
<td>6+, 6, 6-</td>
</tr>
</tbody>
</table>

In terms of test scores, an intermediate level ESL student will have a score of 31-52 on the structure section of the EPT test, a score 8-14 on the listening section of the EPT test (a total of 39-66 on the EPT test). The oral interview score and writing test score will be respectively 49-70, and (3-)-(4+). The cutting scores on EPT at these English Institutes correspond to the cutting scores set by the University of Michigan (46-74 for intermediate level ESL students).
<table>
<thead>
<tr>
<th>Country</th>
<th>Language</th>
<th>Number of Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japan</td>
<td>Japanese</td>
<td>30</td>
</tr>
<tr>
<td>Taiwan</td>
<td>Chinese</td>
<td>14</td>
</tr>
<tr>
<td>Korean</td>
<td>Korean</td>
<td>15</td>
</tr>
<tr>
<td>Thailand</td>
<td>Thai</td>
<td>2</td>
</tr>
<tr>
<td>Saudi Arabia</td>
<td>Arabic</td>
<td>6</td>
</tr>
<tr>
<td>Italy</td>
<td>Italian</td>
<td>1</td>
</tr>
<tr>
<td>Vietnam</td>
<td>Vietnamese</td>
<td>1</td>
</tr>
<tr>
<td>Mexico</td>
<td>Spanish</td>
<td>3</td>
</tr>
<tr>
<td>France</td>
<td>French</td>
<td>1</td>
</tr>
<tr>
<td>Germany</td>
<td>German</td>
<td>2</td>
</tr>
<tr>
<td>Brazil</td>
<td>Portuguese</td>
<td>1</td>
</tr>
<tr>
<td>Peru</td>
<td>Portuguese</td>
<td>1</td>
</tr>
<tr>
<td>El Savado</td>
<td>Spanish</td>
<td>1</td>
</tr>
</tbody>
</table>

The intermediate language proficiency of the participants in this study were validated by the instructors at these institutes and the program academic coordinators as well as with reference to the participants’ scores on the placement tests given when they entered the IEPs.

Originally 85 students participated in this study. After data collection, seven students who did not look up any hyperlinked words during reading were excluded from
the study. As a result, the final sample size was 78. Table 3-3 shows the countries and languages represented by the participants. There were 49 female students and 29 male students. 69% of the participants were in the age group of 18-25, 24% were in the age group of 26-45, and 7% were over 45. In terms of length of stay in the United States, the majority of the participants fell into the category of 0-6 months, 82% of the participants reported 0-6 months, 15% reported 6-12 months and 3% reported over a year. All the participants reported being comfortable using computers and reading a web-based English text on the computer.

**Instruments**

*Questionnaire*

A questionnaire was used to collect demographic information from the participants, general English learning experience, and their experience with using computers. The questionnaire can be found in Appendix A.

The questionnaire contains three major parts: demographic information, English language learning experience, and computer experience. The portion dealing with demographic information was designed to gather information about the participants’ age, gender, national origin, native language, and length of stay in the United States. Another portion of the questionnaire gathered information on the participants’ English language learning experience before and after arriving in the United States. The last part of the questionnaire relates to the participants’ experience and familiarity with technology. Participants were asked to offer information on their general computer skills, access to the Internet, and online reading ability.
Reading Material and Target Words

The reading text, “European Settlers of Australia,” was written by the researcher based on three criteria: text length, syntactic complexity, and content. In terms of length, the text has 449 words (including the title). The reading text can be found in Appendix B. ESL students at the intermediate level are comfortable with reading texts of this length and syntactic complexity. It consists of short, uncomplicated sentences and simple past tense is used throughout the text. There is an average of 6.8 sentences in each paragraph, and an average sentence contains 10.8 words. The percentage of simple sentences in the text is over 80%. With regard to the content, it seems reasonable to assume that ESL students knew more or less the same amount of general information about the European colonization of Australia and have comparable background knowledge of the reading text (i.e., since none has been to Australia and its history is foreign to all participants). The content of the text does not require any specific culturally related knowledge.

The readability of the text is considered to be between grade level 5 and 6 based on the Flesch-Kincaid measure. It tells of the story of the European colonists in Australia in the 1800s. The text was given to experienced ESL instructors who teach intermediate reading/writing classes and was confirmed to be appropriate for intermediate ESL students. A cloze procedure was conducted with non-participating intermediate ESL students to estimate the difficulty of the text. The student’s cloze score of 67% indicate that the reading text was appropriate for intermediate ESL students in terms of difficulty level.
The 20 target words were all nouns. They were selected for frequency. Based on the word frequency corpora of Francis and Kucera (1982), the 20 target words have a mean of 12.7 per million words. The 20 target words can be found (underlined words) in Appendix C.

The 20 target words constitute 4.5% of the total words of the text, leaving 95.5% coverage of the reading text. Nation (2001) points out that 98-99% coverage is desirable for relatively easy reading for ESL students. However, he argues for a smaller coverage of the reading text in annotation studies because the mixed findings of the effects of annotations on reading comprehension might be due to the percentage of unknown words being less than 3% of the running texts in most experimental studies. Therefore, he suggests the use of more unknown words that might initiate the effect of annotation on reading comprehension.

The reading text was modified into two different forms: a text with text-picture annotations, and a text with audio-picture annotations. The 20 target words were highlighted in both texts.

**Program**

The interactive multimedia program used in this study was designed by the researcher to help intermediate ESL students with vocabulary learning and reading comprehension. The program provided students with annotations for unknown words via hypermedia links in two different modes: text-picture and audio-picture. The annotations were used to assist the learning of unknown words and understanding of the reading text.
The program was written in HTML and JavaScript. HTML was chosen as the programming language due to its user-friendly integration of hypermedia and its compatibility for both PC and Mac platforms. The picture annotations were processed with Adobe Photoshop 6.0 (Adobe, 2000), and the audio clips were processed with Vegas 4.0 (Sonic Foundry, 2003). The pictures for the 20 target words can be found in Appendix D. The program was delivered over the Internet and was accessed using either Internet Explorer (version 5.0 or higher) on PCs or Safari on Mac computer.

The screen layout took into consideration instructional design principles. The screen was divided into two frames. The left screen was used for the reading text with the title at the top, and the right screen was reserved for the annotation. In the text-picture version, when participants click on a highlighted word, the right screen offers a textual definition of the words together with a picture that describes the word. In the audio-picture annotation, when participants click on a highlighted word, they could see on the right screen a picture that depicts the meaning of the word and hear an audio clip that explains the meaning of the word. The screenshots of text-picture and audio-picture can be found in Appendix E and F.

In order to count the number of times that a participant accesses an annotated word in the reading text, a JavaScript variable was created for each highlighted word and encoded into the HTML for the reading text. Each time the participant clicked a certain highlighted word, a variable was incremented by one with the “OnClick” JavaScript command. These variables were connected with the participant’s identification code and computer number. Upon finishing reading the text, when the participant clicked the “Submit” button, the PHP sendmail script passed the variables to the researcher’s email.
account. The transferred file contained the participant’s identification code, computer number, annotation type and the clicking behavior recorded during the reading process.

Posttests

*Vocabulary Knowledge Scale (VKS).* The Vocabulary Knowledge Scale (VKS) was modified from Paribakht & Wesche’s (1993, 1997) Vocabulary Knowledge Scale (VKS). The VKS was designed as a measure of the stages in a learner’s developing knowledge of particular words. In the original instrument, a 5-level scale was used to rate vocabulary knowledge that ranges “from complete unfamiliarity, through recognition of the word and some idea of it’s meaning, to the ability to use the word with grammatical and semantic accuracy in a sentence” (Paribakht & Wesche, 1997, p.29).

In this study, an adapted VKS (See Table 3-4) is compiled in consideration of three factors: (1) translation is not encouraged since the participants represent multiple first languages; (2) since the participants only have one exposure to the targeted words from the reading text, therefore Level IV (recognize the word and be able to give the meaning of the word) and Level V (be able to use the word in a sentence) are not required; and (3) the focus of the study is vocabulary learning that is operationalized as the immediate recall of word meaning. Therefore, grammatical accuracy of the explanation is not graded. The VKS test can be found in Appendix G.
**Word Recognition Test (WRT).** The participants were asked to complete a Word Recognition Test (WRT). In this test, the 20 target words were presented in their original context taken from the reading text. For each word, the participants were asked to choose one correct meaning out of four given choices. Of the four choices, one was the correct meaning, and the other three were distractors. The WRT test can be found in Appendix H.

### Table 3-4

*Adapted VKS scoring categories and meaning of scores.*

<table>
<thead>
<tr>
<th>Categories</th>
<th>Description</th>
<th>Possible score</th>
<th>Meaning of score</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>I don’t remember having seeing this word before</td>
<td>0</td>
<td>The word is not familiar at all.</td>
</tr>
<tr>
<td>II</td>
<td>I have seen this word before, but I don’t know what it means</td>
<td>1</td>
<td>The word is recognized, but the meaning is unknown.</td>
</tr>
<tr>
<td>III</td>
<td>I have seen this word before, and I think it means________ (providing explanation)</td>
<td>2</td>
<td>A word is recognized, and a correct meaning is given.</td>
</tr>
</tbody>
</table>

**Multiple-choice Reading Comprehension Test (RC).** The reading comprehension text was comprised of 10 multiple-choice questions. For each question, the participant was asked to choose the best answers from the four given choices. The questions and choices were given to experienced ESL instructors for validation. The questions were confirmed to be easy to understand and reflective of main idea of the reading text. The Reading Comprehension Test can be found in Appendix I.
Written Recall. Written recall was used to measure participants’ general text comprehension of the reading text. Due to the English level of the participants, the written recall was conducted in the participants’ native language. The written recall can be found in Appendix J.

Validity Issue of the Posttests

Validity refers to the degree to which a study accurately reflects or assesses the specific concept that the researcher is attempting to measure. In other words, validity is concerned with the study's success at measuring what the researcher set out to measure. Due to the nature of the achievement test used in this study, content validity will be the focus of discussion here. Content Validity is based on “the extent to which a measurement reflects the specific intended domain of content” (Carmines & Zeller, 1991, p.20).

In this study, vocabulary immediate recall was operationalized as the ability to recognize and recall the meaning of target words when presented. For this particular purpose, the VKS has been proven to be a valid measure as a self-rating report of word knowledge (Paribakht & Wesche, 1993,1997). For WRT, the choice of the right or wrong answer reflects the participants’ knowledge of the target word. Experienced ESL instructors and ESL researchers were consulted in terms of the subject matter of the test items to ensure their valid interpretation.

The written recall and reading comprehension test aimed to assess participants’ general understanding of the reading text. In other words, participants were supposed to recall the main ideas of the text after reading and choose the best answer for each of the
comprehension questions. The multiple-choice questions were given to three experienced ESL instructors for their content validity. A consensus was reached that questions reflect and measure the participants’ general knowledge of the reading text.

Scoring

The researcher rated the scores for Word Recognition Test and Reading Comprehension Test. Two raters were assigned to rate each student’s Vocabulary Knowledge Scale and Written Recall. One rater was the researcher and the other was one experienced ESL instructor from one of the IEPs. Each rater scored the tests individually. Inter-rater reliability, which was calculated by the percentage of agreement between the two raters, was used to examine the reliability of the measurements. Where there were discrepancies in the rating between the two raters, a reexamination of the responses by the two raters was conducted to reach an agreement. If an agreement was not reached, a third rater was used or an average score of two raters was accepted.

Vocabulary Knowledge Scale (VKS). The VKS was scored using a 0-2-point scale adapted from Paribakht & Wesche (1997). A score of 0 indicates total unfamiliarity of the word; a score of 1 stands for recognition of the word, but the meaning of the word is not provided; a score of 1.5 shows recognition of the word and partial meaning provided; and a score of 2 represents familiarity of the wording shown and correct meaning given. The possible total score was 40 points (2 points x 20 words).
Word Recognition Test (WRT). A correct choice received the score of 1 and an incorrect choice received the score of 0. The possible maximum score was 20 points (1 point x 20 words).

Multiple-Choice Reading Comprehension Test (RC). A 0-1-point scale was used for the reading comprehension test. In the 0-1 point scale, 0 represents no response or an incorrect choice; 1 represents a correct choice. The comprehension score for each participant was calculated by combining the score for each of the 10 multiple-choice questions together, with a possible maximum score of 10 (1 point x 10 questions).

Written Recall. The researcher and another experienced ESL instructor independently identified the idea units or propositions in the reading text. Through discussion, it was agreed that there were a total of 52 idea units in the reading text. All the written recall protocols were scored in terms of the 52 idea units. Each unit idea presented in the written recall was given 1 point. Paraphrasing was allowed and counted correct recall of idea units.

Procedures
The study was conducted during the participants’ regular class times, and required two consecutive 50-minute sessions. The participants were randomly assigned to the learning conditions: incidental learning and intentional learning. The two learning conditions were conducted separately at different times. In each of the learning conditions, participants were further randomly divided into two annotation groups (text-
picture, audio-picture). Each student created a unique identification code that they remembered and used in the study. For the identification code, they used the combination of the last letter of their last name and the day of the month in which they were born. For example, Bob Smith, born on the 15th of May, would use the code of H15.

For each learning condition, in the first 50-minute session, the researcher first gave a brief introduction of the study and answered any questions that the participants might have. Then the participants signed the consent form and filled out the questionnaire. The questionnaire asked for students’ demographic information such as gender, age, first language, years of English learning, length of stay in the U.S., and their familiarity with computers.

In the second session, students went to the language lab at their IEPs. The computers in the lab were preset to present different annotations for every two neighboring computers. In other words, two neighboring students had access to different annotations, one was text-picture and the other was audio-picture. In the computer lab, the researcher gave a brief introduction of the online reading activity, and an online tutorial was used to get the participants familiar with the online reading activities. Headsets were used for those who were in the audio-picture annotation group.

During reading, the participants clicked the highlighted unknown words to access available annotations, the text-picture group was able to see textual explanation and pictorial description, and the audio-picture group was able to see pictorial description explanation and hear a spoken explanation. When they finished reading, they raised their hands to receive the posttests.
Participants in the incidental learning condition were told before reading that they would have reading comprehension tests, but not informed of the vocabulary posttests. Participants in the intentional learning condition were told that they would have vocabulary posttests and reading comprehension tests up front. The same reading vocabulary posttest and reading comprehension tests were conducted with these participants after reading. The posttests were given in the following order: Vocabulary Knowledge Scale (VKS), Word Recognition Test (WRT), Multiple-choice Reading Comprehension Test (RC), and Written Recall.

**Data Analysis**

The data were analyzed with the SAS package. A 2 x 2 ANOVA was performed for each of the posttests: vocabulary posttests (VKS, WRT) and reading comprehension tests (Multiple-choice Reading Comprehension Test and Written Recall). The data were analyzed and examined in terms of the effect of annotation type (Text-picture and Audio-picture), the effect of learning condition (Incidental and Intentional), and the interaction effect between the two factors (annotation type and learning condition).

The internal consistency reliability for the dependent measures was estimated using Cronbach’s alpha. Inter-rater reliability was used as a measure for Vocabulary Knowledge Scale and Written Recall.

An additional ANOVA was conducted for lookup frequency behaviors. An interaction was found between the annotation type and learning condition, further correlation analysis was used to examine the relationship between the lookup frequency behavior and dependent measures.
The alpha level was set at .05 for all the analyses. Descriptive statistics for the posttest scores were calculated for each analysis. Relevant statistical data were also presented with tables and graphical figures.

This chapter describes the design, participants and setting, instruments, procedures of data collection and data analysis of this study. The results of data analysis will be presented in the next chapter.
Chapter IV Results

The purpose of this chapter is to communicate the data analysis results as well as report the findings related to each research question.

The first part of this chapter presents the results of a pilot study conducted as a precursor to the dissertation study. The second part of this chapter introduces results obtained from the four instruments --- Vocabulary Knowledge Scale, Word Recognition Test, Multiple-choice Reading Comprehension Test and Written Recall. After the presentation of results, each of the research questions is examined and answered in the third part of this chapter.

**Pilot Study**

A pilot study was conducted in order to validate the instruments and gain some insight into the research design of the dissertation study. The pilot study examined the different effects of two multimedia annotations (text-picture and audio-picture) on L2 vocabulary immediate recall and reading comprehension. Vocabulary Knowledge Scale (VKS) and Word Recognition Test (WRT) were used to measure vocabulary immediate recall. A 10-question multiple-choice reading comprehension test (RC) was used to measure reading comprehension.
Participants

The participants of the pilot study were 20 intermediate level ESL students from an IEP. Five participants were assigned to each of the four treatments. The data were collected during the 2005 fall term.

Procedures

The participants were randomly assigned to one of the four treatments: incidental text-picture, incidental audio-picture, intentional text-picture, and intentional audio-picture. Participants in the intentional learning condition were told in advance of the vocabulary posttests. The researcher explained the project and asked the participants to read the English text of “European Settlers of Australia” on the computer. After reading, the participants were given the posttests. The order of administering the posttests was as follows: Vocabulary Knowledge Scale, Word Recognition Test, and Reading Comprehension Test.

Results

The researcher coded and rated all the three tests; therefore, inter-rater reliability was not obtained. The reliability of the test was measured with Cronbach’s alpha (Cronbach, 1951). For the scores obtained from the three posttests, the reliability was respectively .84 for the Vocabulary Knowledge Scale (VKS), .83 for the Word Recognition Test (WRT), and .51 for the Multiple-choice Reading Comprehension Test (RC).
The descriptive statistical results of the tests can be seen in Table 4-1. Table 4-2 and Table 4-3 present the means and standard deviation for the three tests by annotation type and learning condition respectively. ANOVA results can be seen in Table 4-4.

For the Vocabulary Knowledge Scale (VKS), the audio-picture-intentional group (M=30.5, SD=3.32) obviously outperformed the other three groups (M=24, SD=6.37, M=23.1, SD=8.58, M=25, SD=5.56). If evaluated by annotation type, the audio-picture type (M=27.75, SD=5.20) did better than the text-picture type (M=23.55, SD=7.14); if examined from learning condition, the intentional condition (M=26.8, SD=7.27) performed better than the incidental condition (M=24.5, SD=5.66). However, an ANOVA did not show a statistically significant difference between the two annotation types, \( F (1, 19) = 2.26, p = .1522 > .05 \), or between the two learning conditions, \( F (1, 19) = .68, p = .4224 > .05 \). Nor was there significant interaction between annotation type and learning condition, \( F (1, 19) = 1.31, p = .2688 > .05 \).

Table 4-1

*Descriptive Statistics for Tests of the Pilot Study*

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>VKS M</th>
<th>SD</th>
<th>WRT M</th>
<th>SD</th>
<th>RC M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text-picture/Incidental</td>
<td>5</td>
<td>24</td>
<td>6.37</td>
<td>14.2</td>
<td>1.92</td>
<td>8.2</td>
<td>1.64</td>
</tr>
<tr>
<td>Text-picture/Intentional</td>
<td>5</td>
<td>23.1</td>
<td>8.58</td>
<td>16.6</td>
<td>2.07</td>
<td>7</td>
<td>2.24</td>
</tr>
<tr>
<td>Audio-picture/Incidental</td>
<td>5</td>
<td>25</td>
<td>5.56</td>
<td>17.4</td>
<td>1.52</td>
<td>8.2</td>
<td>1.48</td>
</tr>
<tr>
<td>Audio-picture/Intentional</td>
<td>5</td>
<td>30.5</td>
<td>3.32</td>
<td>17.2</td>
<td>3.03</td>
<td>6.4</td>
<td>.89</td>
</tr>
</tbody>
</table>
Table 4-2

*Means and Standard Deviation for the Tests of the Pilot Study by Annotation Type*

<table>
<thead>
<tr>
<th>Annotation</th>
<th>n</th>
<th>VKS M</th>
<th>SD</th>
<th>WRT M</th>
<th>SD</th>
<th>RC M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text-picture</td>
<td>10</td>
<td>23.55</td>
<td>7.14</td>
<td>15.4</td>
<td>2.27</td>
<td>7.6</td>
<td>1.96</td>
</tr>
<tr>
<td>Audio-picture</td>
<td>10</td>
<td>27.75</td>
<td>5.20</td>
<td>17.4</td>
<td>2.26</td>
<td>7.3</td>
<td>1.49</td>
</tr>
</tbody>
</table>

Table 4-3

*Means and Standard Deviation for the Tests of the Pilot Study by Learning Condition*

<table>
<thead>
<tr>
<th>Learning</th>
<th>n</th>
<th>VKS M</th>
<th>SD</th>
<th>WRT M</th>
<th>SD</th>
<th>RC M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incidental</td>
<td>10</td>
<td>24.5</td>
<td>5.66</td>
<td>15.8</td>
<td>2.35</td>
<td>8.2</td>
<td>1.48</td>
</tr>
<tr>
<td>Intentional</td>
<td>10</td>
<td>26.8</td>
<td>7.27</td>
<td>16.9</td>
<td>2.47</td>
<td>6.7</td>
<td>1.64</td>
</tr>
</tbody>
</table>

Table 4-4

*ANOVA Results of the Pilot Study*

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>VKS F</th>
<th>P</th>
<th>WRT F</th>
<th>P</th>
<th>RC F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annotation</td>
<td>1</td>
<td>2.26</td>
<td>.1522</td>
<td>.01</td>
<td>.9200</td>
<td>.17</td>
<td>.6871</td>
</tr>
<tr>
<td>Learning</td>
<td>1</td>
<td>.68</td>
<td>.4224</td>
<td>.26</td>
<td>.6168</td>
<td>4.21</td>
<td>.0571</td>
</tr>
<tr>
<td>Annotation * Learning</td>
<td>1</td>
<td>1.31</td>
<td>.2688</td>
<td>.09</td>
<td>.7634</td>
<td>.17</td>
<td>.6871</td>
</tr>
<tr>
<td>Error</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
For the Word Recognition Test (WRT), as a whole, the audio-picture-intentional (M=17.4) outscored the other three groups (M=17.2 for audio-picture-intentional group, M=16.6 for text-picture-intentional and M=14.2 for text-picture-incidental). If judged by annotation type, the audio-picture annotation (M=17.4) outperformed the text-picture annotation (M=15.4); if judged by learning condition, the intentional condition (M=16.9) did better than the incidental condition (M=15.8). However, an ANOVA showed the differences were not statistically significant. There was no interaction between annotation type and learning condition.

For the reading comprehension test (RC), as a whole, both annotation types in incidental condition had the highest mean (M=8.2), followed by text-picture annotation group (M=7), and audio-picture-intentional group had the lowest mean (M=6.4). From the perspective of annotation type, the text-picture annotation (M=7.6) scored higher than the audio-picture annotation (M=7.3) in reading comprehension; from the perspective of learning condition, the incidental condition (M=8.2) led to better results in reading comprehension than the intentional condition (M=6.7). However, an ANOVA did not show significant effect for either annotation type or learning condition, although the difference between the two learning conditions was close to being significant (F=4.21, p=.0571). No interaction was found between annotation type and learning condition (F=.17, p=.6871 > .05). The results of the pilot study are summarized in Figure 4-1.

Discussion

The purpose of the pilot study was to validate the scores from the instruments and investigate the applicability of the research design to be used for this dissertation study.
The result of the pilot study showed the three instruments to be reliable. From the ANOVA analysis, there were no statistically significant effects for annotation type or learning condition, nor was there an interaction between annotation type and learning condition. However, the large mean differences along with sizable standard deviations from the pilot study could explain the lack of statistically significant results. This indicated the necessity for further study with a larger sample size.

![Results of Pilot Study](image_url)

*Figure 4-1. Results of Pilot Study*

The following part of this chapter will present the results of the actual study. First the information on annotation look-up behavior is presented, followed by the results of the four posttests will be presented, and finally the results by research questions.
Annotation Look-up Behavior Results

The participants’ annotation look-up behaviors were tracked and recorded when they clicked the hyperlinked words for annotations. One look-up was counted every time a target word was clicked. Seven participants restrained from checking the available annotation and did not look up any of target words and were, therefore, excluded from the data analysis. Among the 7 participants, 2 were from the text-picture annotation group and the other 5 were from the audio-picture group.

Table 4-5 and Table 4-6 show the descriptive statistics for look-up behaviors. As shown by Table 4-5, the mean score for text-picture/incidental was the highest, followed in order by audio-picture/intentional, audio-picture/incidental, and text-picture/intentional. However, as displayed by Table 4-6, participants in the two annotation types looked up on average the same amount of words (24.05 vs. 23.62), and participants in the two learning conditions looked up similar amount of words as well (24.55 vs. 23.18). A two-way (Annotation x Learning) analysis of variance (ANOVA) was performed on the look-up behavior data.

Table 4-5

Descriptive Statistics for Look-up Behavior

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>M</th>
<th>SD</th>
<th>Sk</th>
<th>Kr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text-picture/Incidental</td>
<td>20</td>
<td>26.65</td>
<td>8.76</td>
<td>.23</td>
<td>-1.43</td>
</tr>
<tr>
<td>Text-picture/Intentional</td>
<td>19</td>
<td>21.31</td>
<td>4.99</td>
<td>.47</td>
<td>.01</td>
</tr>
<tr>
<td>Audio-picture/Incidental</td>
<td>20</td>
<td>22.45</td>
<td>6.89</td>
<td>1.25</td>
<td>2.10</td>
</tr>
<tr>
<td>Audio-picture/Intentional</td>
<td>19</td>
<td>24.84</td>
<td>6.88</td>
<td>1.02</td>
<td>.45</td>
</tr>
</tbody>
</table>

Sk = skewness, Kr = kurtosis
Table 4-6

*Means and Standard Deviations for Look-up Behavior*

<table>
<thead>
<tr>
<th>Annotation</th>
<th>n</th>
<th>M</th>
<th>SD</th>
<th>Learning</th>
<th>n</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text-picture</td>
<td>39</td>
<td>24.05</td>
<td>7.58</td>
<td>Incidental</td>
<td>40</td>
<td>24.55</td>
<td>8.07</td>
</tr>
<tr>
<td>Audio-picture</td>
<td>39</td>
<td>23.62</td>
<td>6.90</td>
<td>Intentional</td>
<td>38</td>
<td>23.18</td>
<td>6.19</td>
</tr>
</tbody>
</table>

As indicated by Table 4-7, the differences between annotation types and between learning conditions were not significant, while the interaction between annotation type and learning condition was significant (F = 5.88, p = .0178 < .05).

Table 4-7

*ANOVA for Look-up Behavior*

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Type III SS</th>
<th>Square</th>
<th>F Value</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annotation</td>
<td>1</td>
<td>2.211</td>
<td>2.211</td>
<td>.04</td>
<td>.8332</td>
</tr>
<tr>
<td>Learning</td>
<td>1</td>
<td>42.17</td>
<td>42.17</td>
<td>.85</td>
<td>.3590</td>
</tr>
<tr>
<td>Annotation * Learning</td>
<td>1</td>
<td>290.83</td>
<td>290.83</td>
<td>5.88</td>
<td>.0178**</td>
</tr>
<tr>
<td>Error</td>
<td>74</td>
<td>3662.13</td>
<td>49.49</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

** p < .05
Since there was interaction effect between annotation type and learning condition, a correlation procedure was carried out to check if look-up behaviors were related to the posttest measures. As shown in Table 4-8, look-up behavior was not statistically significantly correlated with any of the posttest measures for the dependent variable. Therefore, further data analysis was conducted without controlling for look-up behavior.

### Posttest Results

Four instruments were used in the posttest: Vocabulary Knowledge Scale (VKS), Word Recognition Test (WRT), Multiple-choice Reading Comprehension Test (RC) and Written Recall. All the four tests were immediate posttests. They were administered right after the participants finished reading the English text on the computer. The tests were given in the following order: VKS, WRT, RC and Written Recall. The results of the Vocabulary Knowledge Scale are discussed first, followed by those from the Word Recognition Test, then the Reading Comprehension Test, and finally the Written Recall. A two-way (Annotation x Learning) analysis of variance (ANOVA) was performed on each immediate posttest. The alpha level was set at .05 for all the analyses.
Vocabulary Knowledge Scale Scores

The results of the Vocabulary Knowledge Scale were based on the participants’ self-rated knowledge of the 20 target words. For each target word, the participants were asked to choose a description of their knowledge of the word. Two points were given for a correct written meaning, 1.5 points were given for a recognized word and partially correct meaning, one point was given for a recognized word without written meaning, and zero point was given for an unrecognized word. The total possible score was 40 points. A two-way ANOVA (Annotation x Learning) was used to examine the scores on the Vocabulary Knowledge Scale. The reliability of the instrument, as measured by internal consistency, was .86 (Cronbach coefficient alpha). The inter-rater reliability was .90.

Table 4-9 shows descriptive statistics for the Vocabulary Knowledge Scale for various combinations of annotation type and learning condition. Table 4-10 presents the means and the standard deviations of the Vocabulary Knowledge Scale scores by the annotation types and learning conditions respectively. The scores were normally distributed for each of the treatment groups. Figure 4-2 illustrates that the audio-picture annotation participants did better on the Vocabulary Knowledge Scale than the text-picture annotation participants. As shown by Figure 4-3, participants in the intentional learning condition performed better than those in the incidental learning conditions. An ANOVA was conducted to establish whether or not the differences were significant.

As seen from Table 4-11 and Figure 4-4, the ANOVA results show that there is no significant interaction between annotation type and learning condition on the
Vocabulary Knowledge Scale. However, there were significant differences between the annotation types as well as between the learning conditions.

Participants in the audio-picture annotation type performed significantly better than those in the text-picture annotation type, $F (1, 74) = 8.40, p = .0049 < .05$. A significant difference was also found for the learning condition factor, with the intentional learning condition outperforming the incidental learning condition, $F (1, 74) = 7.15, p = .0092 < .05$.

Table 4-9

Descriptive Statistics for Vocabulary Knowledge Scale by Annotation and Learning

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>M</th>
<th>SD</th>
<th>Sk</th>
<th>Kr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text-picture/Incidental</td>
<td>20</td>
<td>19.33</td>
<td>7.73</td>
<td>.39</td>
<td>-.32</td>
</tr>
<tr>
<td>Text-picture/Intentional</td>
<td>19</td>
<td>24.13</td>
<td>7.27</td>
<td>.45</td>
<td>-.10</td>
</tr>
<tr>
<td>Audio-picture/Incidental</td>
<td>20</td>
<td>24.48</td>
<td>6.56</td>
<td>.04</td>
<td>-1.01</td>
</tr>
<tr>
<td>Audio-picture/Intentional</td>
<td>19</td>
<td>27.87</td>
<td>5.20</td>
<td>-.30</td>
<td>-.81</td>
</tr>
</tbody>
</table>

Sk = skewness, Kr = kurtosis

Table 4-10

Means and Standard Deviations for Vocabulary Knowledge Scale

<table>
<thead>
<tr>
<th>Annotation</th>
<th>n</th>
<th>M</th>
<th>SD</th>
<th>Learning</th>
<th>n</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text-picture</td>
<td>39</td>
<td>21.67</td>
<td>7.80</td>
<td>Incidental</td>
<td>40</td>
<td>21.9</td>
<td>7.54</td>
</tr>
<tr>
<td>Audio-picture</td>
<td>39</td>
<td>26.13</td>
<td>6.11</td>
<td>Intentional</td>
<td>38</td>
<td>26</td>
<td>6.52</td>
</tr>
</tbody>
</table>
### Table 4-11

**ANOVA for Vocabulary Knowledge Scale**

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Type III SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annotation</td>
<td>1</td>
<td>384.75</td>
<td>384.75</td>
<td>8.40</td>
<td>.0049**</td>
</tr>
<tr>
<td>Learning</td>
<td>1</td>
<td>327.58</td>
<td>327.58</td>
<td>7.15</td>
<td>.0092**</td>
</tr>
<tr>
<td>Annotation * Learning</td>
<td>1</td>
<td>9.73</td>
<td>9.73</td>
<td>.21</td>
<td>.6463</td>
</tr>
<tr>
<td>Error</td>
<td>74</td>
<td>3390.22</td>
<td>45.81</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

** p < .05

---

**Figure 4-2.** Means for the Vocabulary Knowledge Scale by Annotation Types
Word Recognition Test Scores

The results of the Word Recognition Test were based on the participants’ identification of the correct meaning (one out of 4 choices) with the 20 target words. One
point was given for a correct answer, and zero point was given to an incorrect answer. The total possible score was 20 points. A two-way ANOVA (Annotation x Learning) was used to analyze the scores on the Word Recognition Test. The reliability of the instrument, measured by internal consistency (Cronbach coefficient alpha), was .76.

Table 4-12 presents descriptive statistics for the Word Recognition Test in different combinations of annotation type and learning condition. The means and the standard deviations of the Word Recognition Test scores by the annotation types and learning conditions respectively are provided in Table 4-13. The scores were normally distributed for each of the treatment groups.

Figure 4-5 illustrates that participants in the audio-picture annotation did better on the Word Recognition Test than participants in the text-picture annotation. As shown by Figure 4-6, participants in the intentional learning condition performed better than those in the incidental learning conditions. An ANOVA was conducted to establish whether or not the differences were significant.

Table 4-12

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>M</th>
<th>SD</th>
<th>Sk</th>
<th>Kr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text-picture/Incidental</td>
<td>20</td>
<td>12.75</td>
<td>2.88</td>
<td>-.98</td>
<td>1.50</td>
</tr>
<tr>
<td>Text-picture/Intentional</td>
<td>19</td>
<td>16.21</td>
<td>2.35</td>
<td>-.66</td>
<td>-.30</td>
</tr>
<tr>
<td>Audio-picture/Incidental</td>
<td>20</td>
<td>15.75</td>
<td>3.02</td>
<td>-.74</td>
<td>-.50</td>
</tr>
<tr>
<td>Audio-picture/Intentional</td>
<td>19</td>
<td>17.63</td>
<td>2.43</td>
<td>-1.22</td>
<td>.45</td>
</tr>
</tbody>
</table>

Sk = skewness, Kr = kurtosis
Table 4-13

*Means and Standard Deviations for Word Recognition Test*

<table>
<thead>
<tr>
<th>Annotation</th>
<th>n</th>
<th>M</th>
<th>SD</th>
<th>Learning</th>
<th>n</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text-picture</td>
<td>39</td>
<td>14.44</td>
<td>3.14</td>
<td>Incidental</td>
<td>40</td>
<td>14.25</td>
<td>3.29</td>
</tr>
<tr>
<td>Audio-picture</td>
<td>39</td>
<td>16.67</td>
<td>2.88</td>
<td>Intentional</td>
<td>38</td>
<td>16.92</td>
<td>2.46</td>
</tr>
</tbody>
</table>

As seen from Table 4-14 and Figure 4-7, the ANOVA results show that there is no significant interaction between annotation type and learning condition. However, participants in the audio-picture annotation type performed significantly better than those in the text-picture annotation type, $F(1, 74) = 13.12, p = .0005 < .05$. In addition, a significant difference was also found between the two learning conditions, with those in the intentional learning condition outperforming those in the incidental learning condition, $F(1, 74) = 19.16, p < .0001 < .05$.

*Figure 4-5. Means for the Word Recognition Test by Annotation Types*
Figure 4-6. Means for the Word Recognition Test by Learning Conditions

Figure 4-7. WRT scores by Annotation Type and Learning Condition
Table 4-14

ANOVA for Word Recognition Test

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Type III SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annotation</td>
<td>1</td>
<td>95.22</td>
<td>95.22</td>
<td>13.12</td>
<td>.0005**</td>
</tr>
<tr>
<td>Learning</td>
<td>1</td>
<td>139.03</td>
<td>139.03</td>
<td>19.16</td>
<td>&lt; .0001**</td>
</tr>
<tr>
<td>Annotation * Learning</td>
<td>1</td>
<td>12.15</td>
<td>12.15</td>
<td>1.67</td>
<td>.1998</td>
</tr>
<tr>
<td>Error</td>
<td>74</td>
<td>537.08</td>
<td>7.26</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

** p < .05

Multiple-Choice Reading Comprehension Test Scores

The results of the Reading Comprehension Test were obtained from the participants’ scores on the multiple-choice comprehension questions. For each of the 10 comprehension questions, participants were asked to choose a correct answer. One point was given for a correct answer, and zero point was given to an incorrect answer. The total possible score was 10 points. A two-way ANOVA (Annotation x Learning) was used to analyze the scores on the Reading Comprehension Test. The reliability of the instrument was .56 as measured by internal consistency (Cronbach coefficient alpha).

Descriptive statistics of the Reading Comprehension Test were presented in Table 4-15. Table 4-16 shows the means and the standard deviations for the Reading Comprehension Test score respectively by annotation type and learning condition. The scores were normally distributed for each of the treatment groups.
Table 4-15

*Descriptive Statistics for Reading Comprehension Test by Annotation and Learning*

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>M</th>
<th>SD</th>
<th>Sk</th>
<th>Kr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text-picture/Incidental</td>
<td>20</td>
<td>6.25</td>
<td>2.12</td>
<td>-.22</td>
<td>-.49</td>
</tr>
<tr>
<td>Text-picture/Intentional</td>
<td>19</td>
<td>7.32</td>
<td>2.26</td>
<td>-.60</td>
<td>-.87</td>
</tr>
<tr>
<td>Audio-picture/Incidental</td>
<td>20</td>
<td>6.65</td>
<td>2.03</td>
<td>-.51</td>
<td>-.18</td>
</tr>
<tr>
<td>Audio-picture/Intentional</td>
<td>19</td>
<td>6.32</td>
<td>1.45</td>
<td>-.26</td>
<td>.12</td>
</tr>
</tbody>
</table>

Sk = skewness, Kr = kurtosis

Table 4-16

*Means and Standard Deviations for Reading Comprehension Test*

<table>
<thead>
<tr>
<th>Annotation</th>
<th>n</th>
<th>M</th>
<th>SD</th>
<th>Learning</th>
<th>n</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text-picture</td>
<td>39</td>
<td>6.77</td>
<td>2.23</td>
<td>Incidental</td>
<td>40</td>
<td>6.45</td>
<td>2.06</td>
</tr>
<tr>
<td>Audio-picture</td>
<td>39</td>
<td>6.49</td>
<td>1.76</td>
<td>Intentional</td>
<td>38</td>
<td>6.82</td>
<td>1.94</td>
</tr>
</tbody>
</table>

Figure 4-8 and Figure 4-9 indicate that in terms of Reading Comprehension Test, the text-picture annotation participants outscored the audio-picture annotation participants, and the intentional learning condition participants outperformed the incidental learning condition participants. An ANOVA was carried out to verify the significance of the differences.
As presented in Table 4-17, the Reading Comprehension Test results revealed no main effects for annotation type, $F(1, 74) = .44, p = .5089 > .05$, or for learning condition, $F(1, 74) = .66, p = .4209 > .05$. Nor was the interaction between annotation type and learning condition significant, $F(1, 74) = 2.40, p = .1257 > .05$. Figure 4-10 did
show some interaction graphically, but the differences between sample means are not far enough apart to conclude the population means are different, therefore, the effect seen in the sample is not pronounced enough to conclude there is an interaction effect in the population.

Table 4-17

ANOVA for Reading Comprehension Test

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Type III SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annotation</td>
<td>1</td>
<td>1.75</td>
<td>1.75</td>
<td>.44</td>
<td>.5089</td>
</tr>
<tr>
<td>Learning</td>
<td>1</td>
<td>2.61</td>
<td>2.61</td>
<td>.66</td>
<td>.4209</td>
</tr>
<tr>
<td>Annotation * Learning</td>
<td>1</td>
<td>9.55</td>
<td>9.55</td>
<td>2.40</td>
<td>.1257</td>
</tr>
<tr>
<td>Error</td>
<td>74</td>
<td>294.51</td>
<td>3.98</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

** p < .05

Figure 4-10. RC Scores by Annotation Type and Learning Condition
Written Recall Scores

In the Written Recall, each participant was asked to recall as much information as they could remember from the reading passage. One point was given for each idea unit recalled. There were a total of 52 idea units in the reading text. A two-way ANOVA (Annotation x Learning) was used to analyze the scores on the Written Recall. The reliability of the instrument was .64 as measured by internal consistency (Cronbach coefficient alpha). The inter-rater reliability was .92.

Descriptive statistics of the Reading Comprehension Test were presented in Table 4-18. Table 4-19 shows the means and the standard deviations for the Reading Comprehension Test score respectively by annotation type and learning condition. The scores were normally distributed for each of the treatment groups.

Figure 4-11 and Figure 4-12 indicate that in terms of Written Recall, the text-picture annotation participants slightly outscored the audio-picture annotation participants, and the incidental learning condition participants outperformed the intentional learning condition participants. An ANOVA was carried out to verify the significance of the differences.

As presented in Table 4-20 and Figure 4-13, the Written Recall results revealed an interaction between annotation type and learning condition, F (1, 74) = 8.49, p = .0047 < .05. The main effect of learning condition was also significant, F (1, 74) = 8.97, p = .0037 < .05. However, the main effect of annotation type was not significant, F (1, 74) = .90, p = .3464 > .05.
Table 4-18

*Descriptive Statistics for Written Recall by Annotation and Learning*

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>M</th>
<th>SD</th>
<th>Sk</th>
<th>Kr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text-picture/Incidental</td>
<td>20</td>
<td>15.0</td>
<td>6.60</td>
<td>.68</td>
<td>1.11</td>
</tr>
<tr>
<td>Text-picture/Intentional</td>
<td>19</td>
<td>14.89</td>
<td>7.38</td>
<td>.37</td>
<td>-1.17</td>
</tr>
<tr>
<td>Audio-picture/Incidental</td>
<td>20</td>
<td>17.6</td>
<td>4.51</td>
<td>.35</td>
<td>-.55</td>
</tr>
<tr>
<td>Audio-picture/Intentional</td>
<td>19</td>
<td>9.79</td>
<td>4.25</td>
<td>.10</td>
<td>-.65</td>
</tr>
</tbody>
</table>

Sk = skewness, Kr = kurtosis

Table 4-19

*Means and Standard Deviations for Written Recall*

<table>
<thead>
<tr>
<th>Annotation</th>
<th>n</th>
<th>M</th>
<th>SD</th>
<th>Learning</th>
<th>n</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text-picture</td>
<td>39</td>
<td>14.95</td>
<td>6.90</td>
<td>Incidental</td>
<td>40</td>
<td>16.3</td>
<td>5.73</td>
</tr>
<tr>
<td>Audio-picture</td>
<td>39</td>
<td>13.79</td>
<td>5.86</td>
<td>Intentional</td>
<td>38</td>
<td>12.34</td>
<td>6.48</td>
</tr>
</tbody>
</table>
The significant effect of learning condition means that participants from the incidental learning condition did significantly better than participants from the intentional learning condition in recalling the reading text. The significant interaction effect means that effect of learning condition on written recall depends on annotation type. As shown
by Figure 4-13, the incidental learning condition led to better written recall only in the audio-picture annotation group. In the incidental learning condition, participants in the audio-picture annotation did better than participants in the text-picture annotation; while in the intentional learning condition, participants in the text-picture annotation did better than participants in the audio-picture annotation. Independent sample T-tests were conducted to further examine whether the differences were statistically significant ($\alpha=.05$). In incidental learning condition, the value of t ($t=-1.45$) failed to confirm the difference between the two annotations was significant. However, in intentional learning condition, the t value ($t=2.61$) showed that the difference between the two annotations was significant. This means that although the main effect of annotation type is not statistically significant, the text-picture annotation participants did significantly better in written recall than the audio-picture annotation participants in intentional learning condition.

Table 4-20

**ANOVA for Written Recall**

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Type III SS</th>
<th>Square</th>
<th>F Value</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annotation</td>
<td>1</td>
<td>30.58</td>
<td>30.58</td>
<td>.90</td>
<td>.3464</td>
</tr>
<tr>
<td>Learning</td>
<td>1</td>
<td>305.27</td>
<td>305.27</td>
<td>8.97</td>
<td>.0037 **</td>
</tr>
<tr>
<td>Annotation * Learning</td>
<td>1</td>
<td>289.24</td>
<td>289.24</td>
<td>8.49</td>
<td>.0047 **</td>
</tr>
<tr>
<td>Error</td>
<td>74</td>
<td>2519.75</td>
<td>34.05</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

** p < .05
Results by Research Questions

This section of the chapter is organized according to the research questions of this dissertation study. Each research question will be stated and answered based on the results presented above.

**Question 1 – Effects of Annotation Type on Vocabulary Immediate Recall**

The first research question investigated the different effects of text-picture and audio-picture annotations on L2 vocabulary immediate recall. The results of the two vocabulary posttests (Vocabulary Knowledge Scale and Word Recognition Test) were compiled to answer this question. The results are displayed in Table 4-21 and Figure 4-14.

In terms of vocabulary immediate recall, Table 4-21 shows that the audio-picture annotation type led to better results (VKS M=27.75, WRT M=26.13) than the text-picture
annotation type (VKS M=23.55  WRT M=21.67). As indicated by Figure 4-14, the audio-picture annotation type outperformed the text-picture annotation type, regardless of the type of test (Vocabulary Knowledge Scale and Word Recognition Test).

For the Vocabulary Knowledge Scale, there was a statistically significant difference between the two annotation types, $F (1, 74) = 8.40, p = .0049 < .05$. In other words, participants in the audio-picture annotation type did significantly better than participants in the text-picture annotation type on the Vocabulary Knowledge Scale. A statistically significant effect of annotation was also found in the Word Recognition Test, $F (1, 74) = 7.15, p = .0092 < .05$. This confirms that the audio-picture annotation type resulted in better performance in Word Recognition Test than the text-picture annotation type.

Table 4-21

*Means and Standard Deviations for Vocabulary Posttests by Annotation*

<table>
<thead>
<tr>
<th>Annotation</th>
<th>n</th>
<th>VKS</th>
<th>SD</th>
<th>WRT</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text-picture</td>
<td>39</td>
<td>23.55</td>
<td>7.14</td>
<td>21.67</td>
<td>7.80</td>
</tr>
<tr>
<td>Audio-picture</td>
<td>39</td>
<td>27.75</td>
<td>5.20</td>
<td>26.13</td>
<td>6.11</td>
</tr>
</tbody>
</table>
Table 4-22

*Means and Standard Deviations for Reading Comprehension by Annotation*

<table>
<thead>
<tr>
<th>Annotation</th>
<th>n</th>
<th>M</th>
<th>SD</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text-picture</td>
<td>39</td>
<td>6.77</td>
<td>2.23</td>
<td>14.95</td>
<td>6.90</td>
</tr>
<tr>
<td>Audio-picture</td>
<td>39</td>
<td>6.49</td>
<td>1.76</td>
<td>13.79</td>
<td>5.86</td>
</tr>
</tbody>
</table>

*Figure 4-14. Means for Vocabulary Posttests by Annotation Type*

In summary, the results indicate that audio-picture annotation resulted in significantly better performance in vocabulary immediate recall than the text-picture annotation. Thus, hypothesis 1, which states that students in the audio-picture annotation group will perform significantly better than students in the text-picture annotation group was supported.
Question 2 – Effects of Annotation Type on Reading Comprehension

The second research question was focused on the relationship between annotation type and reading comprehension. The results of the two comprehension tests (Reading Comprehension Test and Written Recall) in terms of annotation were presented in Table 4-22 and Figure 4-16.

For reading comprehension, Table 4-22 shows that the results for the Multiple-choice Reading Comprehension Test between the two annotation types were very close (M = 6.77 and M = 6.49 respectively). The difference for the Written Recall between the two annotation types was also very small (M = 14.95 and M = 13.79 respectively). As indicated by Figure 4-15, the text-picture annotation participants slightly outscored the audio-picture annotation participants in terms of reading comprehension.

![Means for Reading Comprehension by Annotation Type](Image)

*Figure 4-15. Means for Reading Comprehension by Annotation Type*
An ANOVA analysis did not show any significant effect for annotation type on reading comprehension in either of the two reading comprehension tests: Multiple-choice Reading Comprehension Test ($F = .44, p = .5089 > .05$) and Written Recall ($F = .90, p = .3464 > .05$). Because the interaction between annotation type and learning condition was statistically significant for Written Recall, independent sample T-tests shows that participants in the text-picture annotation group did significantly better than participants in the audio-picture annotation group in intentional learning condition. Therefore, hypothesis 2, which states that students in the audio-picture annotation group will perform statistically better than participants in the text-picture annotation group on the reading comprehension posttests, was not supported.

**Question 3 – Effects of Learning Condition on Vocabulary Immediate Recall**

The third research question examined if a particular learning condition might affect vocabulary immediate recall when participants were reading an English text for comprehension. The results of the two vocabulary posttests (Vocabulary Knowledge Scale and Word Recognition Test) in terms of learning condition were complied in Table 4-23 and Figure 4-16.

Regarding vocabulary immediate recall, Table 4-23 shows that the intentional learning condition led to better results (VKS $M=26$, WRT $M=16.92$) than the Incidental learning condition (VKS $M=21.9$  WRT $M=14.25$). As indicated by Figure 4-16, the intentional learning condition participants outscored the incidental learning condition participants consistently across the two vocabulary posttests (Vocabulary Knowledge Scale and Word Recognition Test).
For the Vocabulary Knowledge Scale, there was a statistically significant difference between the two learning conditions, $F(1, 74) = 13.12, p = .0009 < .05$. In other words, participants in the intentional learning condition did significantly better than participants in the incidental learning condition in the Vocabulary Knowledge Scale test. For the Word Recognition Test, a statistically significant effect of learning was found, $F(1, 74) = 19.16, p < .0001 < .05$. This means that participants in the intentional learning condition retained significantly more target words than participants in the incidental learning condition.

For immediate vocabulary recall, to summarize, the intentional learning condition resulted in significantly better performance than the incidental learning condition. Thus, hypothesis 3, which states that students in the intentional learning condition will perform significantly better than students in the incidental learning condition on the immediate vocabulary posttests, was supported.

Table 4-23

*Means and Standard Deviations for Vocabulary Posttests by Learning*

<table>
<thead>
<tr>
<th></th>
<th>VKS M</th>
<th>SD</th>
<th>WRT M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning</td>
<td>n</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incidental</td>
<td>40</td>
<td>21.9</td>
<td>7.54</td>
<td>14.25</td>
</tr>
<tr>
<td>Intentional</td>
<td>38</td>
<td>26</td>
<td>6.52</td>
<td>16.92</td>
</tr>
</tbody>
</table>

120
Question 4 – Effects of Learning Condition on Reading Comprehension

The focus of research question four was whether the two learning conditions had different effects on the participants’ reading comprehension. To answer this question, the results of the two reading comprehension measures (Reading Comprehension Test and Written Recall) were presented in Table 4-24 and Figure 4-17.

As indicated by Table 4-24 and Figure 4-17, different results were found for the two reading comprehension measures. In the Multiple-choice Reading Comprehension Test, participants in the two learning conditions had very close scores (incidental M = 6.82, intentional M = 6.45). In the Written Recall, however, participants in the incidental learning condition did better (M = 16.3) than participants in the intentional learning condition (M = 12.34).
For the Multiple-choice Reading Comprehension Test, the effect of learning condition was not found to be significant ($F = .66$, $p = .4209 > .05$). This means that there was no significant difference between the two learning conditions on the multiple-choice reading comprehension test. For the Written Recall, however, there was a significant effect for learning condition ($F = 8.97$, $p = .0037 < .05$). This indicates that participants in the incidental learning condition did significantly better than participants in the intentional learning condition in recalling the reading text.

Therefore, the incidental learning condition resulted in significantly better reading comprehension than the intentional learning condition in the Written Recall measure, but not in the multiple-choice Reading Comprehension Test. Thus, hypothesis 4, which states that students in the incidental learning condition will perform significantly better than participants in the intentional learning condition on the reading comprehension posttest, was partially supported.

Table 4-24

*Means and Standard Deviations for Reading Comprehension by Learning*

<table>
<thead>
<tr>
<th>Learning</th>
<th>n</th>
<th>M</th>
<th>SD</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incidental</td>
<td>40</td>
<td>6.45</td>
<td>2.06</td>
<td>16.3</td>
<td>5.73</td>
</tr>
<tr>
<td>Intentional</td>
<td>38</td>
<td>6.82</td>
<td>1.94</td>
<td>12.34</td>
<td>6.48</td>
</tr>
</tbody>
</table>
The fifth research question examined the degree of interaction between annotation type and learning condition on tasks involving vocabulary immediate recall. In order to answer this question, all the interaction effects between annotation type and learning condition for vocabulary measurements were revisited. Table 4-25 lists the interaction between the two factors for the Vocabulary Knowledge Scale and the Word Recognition Test.

As indicated by Table 4-25, there was no interaction between annotation type and learning condition across the two tests. This helps to explain that the learning condition difference did not have any effect on the annotation type. In other words, regardless of the learning condition, the results found in research question one remained the same, namely that the audio-picture annotation group significantly performed better than the
text-picture annotation group on vocabulary immediate recall. Thus, Hypothesis 5, which states that there will be an interaction between learning condition and annotation type on L2 vocabulary immediate recall, was not supported.

Table 4-25

*A Summary of Interaction between Annotation and Learning on Vocabulary Immediate Recall*

<table>
<thead>
<tr>
<th>Source (Annotation x Learning)</th>
<th>df, df</th>
<th>F Value</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>VKS</td>
<td>1, 74</td>
<td>.21</td>
<td>.6463</td>
</tr>
<tr>
<td>WRT</td>
<td>1, 74</td>
<td>1.67</td>
<td>.1998</td>
</tr>
</tbody>
</table>

*Question 6 – Interaction between Annotation and Learning on Reading Comprehension*

Research question six investigated whether there is interaction between annotation type and learning condition on participants’ reading comprehension. In order to answer this question, all the interaction effects between annotation type and learning condition for reading comprehension measures were revisited and shown in Table 4-26.

As indicated by Table 4-26, there was no interaction between annotation type and learning condition on reading comprehension scores in the Multiple-choice Reading Comprehension Test (F = 2.40, p = .1257 > .05). However, in the Written Recall, there is significant interaction between learning condition and annotation type. This can be interpreted that the difference between the two annotation types in written recall depends on the learning conditions (F = 8.49, p = .0047 < .05). In other words, in the incidental learning condition, the difference between text-picture annotation and audio-picture
annotation was not significant; in the intentional learning condition, participants in text-picture did significantly better than those in audio-picture on Written Recall. Therefore, hypothesis 6, which states that there will be an interaction between learning condition and annotation type on L2 reading comprehension, was partially supported. Whether there is a significant interaction depends on the task type.

Table 4-26

A Summary of Interaction between Annotation and Learning on Reading Comprehension

<table>
<thead>
<tr>
<th>Source (Annotation x Learning)</th>
<th>df, df</th>
<th>F Value</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>RC</td>
<td>1, 74</td>
<td>2.40</td>
<td>.1257</td>
</tr>
<tr>
<td>Written Recall</td>
<td>1, 74</td>
<td>8.49</td>
<td>.0047 **</td>
</tr>
</tbody>
</table>

Summary of Findings

The overarching question addressed in this study concerns how different dual annotations affect L2 vocabulary learning and reading comprehension in both incidental and intentional settings. Text-picture annotation and audio-picture annotation were the two dual annotations under investigation.

Two vocabulary posttests were used to measure the participants’ vocabulary immediate recall: Vocabulary Knowledge Scale and Word Recognition Test. Two comprehension posttests were used to measure the participants’ reading comprehension: Reading Comprehension Test (Multiple-choice) and Written Recall.

The following presents the findings for each of the research questions:
Question 1: What are the different effects of text-picture and audio-picture annotations in facilitating L2 vocabulary immediate recall?

Findings: The audio-picture annotation was the more effective annotation type in facilitating L2 vocabulary immediate recall.

Question 2: What are the different effects of text-picture and text-audio annotations in facilitating L2 reading comprehension?

Findings: Text-picture annotation and audio-picture annotation did not have different effects in facilitating L2 reading comprehension. However in intentional learning condition, the text-picture annotation did lead to significantly better written recall than the audio-picture annotation.

Question 3: Which learning condition (intentional or incidental) results in better L2 vocabulary immediate recall?

Findings: The intentional learning condition resulted in better L2 vocabulary immediate recall.

Question 4: Is there a significant difference in ESL students’ reading comprehension under the two learning conditions (intentional or incidental)?

Findings: the Incidental learning condition resulted in significantly better L2 reading comprehension than the intentional learning condition in the Written Recall measure, but not in the multiple-choice Reading Comprehension Test.

Question 5: Is there an interaction between learning condition and annotation type on immediate L2 vocabulary immediate recall?
Findings: There was no interaction between learning condition and annotation type on L2 vocabulary immediate recall.

Question 6: Is there an interaction between learning condition and annotation type on L2 reading comprehension?

Findings: There was no interaction between learning condition and annotation type on multiple-choice Reading Comprehension test, but the interaction between learning condition and annotation type was significant on Written Recall.

Table 4-27
*A Summary of the Findings in the Study*

<table>
<thead>
<tr>
<th>Posttests</th>
<th>VKS</th>
<th>WRT</th>
<th>RC</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annotation</td>
<td>**</td>
<td>**</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>Audio &gt; Text</td>
<td>Audio &gt; Text</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learning</td>
<td>**</td>
<td>**</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Intentional &gt; Incidental</td>
<td>Intentional &gt; Incidental</td>
<td>ns</td>
<td>Incidental &gt; Intentional</td>
<td></td>
</tr>
<tr>
<td>Interactions</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td>**</td>
</tr>
</tbody>
</table>

(Annotation x Learning)

** = significant (p < .05), ns = not significant, Audio = audio-text annotation, Text = text-picture annotation, Intentional = intentional learning condition, Incidental = incidental learning condition.
The findings for the 6 research questions are summarized in Table 4-27. The next chapter will present discussion on the results, offer theoretical and pedagogical implications, make recommendations for future research, and provide final conclusions.
Chapter V Discussion

This dissertation examined the different effects of text-picture annotation and audio-picture annotation on L2 vocabulary immediate recall and L2 reading comprehension under incidental and intentional learning conditions. After an introduction to the problem in Chapter I, Chapter II reviewed in details the related literature on annotation studies, Chapter III described the method of data collection and analysis, and Chapter IV presented the results of the study. This final chapter will present the interpretations of the results addressing each research question, discuss both theoretical and pedagogical implications, make recommendations for future research, and provide final conclusions.

Interpretations of the Results

Based on the results presented in Chapter IV, the interpretations will address each of the research questions, make references to the literature, and explain possible reasons for the obtained results.

Effect of Annotation Type on L2 Vocabulary Immediate Recall

The first research question addressed the effectiveness of multimedia annotations. On average, participants retained 59.8% of the 20 target words on the Vocabulary Knowledge Scale and 77.7% of the 20 target words on the Word Recognition Test. It was expected that the Vocabulary Knowledge Scale, as a definition-supply test, would
have lower scores than the Word Recognition Test. The average retention rates were comparable to previous multimedia annotation studies (e.g. Al-Seghayer, 2001; Chun & Plass, 1996a; Yoshii, 2000). The results confirmed the effectiveness of multimedia annotation in facilitating second language vocabulary learning. A further explanation is the dual-coding effect (Paivio, 1971, 1986, 1990) that words annotated with both verbal (text or audio) and visual (picture) modes of information lead to effective vocabulary retention.

However, the retention rate of 59.8% (54.8% if only counting incidental vocabulary retention) on the Vocabulary Knowledge Scale was much higher than the results reported in definition-supply tests from previous studies: 25.9% in Chun and Plass (1996a) and 21.4% in Yoshii (2000). This higher rate could be explained by the nature of the task used for this study. The Vocabulary Knowledge Scale combines recognition and production in that points were given to the participants if the target words were recognized in the self-rating scale.

This study was designed to compare the effectiveness of text-picture annotation with audio-picture annotation on L2 vocabulary immediate recall. As shown by Table 4-14 and Figure 4-14, the audio-picture annotation group consistently outperformed the text-picture annotation group, regardless of different posttest measures. The difference was statistically significant for both the Vocabulary Knowledge Scale (effect size $f = .33$) and Word Recognition Test (effect size $f = .41$).

The dual channel assumption, especially the modality principle, of the cognitive theory of multimedia learning can be used to explain this finding (Mayer, 2001). Mayer distinguishes the two separate channels for processing visual/pictorial and auditory/verbal
information. The modality effect articulates that working memory has partially independent processors for handling visual and auditory information. The effective capacity of working memory could be increased by using both visual and auditory channels (Mayer & Moreno, 1998; Mousavi, Low & Sweller, 1995; Penny, 1989).

Text annotation and audio annotation are both verbally-presented information, and picture annotation is non-verbally-presented (pictorial) information; thus both multimedia annotations contain a combination of verbal and non-verbal information. Based on the modality principle (Baddeley, 1986, 1992, 1999; Mayer, 2001), text annotation and picture annotation will be processed by the visual channel, while audio annotation will be processed by the auditory channel. Therefore, in text-picture annotations, the simultaneous register of both text and picture caused the visual channel to be overloaded. This led to an information processing that was, at least initially, carried out solely in the visual working memory. Thus, the cognitive resources available in the visual working memory had to be divided between textual and pictorial information, whereas the auditory (phonological) working memory was left unused.

In comparison, in audio-picture annotations, the audio was registered by the auditory channel and processed in the phonological working memory, while the picture was registered by the visual channel and processed in the visual working memory. This combination allowed cognitive resources in both working memories to be used. In other words, more cognitive resources were utilized in audio-picture annotations than in text-picture annotations.

The superiority of audio-picture annotation on L2 vocabulary immediate recall can also be explained with the split-attention principle (Mayer & Moreno, 1998;
Mousavi, Low, & Seller, 1995). Participants with access to text-picture annotations had to split their attention in the visual working memory between multiple visual resources (written text and picture). Participants with access to audio-picture annotations approached the audio as an auditory resource and the picture as a visual resource through auditory working memory and visual working memory respectively, which did not require an attention split in either of the working memories. In this way, effective working memory might be increased by presenting information in a mixed (visual and auditory) rather than a unitary mode (visual only). Hence, audio-picture annotation resulted in higher vocabulary immediate recall than text-picture annotation.

Effect of Annotation Type on L2 Reading Comprehension

The second research question focused on the effects of different multimedia annotations on L2 reading comprehension. The two multimedia annotations compared were text-picture and audio-picture. The ANOVA results did not show any significant effect for annotation type on L2 reading comprehension for either of the two comprehension measures (Reading Comprehension Test, $F = .44$, $p = .5089$; and Written Recall, $F = 3.38$, $p = .0700$). This fails to support the modality effect of the cognitive theory of multimedia learning (Mayer, 2001).

Studies have reported the superiority of a combination of audio and picture in comparison to a combination of text and picture when presenting new knowledge (Mayer & Anderson, 1991; Mayer & Moreno, 1998; Moreno & Mayer, 1999). According to the modality principle, audio-picture annotation simultaneously engages both the visual working memory and auditory working memory, while text-picture annotation involves
only the visual working memory; therefore, audio-picture annotation enables more application of available cognitive capacity and should consequently lead to more content recall. The results from this study did not indicate the superiority of audio-picture annotations over text-picture annotations.

If reexamined from the data results, as indicated by Table 4-22 and Figure 4-15, text-picture annotations, in contrast to previous findings, resulted in better results (RC M = 6.77, Recall M = 14.95) than audio-picture annotations (RC M = 6.49, Recall M = 13.79), with larger differences in the Written Recall measure. The difference was statistically significant in the intentional learning condition. This could be tentatively explained from the perspective of learning style. Most of the participants (78%) were from Asian countries (Japan, Taiwan, Korean, etc). Previous studies have reported on the preference of visual learning styles over auditory learning styles among Asian students (e.g. Ye & Wang, 2003). For this study, the visual information presented via text-picture annotation might act as better retrieval cues, in comparison to the visual and auditory information presented via audio-picture annotation, to help the participants when taking the comprehension tests.

**Effect of Learning Condition on L2 Vocabulary Immediate Recall**

The third research question dealt with the effects of learning condition on L2 vocabulary immediate recall. In particular, two learning conditions, incidental and intentional, were compared in this study. As shown by Table 4-23 and Figure 4-16, the intentional learning condition led to better vocabulary immediate recall (VKS M = 26,
WRT M= 16.92) than the incidental learning condition (VKS M = 21.9, WRT M = 14.95). The differences in both tests were statistically significant (effect size f was respectively .30 for VKS and .50 for WRT). The finding confirms Konopak et al (1987) who reported significantly better vocabulary learning in the intentional learning than in the incidental learning condition.

However, this differs from the results obtained by Ko (1995). In Ko’s (1995) study, intentional learning was only found to be slightly better than incidental learning (M=18.70 vs. M=18.09), and the difference was not significant. Ko (1995) explained the non-significant difference between learning conditions in terms of the attention that students paid to the annotations. Student reaction to the annotation as collected via a questionnaire indicated a small gap between the two learning conditions in the amount of effort they put into the annotation. In other words, participants in the incidental learning condition paid as much attention to the target words as their counterparts in the intentional learning condition.

Schmidt (1994) points out that incidental learning is the condition in which an individual learns “without the intent to learn or the learning of one thing (e.g. grammar) when the learner’s primary objective is to do something else (e.g. communicate)” (p. 9). This study followed exactly the division between incidental learning and intentional as defined by Hulstijn (1993) by providing different objectives for different groups. In incidental learning, learner attention was not focused on the target words, while in intentional learning, learner attention was focused on both the target words and the reading text. Therefore, although in incidental learning, learners might pick up some new words as a by-product while reading, their attention was to understand the meaning of the
reading text; hence, those new words could be temporarily held in short-term memory, but could not be processed into long-term memory due to the lack of attention. In contrast, however, intentional learning, in which the intention was generated by a per-determined objective, motivated the learner’s attention to be focused on the target words, thus leading to better word retention after reading.

The learning of a new word in a second language depends on how much mental effort the reader devotes to that particular word, especially for productive vocabulary. An interesting finding worth mentioning is that the gap between the two learning conditions was larger in the Vocabulary Knowledge Scale than in the Word Recognition Test (as seen from Figure 4-11). This could be explained by the different nature of the two tasks. A Word Recognition Test is used to measure receptive vocabulary, while a Vocabulary Knowledge Scale tends to measure both receptive and productive knowledge of the target words.

**Effects of Learning Condition on L2 Reading Comprehension**

This study attempted to compare two learning conditions, incidental and intentional, in their effects on L2 reading comprehension. Since multimedia annotations were provided for target words in the reading text, the two conditions were predetermined on the objective for vocabulary immediate recall. In addition, vocabulary knowledge is regarded as an important predictor for L2 reading comprehension (Bossers, 1992, cited in Pulido, 2004; Haynes & Baker, 1993; Laufer, 1992). Therefore, the discussion is actually focused on comparing the effects of incidental vocabulary learning and intentional vocabulary learning on L2 reading comprehension.
It was expected that the incidental vocabulary learning condition would lead to better reading comprehension because the participants were not responsible for vocabulary retention and their attention was steadily focused on comprehending the text. While in intentional vocabulary learning, the participants were responsible for vocabulary learning and their attention was focused on learning the target words. Due to limited cognitive capacity, participants would shift more attention to vocabulary learning and pay less attention to comprehending the text. In other words, the intentional vocabulary learning condition more seriously interrupts the reading process and results in worse reading comprehension.

As shown by Table 4-24 and Figure 4-17, the results for multiple-choice Reading Comprehension Test were very close for the two learning conditions (incidental $M = 6.45$, intentional $M = 6.82$). However, the results for the Written Recall presented a better performance by the incidental learning condition participants ($M = 16.3$) than by the intentional learning condition participants ($M = 12.34$), and the difference was statistically significant (effect size $f = .34$).

This could be explained from the perspective of focused attention (Ko, 1995). In the incidental learning condition, participants were told in advance that they would have reading comprehension tests; it was anticipated that they would focus their attention solely on the content. Whereas in intentional learning condition, participants were told in advance that they would have both vocabulary tests and reading comprehension tests, thus focusing their attention on both the content and vocabulary simultaneously. The focused attention in the incidental learning condition resulted in better reading comprehension than divided attention in the intentional learning condition.
From another point of view, the Multiple-choice Reading Comprehension test and The Written Recall are two different tasks. Different tasks may, by their nature, influence an L2 reader’s performance on reading comprehension (Lee, 1986; Riley & Lee, 1996; Wolf, 1996). A multiple-choice test is a discrete-point task that focuses on isolated bits of information. In contrast, written recall is an integrative task that focuses on global understanding of the reading text. Researchers (Bernhardt, 1983, 1991; Swaffer et al., 1991) have argued for the use of recall protocol, instead of multiple-choice test, as a measure of global or holistic reading comprehension. A multiple-choice test itself suggests possible answers and it sometimes tends to be independent of the reading text (Berhardt, 1983). In addition, the questions in the multiple-choice test, especially those distractors, might cause the reader to change his or her representation of the reading text. In comparison, written recall seems to more accurately reflect comprehension, especially in its propositional content of the text reconstruction. It is regarded to be more text-dependent as articulated by Berhardt (1991) that recall is a good measure because “generating recall data does not influence the reader’s understanding of a text” (p. 200).

Interaction between Annotation Type and Learning Condition

The fifth and sixth research questions focused on the interaction between annotation type and learning condition on L2 vocabulary immediate recall and reading comprehension. As shown by Table 4-25, there was no interaction between annotation type and learning condition on L2 vocabulary immediate recall. This means that the superiority of audio-picture over text-picture in facilitating L2 vocabulary immediate recall does not depend on learning condition.
The interaction between annotation type and learning condition on L2 reading comprehension needs to be addressed by the different comprehension measures. For the Multiple-choice Reading Comprehension Test, there was no interaction between the two factors. For the Written Recall, however, the interaction between annotation type and learning condition was found to be significant. This means that the performance of the participants on the written recall in different annotation types depends on learning condition. Upon closer examination of the results for research question 6 (see Figure 4-13), the scores on written recall by the participants in the audio-picture annotations declined as the learning condition changed from incidental to intentional, while the score by the participants in the text-picture annotations did not change much in the two learning conditions. The decline in audio-picture reflected the focus-attention principle used to explain the better reading comprehension performance in the incidental learning group, but the performance in text-picture did not reflect this principle.

Additionally, whether the effects of learning conditions on L2 reading comprehension are related to annotation type depends on the task type. As indicated by Table 4-26, the multiple-choice reading comprehension Test, as a discrete-point testing method, did not reveal any interaction between learning condition and annotation type. However, the written recall protocol, as an overall comprehension testing method, did present a significant interaction between learning condition and annotation type. This reflects that the choice of different task types might cause different findings.
**Theoretical Implications**

This dissertation adds to the growing body of research in multimedia annotation studies in second language acquisition. Previous multimedia annotation studies have focused on the comparison of text-picture annotation to text-only annotation or picture-only annotation (Jones, 2003; Yoshii, 2000) or on the differences between text-picture annotation and text-video annotation (Al-Seghayer, 2001; Chun & Plass, 1996a, 1996b). However, audio annotation, as a different sensory modality from visual (text, picture), has never been studied before. The present study fills this gap in the literature.

This study provided the much-needed information on the effect of audio annotation on L2 vocabulary learning. By comparing audio-picture annotation to text-picture annotation, it shed light on the use of different dual annotations for multimedia L2 learning. The dissertation has established that audio-picture annotation is superior over text-picture annotation in facilitating L2 vocabulary immediate recall. This contributes to the extension of the cognitive theory of multimedia learning to second language learning by verifying both the modality effect and split-attention effect. In addition, the present study has determined that in multimedia environments, the intentional learning condition leads to better vocabulary immediate recall than the incidental learning condition. This adds to the existing literature of promoting intentional vocabulary learning.

In terms of L2 reading comprehension, the incidental learning condition was found to be more facilitative of L2 reading comprehension, as reflected in the written recall protocol. However, the learning condition was not found to have any effect on the multiple-choice reading comprehension test.
Although this dissertation has made important contributions to multimedia annotation research in second language acquisition, some questions still remain unanswered, such as the long-term effects of different multimedia annotations, the effects of different dual multimedia annotations on L2 reading comprehension, and the effects of incidental and intentional vocabulary learning conditions on L2 reading comprehension as measured by different tasks.

*Pedagogical Implications*

In addition to the contributions and implications for the field of second language acquisition, especially in the area of multimedia annotation research, this study carried pedagogical implications.

First of all, the study provides some insights for CALL material designers in choosing the right combination of modalities in facilitating L2 vocabulary learning. This study confirmed that the use of audio-picture combinations facilitates L2 vocabulary immediate recall in a more effective manner than text-picture annotation. In designing multimedia courseware or materials, this finding could be taken into consideration when making decisions about presenting information in different modes. This could also inform language teachers and administrators in making decisions about the most effective multimedia programs to enhance L2 vocabulary learning.

Participants in the intentional learning condition outperformed participants in the incidental learning condition in L2 vocabulary immediate recall. Different learning conditions stimulate different allocation of students’ attention depending on what is regarded as important as defined by the learning objective. With respect to vocabulary
learning, intentional learning should be encouraged if the final learning goal is vocabulary acquisition. This implies that drawing students’ attention to a specific learning objective when they are engaged in a task can have a positive impact on their learning performance.

In terms of measuring L2 reading comprehension, this study shows that written recall might be a more reliable and accurate testing method to gauge L2 learners’ overall or global comprehension of the reading text. This informs language teachers in choosing the most appropriate tasks based on the purpose of testing, the nature of the reading text and proficiency of the learners. In addition, it seems that incidental vocabulary learning results in better reading comprehension because the students’ attention is focused on the content of the reading text.

**Directions for Future Research**

Due to the nature of this study, only immediate posttests were used to measure L2 vocabulary recall. Future research should involve delayed posttests to attest whether the superior effect of audio-picture annotation can be retained over time. In addition, instead of conducting comparative studies, research could focus on a particular multimedia dual annotation such as audio-picture to examine its effects on L2 vocabulary immediate recall as well as long-term retention.

This study tracked the frequency records of participants’ look-up behaviors. Although the look-up behavior was not found to be correlated to the posttest measures, it is worthwhile for future research to use a device to keep track of the length of time participants spent on the annotations and include time-on-task as a possible variable. In
addition, possible Hawthorne Effect should be cautioned when participants getting
different treatments are placed in one room.

Further study should be carried out to investigate the effects of different
annotations on L2 reading comprehension. Since this study did not find any significant
difference between text-picture annotation and audio-picture annotation on L2 reading
comprehension, it might be a better approach to exclusively examine their effects on
reading comprehension.

This study involved participants from multiple first languages. The L1 written
recall task might, in some way, have skewed the recall data during the translation process
because different languages have different rhetorical styles. Future research could
engage more linguistically homogeneous participants in order to alleviate this
shortcoming. Furthermore, this study had different results for reading comprehension
from the different tasks. Other holistic comprehension measures such as summary or
think-aloud protocol could be used as complementary methods to measure
comprehension.

This dissertation used intermediate ESL students as the participants. More cross-
sectional studies should be carried out to examine the effects of multimedia annotations.
As well, instead of using a between-subjects design, a within-subjects design could be
employed to decrease the influence of individual differences.

Conclusions

Previous studies have examined the effects of multimedia annotations on L2
vocabulary learning and reading comprehension. These studies have supported the
effectiveness of multimedia annotations in facilitating L2 vocabulary learning. However, no study in second language acquisition has examined audio annotation in combination with text as a dual multimedia annotation type. This dissertation focused on this issue by comparing audio-picture annotation to text-picture annotation in their effects on L2 vocabulary immediate recall and reading comprehension under incidental and intentional learning conditions.

The results of the study demonstrate that audio-picture annotation is more effective than text-picture annotation in facilitating L2 vocabulary immediate recall, an intentional learning condition is more effective than an incidental learning condition in promoting L2 vocabulary immediate recall. Furthermore, an incidental vocabulary learning condition results in better reading comprehension in written recall than in an intentional vocabulary learning condition.
References


Merriam-Webster Online Dictionary. Available online: http://www.m-w.com/


Language Teaching and Learning (Technical Report No.9) (pp. 1-64). Honolulu: University of Hawai’i at Manoa.


Appendices
Appendix A: Questionnaire

**Personal Information:**
Identification code (ID): __________ Computer Number: __________
Gender: MALE______ FEMALE______
Age Range: 18-25____ 26-35____ 36-45____
What is your home country? ____________
What is your native language? ___________
What other language(s) do you speak? (List all) ________________
How long have you been studying in America? ________________

Please check one for each of the following questions:

1. Do you like to use computers for your study?
   - [ ] Very much  [ ] Yes  [ ] It is ok  [ ] Not very much  [ ] Not at all

2. Is it easy for you to use computers?
   - [ ] Very much  [ ] Yes  [ ] It is ok  [ ] Not very much  [ ] Not at all

3. Do you like to read on computers?
   - [ ] Very much  [ ] Yes  [ ] It is ok  [ ] Not very much  [ ] Not at all

4. Are you good at using computers?
   - [ ] Very much  [ ] Yes  [ ] It is ok  [ ] Not very much  [ ] Not at all

5. Do you like to look new words up in a dictionary?
   - [ ] Very much  [ ] Yes  [ ] It is ok  [ ] Not very much  [ ] Not at all

6. How often do you use computers in a week?
   - [ ] Almost every day  [ ] 3-6 times  [ ] 1-3 times  [ ] never

7. How often do you use Internet in a week?
   - [ ] Almost every day  [ ] 3-6 times  [ ] 1-3 times  [ ] never

8. How often do you do reading in English online in a week?
   - [ ] Almost every day  [ ] 3-6 times  [ ] 1-3 times  [ ] never

9. How often do you look it up in a dictionary when meeting a new word in online reading?
   - [ ] All the time  [ ] Often  [ ] sometimes  [ ] never

10. How often do you use emails in a week?
    - [ ] Almost every day  [ ] 3-6 times  [ ] 1-3 times  [ ] never
Appendix B: Reading Text

The European Settlers of Australia

During the early 1800s, European people began to settle in Australia. They lived along the coast. Most of them thought that the inland of Australia was a prairie. In fact, a mountain range separates the coast from the inland. The mountain range is called the Great Dividing Range. To the new settlers, it seemed like a labyrinth of unexplored valleys and mountains. Burke and Wills successfully crossed the Great Dividing Range in 1830.

After that, European settlers started to settle beyond the mountains. Most of them were shepherds. They raised animals for meat. They wore old clothes, ate simple foods, and lived in small houses. Sometimes the native people attacked them. Other times, animals such as emus attacked them with sharp beaks. They didn’t have armor to protect themselves.

Banjo Patterson was one of the most famous early settlers. He was a songwriter. He wrote Australia’s most famous song: Waltzing Matilda. The song is about a young traveler. With nothing to do, the young man sat under a tree, put his satchel on the ground and played his accordion. Then he saw a herd of cows and decided to steal one to eat.

At that time, life was hard and people hated stealing. The owner of the cow reported the loss to the police. A local policeman caught the young traveler. The policeman took away the young man’s saddle and horse, and put him in prison. He used his cornet to call for a meeting at the center of the town. After the meeting, the young man was made to stand on a barrel. The policeman put a noose around the young man’s neck and killed him.

Most Australian songs are not that sad, but Waltzing Matilda tells the stories of the early colonists in the first part of the 1800s in Australia. By the 1850s, however, life for these early settlers became better. Gold was discovered in the colony of Victoria. As a result, thousands of new settlers came and made Australia their home. Many of them became rich through the gold rush. In many stories of that time, a damsel had nothing better to do than to be beautiful. She just drank from a golden goblet, looked pretty for handsome young men and lived in a big mansion.

More people came to Victoria for gold. The early settlers were afraid of the newcomers. They organized soldiers to keep order. At times, the newcomers walked on the streets and waved placards as a sign of protest. On the placards, they drew pictures of phantoms. Sometimes they threw twigs at the soldiers. However, the soldiers, each armed with a bayonet, easily ended the protests.
Appendix C: Target Words

This is a list of the 20 target words. The definitions are used in the text annotation and audio annotation.

1. prairie: a larger area of treeless grassland.
2. labyrinth: a confusing set of connecting paths.
3. shepherd: a person who takes care of sheep.
4. beak: the hard pointed part of a bird’s mouth.
5. armor: metal clothing for protection in battles.
6. satchel: a small bag with a shoulder strap.
7. accordion: a box-shaped musical instrument.
8. herd: a large number of animals living together.
9. saddle: a leather seat for the rider of a horse.
10. cornet: a musical instrument that you play by blowing into it.
11. barrel: a large container with flat top and bottom.
12. noose: a rope tied in a circle.
13. colonist: one who settles in a new country.
15. goblet: a drinking glass with a base and a stem.
16. mansion: a large and impressive house.
17. placard: a sign people hold in a demonstration.
18. phantom: a ghost.
19. twig: a small branch of a tree.
20. bayonet: a long sharp blade fixed at the end of a gun.
## Appendix D: Pictures for Target Words

<table>
<thead>
<tr>
<th>prairie</th>
<th>labyrinth</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Prairie" /></td>
<td><img src="image" alt="Labyrinth" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>shepherd</th>
<th>beak</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Shepherd" /></td>
<td><img src="image" alt="Beak" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>armor</th>
<th>satchel</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Armor" /></td>
<td><img src="image" alt="Satchel" /></td>
</tr>
<tr>
<td>accordion</td>
<td>Herd</td>
</tr>
<tr>
<td>-----------</td>
<td>------</td>
</tr>
<tr>
<td><img src="image1.png" alt="Accordion" /></td>
<td><img src="image2.png" alt="Herd" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Saddle</th>
<th>Cornet</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image3.png" alt="Saddle" /></td>
<td><img src="image4.png" alt="Cornet" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>barrel</th>
<th>Colonist</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image5.png" alt="Barrel" /></td>
<td><img src="image6.png" alt="Colonist" /></td>
</tr>
<tr>
<td>noose</td>
<td>Damsel</td>
</tr>
<tr>
<td>------------</td>
<td>-------------</td>
</tr>
<tr>
<td><img src="image" alt="noose" /></td>
<td><img src="image" alt="Damsel" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>goblet</th>
<th>Mansion</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="goblet" /></td>
<td><img src="image" alt="Mansion" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>placard</th>
<th>Phantom</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="placard" /></td>
<td><img src="image" alt="Phantom" /></td>
</tr>
</tbody>
</table>
The European Settlement of Australia

During the early 1800s, European people began to settle in Australia. They lived along the coast. Most of them thought that the island of Australia was a swamp. In fact, a mountain range separates the coast from the island. The mountain range is called the Great Dividing Range. To the new settlers, it seemed like a jungle, of unexplored valleys and mountains. Burke and Wills successfully crossed the Great Dividing Range in 1860.

After that, European settlers started to settle beyond the mountains. Most of them were shepherds. They raised sheep. They wore old clothes, ate simple foods, and lived in small houses. Sometimes the native people attacked them. Other times, animals such as emus attacked them with厉害. They didn’t have arms to protect themselves.

Banjo Patterson was one of the most famous early settlers. He was a composer. He wrote Australia’s most famous song, Waltzing Matilda. The song is about a young miner. With nothing to do, the young man sat under a tree, put his guitar on the ground, and played his accordion. Then he saw a herd of cows and decided to steal one to eat.

At that time, life was hard and people hated stealing. The owner of the cow reported the theft to the police. A local policeman caught the young miner. The policeman took away the young man’s guitar, and drove, and put him in prison. He used it to call for a meeting at the center of the town. After the meeting, the young man was made to stand on a ladder. The policeman put a musket around the young man’s neck, and killed him.

Most Australian ranges are not like that, but Waltzing Matilda tells the stories of the early settlers in the first part of the 1800s in Australia. By the 1800s, however, life for those early settlers became better. Gold was discovered in the colony of Victoria. As a result, thousands of new settlers came and made Australia their home. Many of them became rich through the gold rush. In many stories of that time, a farmer had nothing better to do than to be happy. He just drank from a golden goblet, looked pretty for handsome young men and lived in a big mansion.

Most people came to Victoria for gold. The early settlers were afraid of the movements. They formed soldiers.
During the early 1800s, European people began to settle in Australia. They lived along the coast. Most of them thought that the island of Australia was a big rock. In fact, a mountain range separates the coast from the island. The mountain range is called the Great Dividing Range. To the new settlers, it seemed like a nightmare of unexplored valleys and mountains. Burke and Wills successfully crossed the Great Dividing Range in 1839.

After that, European settlers started to settle beyond the mountains. Most of them were farmers. They raised animals for meat. They wore old clothes, ate simple foods, and lived in small houses. Sometimes the native people attacked them. Other times, animals such as snakes attacked them with sharp fangs. They didn’t have enough to protect themselves.

Ralph’s mother was one of the first farmers. He was a young farmer. He wrote Australia’s most famous song, Waltzing Matilda. The song is about a young worker. With nothing to do, the young man sat under a tree, put his guitar on the ground, and played his guitar. Then he saw a bird of prey and decided to steal one to eat.

At that time, life was hard and people hated working. The owner of the farm reported the man to the police. A local policeman caught the young man. The policeman took away the young man’s guitar and horse, and put him in prison. He used him to call for a meeting at the center of the town. After the meeting, the young man was made to stand on a horse. The policeman put it around the young man’s neck and killed him.

Most Australian songs are not that bad, but Waltzing Matilda tells the stories of the early colonists in the first part of the 1800s in Australia. By the 1800s, however, life for these early settlers became better. Gold was discovered in the colony of Victoria. As a result, thousands of new settlers came and made Australia their home. Many of them became rich through the gold rush. In many stories of that time, a young man had nothing to do except to be brave. She just drank from a golden goblet, looked pretty for bands, and lived in a big manor.

Most people came to Victoria for gold. The early settlers were afraid of the newcomers. They remained soldiers.
Appendix G: Vocabulary Knowledge Scale (VKS)

Directions: For each word, there are three choices of how much you know about the word, please circle the one that fits you most. If you choose III, please written down the meaning of the word.

<table>
<thead>
<tr>
<th>1. shepherd</th>
<th>Categories</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>I don’t remember having seeing this word before</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>I have seen this word before, but I don’t know what it means</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>I have seen this word before, and I think it means</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2. placard</th>
<th>Categories</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>I don’t remember having seeing this word before</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>I have seen this word before, but I don’t know what it means</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>I have seen this word before, and I think it means</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3. prairie</th>
<th>Categories</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>I don’t remember having seeing this word before</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>I have seen this word before, but I don’t know what it means</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>I have seen this word before, and I think it means</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4. phantom</th>
<th>Categories</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>I don’t remember having seeing this word before</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>I have seen this word before, but I don’t know what it means</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>I have seen this word before, and I think it means</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>5. noose</th>
<th>Categories</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>I don’t remember having seeing this word before</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>I have seen this word before, but I don’t know what it means</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>I have seen this word before, and I think it means</td>
<td></td>
</tr>
</tbody>
</table>
### Appendix G: Vocabulary Knowledge Scale (VKS) (Continued)

<table>
<thead>
<tr>
<th>6. herd</th>
<th>Categories</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>I don’t remember having seeing this word before</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>I have seen this word before, but I don’t know what it means</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>I have seen this word before, and I think it means __________________________</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>7. twig</th>
<th>Categories</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>I don’t remember having seeing this word before</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>I have seen this word before, but I don’t know what it means</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>I have seen this word before, and I think it means __________________________</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>8. damsel</th>
<th>Categories</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>I don’t remember having seeing this word before</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>I have seen this word before, but I don’t know what it means</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>I have seen this word before, and I think it means __________________________</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>9. colonist</th>
<th>Categories</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>I don’t remember having seeing this word before</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>I have seen this word before, but I don’t know what it means</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>I have seen this word before, and I think it means __________________________</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>10. saddle</th>
<th>Categories</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>I don’t remember having seeing this word before</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>I have seen this word before, but I don’t know what it means</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>I have seen this word before, and I think it means __________________________</td>
<td></td>
</tr>
<tr>
<td>11. <strong>labyrinth</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>Categories</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>I don’t remember having seeing this word before</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>I have seen this word before, but I don’t know what it means</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>I have seen this word before, and I think it means</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>12. <strong>cornet</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Categories</td>
<td>Description</td>
</tr>
<tr>
<td>I</td>
<td>I don’t remember having seeing this word before</td>
</tr>
<tr>
<td>II</td>
<td>I have seen this word before, but I don’t know what it means</td>
</tr>
<tr>
<td>III</td>
<td>I have seen this word before, and I think it means</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>13. <strong>armor</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Categories</td>
<td>Description</td>
</tr>
<tr>
<td>I</td>
<td>I don’t remember having seeing this word before</td>
</tr>
<tr>
<td>II</td>
<td>I have seen this word before, but I don’t know what it means</td>
</tr>
<tr>
<td>III</td>
<td>I have seen this word before, and I think it means</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>14. <strong>bayonet</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Categories</td>
<td>Description</td>
</tr>
<tr>
<td>I</td>
<td>I don’t remember having seeing this word before</td>
</tr>
<tr>
<td>II</td>
<td>I have seen this word before, but I don’t know what it means</td>
</tr>
<tr>
<td>III</td>
<td>I have seen this word before, and I think it means</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>15. <strong>beak</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Categories</td>
<td>Description</td>
</tr>
<tr>
<td>I</td>
<td>I don’t remember having seeing this word before</td>
</tr>
<tr>
<td>II</td>
<td>I have seen this word before, but I don’t know what it means</td>
</tr>
<tr>
<td>III</td>
<td>I have seen this word before, and I think it means</td>
</tr>
</tbody>
</table>
# Appendix G: Vocabulary Knowledge Scale (VKS) (Continued)

## 16. barrel

<table>
<thead>
<tr>
<th>Categories</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>I don’t remember having seen this word before</td>
</tr>
<tr>
<td>II</td>
<td>I have seen this word before, but I don’t know what it means</td>
</tr>
<tr>
<td>III</td>
<td>I have seen this word before, and I think it means</td>
</tr>
</tbody>
</table>

## 17. accordion

<table>
<thead>
<tr>
<th>Categories</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>I don’t remember having seen this word before</td>
</tr>
<tr>
<td>II</td>
<td>I have seen this word before, but I don’t know what it means</td>
</tr>
<tr>
<td>III</td>
<td>I have seen this word before, and I think it means</td>
</tr>
</tbody>
</table>

## 18. satchel

<table>
<thead>
<tr>
<th>Categories</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>I don’t remember having seen this word before</td>
</tr>
<tr>
<td>II</td>
<td>I have seen this word before, but I don’t know what it means</td>
</tr>
<tr>
<td>III</td>
<td>I have seen this word before, and I think it means</td>
</tr>
</tbody>
</table>

## 19. goblet

<table>
<thead>
<tr>
<th>Categories</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>I don’t remember having seen this word before</td>
</tr>
<tr>
<td>II</td>
<td>I have seen this word before, but I don’t know what it means</td>
</tr>
<tr>
<td>III</td>
<td>I have seen this word before, and I think it means</td>
</tr>
</tbody>
</table>

## 20. mansion

<table>
<thead>
<tr>
<th>Categories</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>I don’t remember having seen this word before</td>
</tr>
<tr>
<td>II</td>
<td>I have seen this word before, but I don’t know what it means</td>
</tr>
<tr>
<td>III</td>
<td>I have seen this word before, and I think it means</td>
</tr>
</tbody>
</table>
Appendix H: Word Recognition Test (WRT)

For each underlined word, please circle the correct meaning from the four choices.

1. Most of these settlers were shepherds.
   a. a person who travels in the forest
   b. a person who takes care of sheep
   c. a person who digs gold for a living
   d. a person who buys and sells sheep

2. To the new settlers, it seemed like a labyrinth.
   a. a group of scattered houses
   b. a set of connecting roads on a map
   c. a confusing set of connecting paths
   d. a group of animals living together

3. The policeman took away the young man’s saddle.
   a. a small bag
   b. a leather seat
   c. a handgun
   d. a musical instrument

4. Most of them thought that the inland of Australia was a prairie.
   a. a large area of wetland
   b. a large area of grassland
   c. a large area of desert
   d. a large area of forest

5. They didn’t have armor to protect themselves.
   a. wooden weapons
   b. metal tools
   c. wooden housing
   d. metal clothing

6. The young traveler put his satchel on the ground.
   a. a small bag with a shoulder strap
   b. a wine glass with a base and a handle
   c. a leather seat used to ride horses
   d. a long sharp knife used by soldiers

7. In many stories of that time, a damsel does nothing all the time.
   a. an unmarried girl
   b. a housewife
   c. a young soldier
   d. a rich man
Appendix H: Word Recognition Test (WRT) (Continued)

8. They drew a picture of a **phantom** on the wall.
   a. a horse
   b. a ghost
   c. a person
   d. a sheep

9. The policeman put a **noose** around the traveler’s neck.
   a. a long necktie
   b. a long metal chain
   c. a rope tied in a circle
   d. a thin piece of cloth

10. The newcomers threw **twigs** at the soldiers.
    a. a tree branch
    b. a wood chip
    c. a small stone
    d. a smelly plant

11. She drank from a golden **goblet**.
    a. a drinking glass
    b. a water bowl
    c. a table spoon
    d. a metal plate

12. The policeman used his **cornet** to call for a meeting.
    a. a box-shaped musical instrument that you play with hands
    b. a whistle used by policemen to get people’s attention
    c. a musical instrument that you play by blowing into it
    d. a big bell used to inform people of the time in the past

13. The traveler saw a **herd** of cows.
    a. a number of people living together
    b. a number of animals living together
    c. a number of travelers on the desert
    d. a number of policemen at the station

14. The **colonist** arrived in Australia in 1800s.
    a. a person who writes songs for travelers
    b. a person who takes care of sheep
    c. a person who settles in a new country
    d. a person who fights for his country
15. The young girl lived in a mansion.
   a. a large house
   b. an old castle
   c. a tall tower
   d. a small cottage

16. Each soldier is armed with a bayonet.
   a. a handgun used by policemen to protect people
   b. a long and sharp blade fixed at the end of a gun
   c. a whistle used by policemen to get attention
   d. a knife used by soldiers to kill animals

17. A gold digger waved a placard at the soldiers.
   a. a flag people wave in public to get attention
   b. a sign people hold in public in a demonstration
   c. a piece of cloth people use to cover their heads
   d. a cotton scarf people wear in cold weather

18. The young man was made to stand on a barrel.
   a. a large tree trunk for people to stand on it
   b. a wooden bench for people to sit on it
   c. a large container with two handles
   d. a large container with flat top and bottom

19. Emus attacked the settlers with sharp beaks.
   a. the long and sharp knife used by hunters
   b. the hard pointed part of a bird’s mouth
   c. the sharp blade at the end of a gun
   d. the pointed horn of a large animal

20. The young traveler played his accordion under the tree.
   a. a whistle used to get attention
   b. a big bell used to tell people time
   c. a box-shaped musical instrument
   d. a musical instrument like a trumpet
Appendix I: Multiple-Choice Reading Comprehension Questions

Based on the reading passage, please circle the best answer for each question:

1. When the European settlers arrived in Australia in the 1800s, they had a misunderstanding about the inland of Australia. What is the misunderstanding?
   A. They thought the inland of Australia was a large forest.
   B. They thought the inland of Australia was a large prairie.
   C. They thought the inland of Australia was a large mountain.
   D. They thought the inland of Australia was a large lake.

2. Which of the following is mentioned as one of the dangerous things the early settlers had to face?
   A. bad weather of the inland
   B. thieves who steal cows
   C. bayonets of the newcomers
   D. animals with sharp beaks

3. The early settlers did NOT have which of the following?
   A. old clothes
   B. metal armor
   C. small houses
   D. animal meat

4. The main character in the song “Waltzing Matilda” is a ______.
   A. shepherd
   B. soldier
   C. songwriter
   D. traveler

5. According to the passage, the character in the song did NOT have ______.
   A. a small bag
   B. an accordion
   C. a horse and saddle
   D. a sharp blade

6. What did the policeman use to call for a meeting at the center of the town?
   A. an accordion
   B. a cornet
   C. a whistle
   D. a goblet
Appendix I: Multiple-choice Reading Comprehension Questions (Continued)

7. How did the main character in the song die?
   A. He was killed by a stone.
   B. He was killed by a gun.
   C. He was killed by a noose.
   D. He was killed by a bayonet.

8. In many stories about the European settlers after the 1850s, a character is often used to describe the life at that time. Who is that character?
   A. a young woman
   B. a settler’s wife
   C. a young soldier
   D. a young shepherd

9. What did the newcomers do to show their dislike of the soldiers?
   A. They moved away from the center of the town.
   B. They walked on the streets and waved signs.
   C. They threw small stones at the soldiers.
   D. They used bayonets to fight the soldiers.

10. What pictures did the newcomers draw on the placards?
    A. pictures of prairies
    B. pictures of satchels
    C. pictures of ghosts
    D. pictures of animals
Appendix J: Written Recall

Directions: Please use your first language to write down as much as you know about the passage you just read.
About the Author

Zhaohui Chen obtained his Bachelor’s degree in English Education from Suzhou Railway Teachers College in 1994 and Master’s degree in English Language & Literatures from Suzhou University in 1997. He worked as an English lecturer in the School of Foreign Language at Suzhou University from 1997 to 2001. In August of 2001, he started his doctoral program in Second Language Acquisition & Instructional Technology at the University of South Florida. While in the PhD program, Mr. Chen taught ESOL classes in both face-to-face and distance learning formats, worked as the treasurer of SLAQ (Student Organization for Second Language Acquisition & Instructional Technology), and presented at conferences.